HHS Public Access

Author manuscript

Nat Genet. Author manuscript; available in PMC 2012 July 01.

Published in final edited form as:

Nat Genet. ; 44(1): 89–93. doi:10.1038/ng.1006.

Large-Scale Discovery of Enhancers from Human Heart Tissue

Dalit May[1], **Matthew J. Blow**[1,2], **Tommy Kaplan**[3,4], **David J. McCulley**[5], **Brian C. Jensen**[6], **Jennifer A. Akiyama**[1], **Amy Holt**[1], **Ingrid Plajzer-Frick**[1], **Malak Shoukry**[1], **Crystal Wright**[2], **Veena Afzal**[1], **Paul C. Simpson**[5,7], **Edward M. Rubin**[1,2], **Brian L. Black**[5], **James Bristow**[1,2], **Len A. Pennacchio**[1,2,*], and **Axel Visel**[1,2,*]

[1]Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720 USA

[2]U.S. Department of Energy Joint Genome Institute, Walnut Creek, CA 94598 USA

[3]Department of Molecular and Cell Biology, California Institute of Quantitative Biosciences, University of California Berkeley, Berkeley, CA 94720 USA

[4]School of Computer Science and Engineering, The Hebrew University, Jerusalem, Israel

[5]Cardiovascular Research Institute, University of California, San Francisco, CA 94158 USA

[6]Division of Cardiology, University of California, San Francisco, CA 94143

[7]Cardiology Division, VA Medical Center, San Francisco, CA 94121

Abstract

Development and function of the human heart depend on the dynamic control of tissue-specific gene expression by distant-acting transcriptional enhancers. To generate an accurate genome-wide map of human heart enhancers, we used an epigenomic enhancer discovery approach and identified ~6,200 candidate enhancer sequences directly from fetal and adult human heart tissue. Consistent with their predicted function, these elements were markedly enriched near genes implicated in heart development, function and disease. To further validate their *in vivo* enhancer activity, we tested 65 of these human sequences in a transgenic mouse enhancer assay and observed that 43 (66%) drove reproducible reporter gene expression in the heart. These results support the discovery of a genome-wide set of non-coding sequences highly enriched in human heart enhancers which is likely to facilitate down-stream studies of the role of enhancers in development and pathological conditions of the heart.

Heart disease is a leading cause of morbidity and mortality in both children and adults and significantly depends on genetic factors [1-5]. Genome-wide association studies indicate that variation in non-coding sequences, including distant-acting transcriptional enhancers, affects

*Correspondence should be addressed to L.A.P. (lapennacchio@lbl.gov) or A.V. (avisel@lbl.gov).
Current address: Division of Cardiology, University of North Carolina, Chapel Hill, NC 27599.

Author Contributions: D.M., E.M.R., J.B., L.A.P. and A.V. conceived and designed the experiments. D.M., M.J.B., T.K., D.J.M., B.C.J., J.A.A., A.H., I.P.-F., M.S., C.W., and V.A. performed experiments and data analysis. P.C.S. and B.L.B. provided reagents/ materials and performed data analysis. All authors contributed to the writing of the paper.

Author Information: The authors declare no competing financial interests.

the susceptibility to many types of human disease [6-10]. However, the possible role of enhancers in heart disease has been difficult to evaluate due to the lack of a human cardiac enhancer catalogue. Mapping of enhancer-associated epigenomic marks via chromatin immunoprecipitation coupled to massively parallel sequencing (ChIP-seq) represents a conservation-independent strategy to discover tissue-specific enhancers [11-14]. It has previously been shown that genome-wide binding profiles of an enhancer-associated co-activator protein, p300, in mouse heart tissue can correctly predict the genomic location of heart enhancers in the mouse genome [15]. However, the sequences identified by this approach tend to be poorly conserved in evolution, suggesting that mouse-derived ChIP-seq data sets are of limited value for accurate annotation of heart enhancers in the human genome.

To generate genome-wide maps of predicted cardiac enhancers in the human genome, we determined the occupancy of two enhancer-associated co-activator proteins in human fetal (gestational week 16) and adult heart. We performed chromatin immunoprecipitation with a pan-specific antibody that recognizes both p300 and the closely related CBP co-activator protein [16-18]. Massively parallel sequencing and enrichment analysis [19] of the aligned sequences from fetal heart tissue identified 5,047 p300/CBP-bound regions (peaks) genome-wide that were located at least 2.5kb from the nearest transcript start site (Fig. 1a/c, Supplementary Table 1, Supplementary Fig. 1, Methods). Likewise, 2,233 regions were identified from adult human heart. Nearly half of the adult human heart enhancer candidates (1,082; 48%) coincided with candidate enhancers derived from fetal human heart. In addition, many peaks identified in one of the samples exhibit read densities above background, but below the peak significance threshold in the respective other sample. In total, 4,257 (84%) of fetal peaks and 2,113 (95%) of adult peaks show significantly or sub-significantly increased read densities in the adult and fetal data set, respectively. This remarkable overlap in data from the two samples suggests that many cardiac p300/CBP binding sites are maintained from prenatal stages of heart development into adulthood (Fig. 1b, Supplementary Fig. 2). These results indicate that thousands of distal p300/CBP binding sites (candidate enhancers) exist in fetal and adult human heart tissue.

Tissue-specific enhancers typically act over distances of tens or hundreds of kilobases [9], therefore authentic cardiac enhancers are expected to be detectably enriched in the larger genomic vicinity of genes that are expressed and functional in the heart. To assess this, we examined the cardiac expression and function of genes located near the ChIP-seq-identified regions. First, we compared the genome-wide set of candidate enhancers to genome-wide gene expression data from human heart tissue. We observed a 4.7-fold enrichment in heart p300/CBP peaks within 2.5-10kb of the transcript start sites of genes highly expressed in fetal human heart ($p < 1e\text{-}40$, binomial distribution), with significant enrichment up to 200kb away from promoters (cumulative fold enrichment 2.6, $p < 0.001$, binomial distribution, Fig. 2a). Enrichment was also observed for human adult candidate enhancers or when RNA polymerase II-binding to gene promoters was used as a complementary approach to identify active genes (Supplementary Note and Supplementary Figs. 3 and 4). In contrast, no enrichment of p300/CBP binding sites was observed near genes highly expressed in other tissues (Fig. 2b). Second, to examine if candidate cardiac enhancers are enriched near genes with known cardiac functions, we performed an unsupervised statistical enrichment analysis

of functional gene annotations [20]. Candidate heart enhancers are indeed associated with genes that have been linked to cardiovascular functions in mouse deletion studies and by gene ontology (GO) annotations [21] (Table 1, Supplementary Table 2). These results indicate that candidate enhancers identified by p300/CBP binding in the human heart are enriched near genes that are expressed and functional in the cardiovascular system, supporting their function as human heart enhancers in the genome.

To assess the potential relevance of predicted heart enhancers to cardiac diseases, we focused on candidate enhancers located in the genomic vicinity of genes associated with different types of congenital and adult heart disorders. We compiled a panel of 73 genes used in genetic diagnosis of heart diseases and found that candidate heart enhancers are significantly enriched near these genes (Supplementary Fig. 5). We identified 81 human fetal heart candidate enhancers that are located within 50kb of 30 of these cardiac genes, an enrichment of 4.5-fold over what would be expected by chance ($p < 7e\text{-}15$, hypergeometric distribution, see Methods). These genes are associated with a variety of cardiac diseases, including conduction disorders (potassium channels *KCNQ1* and *KCNE2*), cardiomyopathies (muscle structural proteins *MYL2*, *ACTN2* and *TNNC1*), and congenital defects (cardiac transcription factors *GATA4*, *NKX2*.5 and *TBX5*) (Supplementary Table 3). To facilitate the study of genetic variation in candidate enhancers, we identified 1,546 known single nucleotide polymorphisms (SNPs) within these 81 candidate enhancers (Supplementary Table 3). In addition, we have identified genetic variants in 11 candidate enhancers that have been linked to changes in expression of genes in human cell types [22] (Supplementary Table 4). While in-depth studies will be required to clarify how these sequence variants affect the *in vivo* activity of individual enhancers, these data highlight the usefulness of a genome-wide set of heart enhancers as an entry-point for the functional exploration of non-coding enhancer variants contributing to cardiac disease.

It was previously shown that embryonic mouse heart enhancers tend to be poorly conserved in evolution [15], raising the possibility that many human heart enhancers might also be under weak evolutionary constraint. Comparison with previously published mouse enhancer data sets confirmed that human candidate heart enhancer sequences also tend to be under relatively weak evolutionary constraint [13,15,23] (Supplementary Fig. 6). Moreover, the vast majority of human peaks (86%) identified in this study were not predicted in previous computational screens combining evolutionary conservation with motif-based prediction of heart enhancers [24]. These observations emphasize the limitations of comparative genomic and computational methods for the discovery of heart enhancers in the human genome.

In light of the apparently limited sequence conservation of human heart candidate enhancers, we also re-examined the utility of mouse-derived ChIP-seq data sets for accurate annotation of human heart enhancers. We performed p300/CBP ChIP-seq on mouse heart tissue at postnatal day 2, a stage at which the developmental progression and gene expression profile of the mouse heart is broadly similar to that of human fetal heart at gestational week 16 [25,26]. Using identical methods as for human heart tissue, we identified 6,564 candidate enhancers (distal ChIP-seq peaks, Supplementary Table 1) and compared them to the human fetal candidate enhancer sets. We find that merely 21% of fetal human heart enhancer candidates coincided with significant peaks at the orthologous site in the mouse genome,

whereas the location of the majority of human peaks could not have been predicted from the mouse data set (Supplementary Fig. 7, Supplementary Note, Methods). While some candidate enhancers may have been missed in the mouse data set due to technical limitations of the approach, these observations suggest that lineage-specific differences in genome-wide enhancer architecture represent a major obstacle for the identification of human heart enhancers and highlight the value of performing ChIP-seq directly on human tissue samples.

To further validate and characterize the ChIP-seq-predicted human heart enhancers, we utilized a transgenic mouse enhancer assay that was previously demonstrated to be an effective approach to define the *in vivo* activity patterns of human and mouse enhancers [13,15,27,28]. We tested 65 candidate human heart enhancers in transgenic mice at embryonic day 11.5 (e11.5). These candidates were selected blind to the identity of nearby genes, focusing on sequences that showed high-confidence p300/CBP-binding in the fetal heart data set. The selected regions were a sampling of sequences from all categories of evolutionary and functional conservation (Supplementary Tables 5 and 6). In total, 43 of 65 (66%) tested sequences drove reproducible expression in the heart or in the vasculature; either exclusively (28, 43%) or as a part of reproducible compound patterns that included the heart (15, 23%) (Fig. 3, selected examples in Fig. 4, Supplementary Table 7). The *in vivo* validation rate of the tested human sequences was significantly higher than the frequency of *in vivo* heart enhancers among conserved sequences near heart-expressed genes or among enhancer predictions derived from human cell lines (Supplementary Table 8, Supplementary Figures 8 and 9) and similar to that of mouse-derived candidate heart enhancer sequences previously tested in this assay (compared to data sets described in ref. [15]; Supplementary Fig. 9). While the tested human sequences in the present study were on average longer than the previously tested mouse sequences (3,719bp vs. 1,648bp, Supplementary Table 9), the predicted human candidate heart enhancers have overall robust cardiac activity and potential human-mouse species incompatibilities have only minor effects on the reproducibility of reporter patterns (Supplementary Table 10). Importantly, we did not observe significant differences in the reproducible activity of the tested elements between the different classes of conservation (Fig. 3b, p=0.9, Chi-square test). The reproducible activity of even weakly conserved human heart enhancers in this assay may be at least partly attributable to enrichment of the same transcription factor binding sites in human and mouse heart enhancer sequences (Supplementary Figure 10). The considerable validation rate in this first-pass screen at e11.5 suggests that the ChIP-seq data sets derived from human heart tissue identify a population of human non-coding sequences that are highly enriched for bona fide human heart enhancers.

To examine whether human ChIP-seq data can be used to predict *in vivo* heart enhancers beyond prenatal stages of development, we also studied sequences predicted to be enhancers active in the adult human heart in adult transgenic mice. In total, 41 of the 65 (63%) candidate enhancers tested at e11.5 are also bound by p300/CBP in the adult human heart and were therefore considered as adult human heart candidate enhancers (Supplementary Fig. 11). We selected eight sequences that had been validated as heart enhancers at e11.5 and examined their *in vivo* activity patterns in four week-old (adult) transgenic mice. All eight enhancers showed reproducible activity in hearts from the adult animals, regardless of

their sequence and binding conservation (Supplementary Table 11). To explore the spatial activity of these enhancers, we examined transverse sections from representative whole-mount stained embryos and longitudinal sections from adult hearts and annotated reproducible reporter staining patterns. While different enhancers drive expression to distinct subregions of the developing and adult heart, we observed a remarkable general concordance between the regions targeted by the same enhancer in the embryonic and adult heart in seven of the eight cases examined (Fig. 4 and Supplementary Fig. 12). These results further support that the ChIP-seq data sets obtained in this study correctly predict the genomic location of human non-coding sequences that are active developmental and adult heart enhancers *in vivo*.

Deciphering the gene regulatory architecture required for development and function of the human heart represents a daunting but pressing challenge. In the present study, we have used an epigenomic approach to identify a genome-wide set of several thousand putative human heart enhancers by performing ChIP-seq for the enhancer-associated proteins p300/CBP on human fetal and adult *ex vivo* heart tissue [11-14]. Through transgenic mouse reporter experiments, we also demonstrated that the majority of these sequences are authentic *in vivo* heart enhancers. Importantly, comparison of these human-derived epigenomic data sets with ChIP-seq data from mouse heart tissue revealed that many heart enhancers are neither evolutionarily nor functionally conserved between human and mouse. Our studies demonstrate that even human heart enhancers whose sequence is not conserved in mice have robust cardiac activity in this assay, suggesting that mouse transcription factors are sufficiently similar to those of humans to correctly report the *cis*-regulatory function of poorly conserved human enhancers. The results from this study add an experimentally defined layer of functional annotation to the human genome and are expected to be critical for elucidating the role of distant-acting enhancer sequences in human heart disorders.

Data availability

All raw sequences and processed data from ChIP-seq experiments are publicly available from NCBI under accession numbers GSE32587. Complete *in vivo* data sets are available from the Vista Enhancer Browser at http://enhancer.lbl.gov

Methods

Human and mouse heart tissue collection and preparation

Adult heart tissue (ischemic, failing, male, age 45y, ejection fraction 31%) was obtained from a heart removed at the time of transplant at the University of California, San Francisco (UCSF) with the approval of the UCSF Committee for Human Research. Full informed consent was obtained from the transplant recipient prior to surgery. Cold cardioplegic solution was perfused antegrade prior to cardiectomy, and the explanted heart was placed immediately in ice-cold physiologic solution. Samples were cleaned rapidly of all epicardial fat, flash frozen in liquid nitrogen, and stored at −80°C. Fresh fetal heart tissues from gestational week 16 were obtained from ABR, Inc. (Alameda, CA) in compliance with applicable state and federal laws and with full informed consent. Postnatal day 2 mouse heart tissues were isolated and pooled from approximately 15 CD-1 strain mice. All

procedures of this study involving human and animal tissue samples were reviewed and approved by the respective institutional Human and Animal Regulatory Committees at Lawrence Berkeley National Laboratory.

Tissue samples were processed for ChIP and DNA sequencing as described previously [13,15] except for the following modifications: following cross-linking in 1% formaldehyde, fresh samples were dissociated in a glass dounce homogenizer and frozen tissues were homogenized using a Polytron homogenizer. Chromatin was sheared using a Bioruptor (Diagenode) and immunoprecipitated using 40 ul anti-acetyl-CBP/p300 antibody (rabbit polyclonal #4771, Cell Signaling Technology) that recognizes these proteins in their active, acetylated state[30,31] or 5 ug anti-RNA polymerase II antibody (mouse monoclonal ab817, Abcam).

Processing of ChIP-sequence data

Sequence reads (36 bp) were aligned to the reference genomes (NCBI build 36, hg18 and NCBI build 37, mm9) using bwa [32]. Repetitively mapped and duplicate reads were excluded and alignments were extended to 300 bp to account for the estimated sequenced DNA fragment length. Cumulative genome-wide coverage of extended reads was calculated at 25 bp resolution.

High confidence data sets of CBP/p300 bound regions were identified from the intersection of MACS [11] (version 1.4, with default settings except --bw=300 --p=1e-5 --mfold=10,30 --llocal=20000 --off-auto --shiftsize=100) and the Grizzly Peak Fitting algorithm (http://eisenlab.org/software/grizzly, modified from Capaldi et al. [19]). Grizzly peaks with a separating gap of less than the length of either of the two peaks or by less than 1kb were merged into single regions. Peaks were scored and sorted based on maximal occupancy, as called by Grizzly. Bound regions mapped to unassembled chromosomal contigs, centromeric regions, telomeric regions, segmental duplications, as well as regions where >50% of the contributing reads originate in repeats were removed from further analyses as likely artifacts.

p300/CBP-bound regions were classified according to their proximity to transcription start sites of UCSC known genes [33]. Regions within 2.5kb of the nearest transcript start site were defined as promoter peaks. The remaining peaks represent candidate distant-acting heart enhancers.

To assess the reproducibility of p300/CBP ChIP-seq, we performed ChIP-seq from a second human fetal heart sample. Enrichment of reads from human fetal heart ChIP-seq samples in the vicinity of candidate adult heart enhancers was performed using custom scripts. For comparison, a generic human input DNA sample was generated by combining 1 million randomly sampled reads from input DNA samples from each of 9 human cell lines, generated as part of the ENCODE project [34,35].

Enrichment of candidate enhancers near cardiovascular-related genes

Enrichment of human heart candidate enhancers near heart-expressed genes was determined by comparing the frequency of human heart candidate enhancers with the frequency of

matched random regions near heart over- and under-expressed genes. Differentially expressed genes were identified using publicly available gene expression data from human fetal heart (GEO data set GSE1789, average expression across 5 ∼20wk normal fetal heart samples [36]), human adult heart (http://www.cardiogenomics.org [*retrieved October 2010*], average expression across 32 ischemic human heart samples) and human non-heart tissues (http://www.affymetrix.com/Auth/support/downloads/demo_data/HG-U133_Plus_2.tissue-mixture-data-set.apt-results.zip, average expression across ten non-heart tissues). Heart over- and under-expressed genes were identified by comparison with the non-heart tissue data set. Genes expressed at levels less than 100 in both heart and non-heart were excluded from further analysis. Heart over- and under-expressed genes were identified as the top 1000 genes ranked by the ratio of expression in heart versus non-heart and non-heart versus heart, respectively. To determine the expected frequency of candidate enhancers near heart over- and under-expressed genes, matched randomized data sets were generated by moving each candidate enhancer to a random location on the same chromosome, excluding regions that fail the peak filtering procedure described above. Enrichment of candidate heart enhancers was determined by comparison with the average frequency of randomized regions over 1,000 iterations.

Unsupervised enrichment analysis of gene annotation in the proximity of p300/CBP human fetal heart candidate enhancers was performed using Genomic Regions Enrichment of Annotations Tool (GREAT) [20]. Enrichment of p300/CBP candidate enhancers near RNA polymerase II proximal peaks was determined by comparing the frequency of human heart candidate enhancers with the frequency of matching random regions near proximal RNA polymerase II proximal (<2.5Kb away from nearest transcription start site) peaks (RNA polymerase II peaks were called using the same method as for p300/CBP ChIP-seq).

The genetest@NIH database (http://www.ncbi.nlm.nih.gov/sites/GeneTests/) was used to identify 73 genes known to be involved in heart disease: *CASQ2*, *GBA*, *LMNA*, *TNNT2*, *PSEN2*, *ACTA1*, *AGT*, *ACTN2*, *RYR2*, *SLC29A3*, *VCL*, *LDB3*, *ANKRD1*, *CSRP3*, *KCNQ1*, *ANO5*, *MYBPC3*, *MYL2*, *TBX5*, *ABCC9*, *PKP2*, *TMPO*, *SGCG*, *MYH6*, *MYH7*, *PSEN1*, *TGFB3*, *ACTC1*, *CAPN3*, *TPM1*, *TCAP*, *JUP*, *SGCA*, *ACE*, *DSC2*, *DSG2*, *TTR*, *FKRP*, *TNNI3*, *CFC1*, *TTN*, *DES*, *DYSF*, *JAG1*, *KCNE2*, *KCNE1*, *PPARG*, *TMEM43*, *AGTR1*, *DNAJC19*, *SCN5A*, *MYL3*, *TNNC1*, *CAV3*, *SGCB*, *MYOT*, *SGCD*, *NKX2-5*, *PLN*, *GJA1*, *EYA4*, *DSP*, *PRKAG2*, *GATA4*, *CHD7*, *TMEM70*, *FKTN*, *TRIM32*, *NOTCH1*, *LAMP2*, *FLNA*, *TAZ*, *DMD*. The human heart candidate enhancer data set was intersected with the regions flanking 50kb downstream and upstream heart disease-associated genes using Galaxy server [37,38].

The list of SNPs and their locations was retrieved from dbSNP build 130 using the table function of UCSC Genome Browser [33,39].

The list of functional SNPs affecting gene expression in primary human cells was retrieved from ref. [22] and 1637 eQTL SNPs for three cell types of 75 individuals [22] were intersected with 5,047 fetal human heart candidate enhancers.

Sequence constraint analyses

Evolutionary constraint of candidate enhancers was evaluated by comparison with conserved elements (phastCons [23]) identified from multiple alignments of vertebrate genome sequences to human or mouse. Candidate enhancers were assigned the score of the highest-scoring phastCons element overlapping either the whole peak or the 1kb genomic interval centered on the peak maximum (Supplementary Table 5).

Human-mouse alignments and conservation

To identify human-specific candidate heart enhancers, we aligned the human candidate enhancers to the mouse genome (mm9) using the liftOver tool [40]. We then analyzed the *in vivo* occupancy of p300/CBP in mouse and classified the human candidate enhancers as one of four classes: (class 1) human peaks mapped to the mouse genome, ≤2 kb apart from a mouse p300/CBP peak; (class 2) human peaks whose mouse homologous regions were not called as p300/CBP peaks but exceeded the genome-wide background (coverage threshold of 11 overlapping extended reads); (class 3) human peaks mapped to mouse regions with no p300/CBP binding; and (class 4) human peaks with no unique homologous locus in the mouse genome.

Transcription factor binding site analysis

Transcription factor binding site analysis of the top 500 acCBP/p300 peaks from human and mouse heart was performed as described previously [13,15]

Transgenic mouse enhancer assay

We selected regions for *in vivo* testing from the 5,047 fetal human candidate heart enhancer data set based on their Grizzly peak score. The high scoring 65 *in vivo* tested sequences also include three elements (hs1913, hs1948, and hs1959) that were initially called by the Grizzly Peak Fitting algorithm, but subsequently removed from the final genome-wide data set by an increased-stringency filtering process. Enhancer candidate regions consisting of ∼3.7kb (Supplementary Table 10) of human genomic DNA flanking the p300/CBP peak were amplified by PCR from human genomic DNA (Clontech) and cloned into the *Hsp68-promoter-LacZ* reporter vector as previously described [27,41]. Genomic coordinates of amplified regions are reported in Supplementary Table 12. Transgenic mouse embryos were generated as previously described [41]. Only patterns observed in at least three different embryos resulting from independent transgenic integration events of the same construct were considered reproducible [42]. For histological analysis, embryos were embedded in paraffin and sectioned using standard methods.

The activity of 8 selected elements was tested in adult heart (4 weeks of age). These elements were selected based on the reproducibility of LacZ cardiac staining at e11.5 and the peak score in the adult human heart candidate enhancer data set. The transgenic mouse enhancer assays were carried out as detailed above, except that F0 transgenic mice were sacrificed at 4 weeks of age. Hearts were X-Gal stained and sectioned using standard methods. All animal work was performed in accordance with protocols reviewed and

approved by the Lawrence Berkeley National Laboratory Animal Welfare and Research Committee.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors wish to thank Roya Hosseini and Sengthavy Phouanenavong for technical support, Samuel Deutsch for help in retrieving human genetic data, Catia Attanasio and Diane Dickel for critical comments on the manuscript. L.A.P. and A.V. were supported by grant HG003988 funded by National Human Genome Research Institute. B.L.B. was supported by grants HL64658 and HL89707 from the National Heart, Lung, and Blood Institute. D.M. and T.K. were supported by European Molecular Biology Organization (EMBO) long-term post-doctoral fellowships. B.C.J. was supported by GlaxoSmithKline Research and Education Foundation for Cardiovascular Disease, UCSF Foundation for Cardiac Research, and HL096836 from the National Heart, Lung, and Blood Institute. P.C.S was supported by the NHLBI and the Department of Veterans Affairs. Research was performed at Lawrence Berkeley National Laboratory and at the U.S. Department of Energy Joint Genome Institute under Department of Energy Contract DE-AC02-05CH11231, University of California.

References

1. Hoffman JI, Kaplan S, Liberthson RR. Prevalence of congenital heart disease. Am Heart J. 2004; 147:425–39. [PubMed: 14999190]
2. Lloyd-Jones D, et al. Heart disease and stroke statistics--2010 update: a report from the American Heart Association. Circulation. 2010; 121:e46–e215. [PubMed: 20019324]
3. Bentham J, Bhattacharya S. Genetic mechanisms controlling cardiovascular development. Ann N Y Acad Sci. 2008; 1123:10–9. [PubMed: 18375573]
4. Bruneau BG. The developmental genetics of congenital heart disease. Nature. 2008; 451:943–8. [PubMed: 18288184]
5. Pierpont ME, et al. Genetic basis for congenital heart defects: current knowledge: a scientific statement from the American Heart Association Congenital Cardiac Defects Committee, Council on Cardiovascular Disease in the Young: endorsed by the American Academy of Pediatrics. Circulation. 2007; 115:3015–38. [PubMed: 17519398]
6. Helgadottir A, et al. A common variant on chromosome 9p21 affects the risk of myocardial infarction. Science. 2007; 316:1491–3. [PubMed: 17478679]
7. McPherson R, et al. A common allele on chromosome 9 associated with coronary heart disease. Science. 2007; 316:1488–91. [PubMed: 17478681]
8. Schunkert H, et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. Nat Genet. 2011
9. Visel A, Rubin EM, Pennacchio LA. Genomic views of distant-acting enhancers. Nature. 2009; 461:199–205. [PubMed: 19741700]
10. Visel A, et al. Targeted deletion of the 9p21 non-coding coronary artery disease risk interval in mice. Nature. 2010; 464:409–12. [PubMed: 20173736]
11. Chen X, et al. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell. 2008; 133:1106–17. [PubMed: 18555785]
12. Heintzman ND, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet. 2007; 39:311–8. [PubMed: 17277777]
13. Visel A, et al. ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature. 2009; 457:854–8. [PubMed: 19212405]
14. Xi H, et al. Identification and characterization of cell type-specific and ubiquitous chromatin regulatory structures in the human genome. PLoS Genet. 2007; 3:e136. [PubMed: 17708682]
15. Blow MJ, et al. ChIP-Seq identification of weakly conserved heart enhancers. Nat Genet. 2010; 42:806–810. [PubMed: 20729851]

16. Arany Z, Sellers WR, Livingston DM, Eckner R. E1A-associated p300 and CREB-associated CBP belong to a conserved family of coactivators. Cell. 1994; 77:799–800. [PubMed: 8004670]
17. Eckner R, et al. Molecular cloning and functional analysis of the adenovirus E1A-associated 300-kD protein (p300) reveals a protein with properties of a transcriptional adaptor. Genes Dev. 1994; 8:869–84. [PubMed: 7523245]
18. Ogryzko VV, Schiltz RL, Russanova V, Howard BH, Nakatani Y. The transcriptional coactivators p300 and CBP are histone acetyltransferases. Cell. 1996; 87:953–9. [PubMed: 8945521]
19. Capaldi AP, et al. Structure and function of a transcriptional network activated by the MAPK Hog1. Nat Genet. 2008; 40:1300–6. [PubMed: 18931682]
20. McLean CY, et al. GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol. 2010; 28:495–501. [PubMed: 20436461]
21. Ashburner M, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000; 25:25–9. [PubMed: 10802651]
22. Dimas AS, et al. Common regulatory variation impacts gene expression in a cell type-dependent manner. Science. 2009; 325:1246–50. [PubMed: 19644074]
23. Siepel A, et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res. 2005; 15:1034–50. [PubMed: 16024819]
24. Narlikar L, et al. Genome-wide discovery of human heart enhancers. Genome Res. 2010; 20:381–92. [PubMed: 20075146]
25. Henderson DJ, Anderson RH. The development and structure of the ventricles in the human heart. Pediatr Cardiol. 2009; 30:588–96. [PubMed: 19225828]
26. Coppen SR, et al. Comparison of connexin expression patterns in the developing mouse heart and human foetal heart. Mol Cell Biochem. 2003; 242:121–7. [PubMed: 12619874]
27. Kothary R, et al. A transgene containing lacZ inserted into the dystonia locus is expressed in neural tube. Nature. 1988; 335:435–7. [PubMed: 3138544]
28. Pennacchio LA, et al. In vivo enhancer analysis of human conserved non-coding sequences. Nature. 2006; 444:499–502. [PubMed: 17086198]
29. Blake JA, Bult CJ, Eppig JT, Kadin JA, Richardson JE. The Mouse Genome Database genotypes: :phenotypes. Nucleic Acids Res. 2009; 37:D712–9. [PubMed: 18981050]
30. Morimoto T, Sunagawa Y, Fujita M, Hasegawa K. Novel heart failure therapy targeting transcriptional pathway in cardiomyocytes by a natural compound, curcumin. Circ J. 2010; 74:1059–66. [PubMed: 20467147]
31. Thompson PR, et al. Regulation of the p300 HAT domain via a novel activation loop. Nat Struct Mol Biol. 2004; 11:308–15. [PubMed: 15004546]
32. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009; 25:1754–60. [PubMed: 19451168]
33. Kuhn RM, et al. The UCSC Genome Browser Database: update 2009. Nucleic Acids Res. 2009; 37:D755–61. [PubMed: 18996895]
34. Birney E, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature. 2007; 447:799–816. [PubMed: 17571346]
35. Ernst J, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature. 2011; 473:43–9. [PubMed: 21441907]
36. Conti A, et al. Altered expression of mitochondrial and extracellular matrix genes in the heart of human fetuses with chromosome 21 trisomy. BMC Genomics. 2007; 8:268. [PubMed: 17683628]
37. Blankenberg D, et al. Galaxy: a web-based genome analysis tool for experimentalists. Curr Protoc Mol Biol. 2010; Chapter 19(Unit 19 10):1–21.
38. Goecks J, Nekrutenko A, Taylor J. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol. 2010; 11:R86. [PubMed: 20738864]
39. Sherry ST, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 2001; 29:308–11. [PubMed: 11125122]
40. Hinrichs AS, et al. The UCSC Genome Browser Database: update 2006. Nucleic Acids Res. 2006; 34:D590–8. [PubMed: 16381938]

41. Nobrega MA, Ovcharenko I, Afzal V, Rubin EM. Scanning human gene deserts for long-range enhancers. Science. 2003; 302:413. [PubMed: 14563999]

42. Visel A, Minovitsky S, Dubchak I, Pennacchio LA. VISTA Enhancer Browser--a database of tissue-specific human enhancers. Nucleic Acids Res. 2007; 35:D88–92. [PubMed: 17130149]

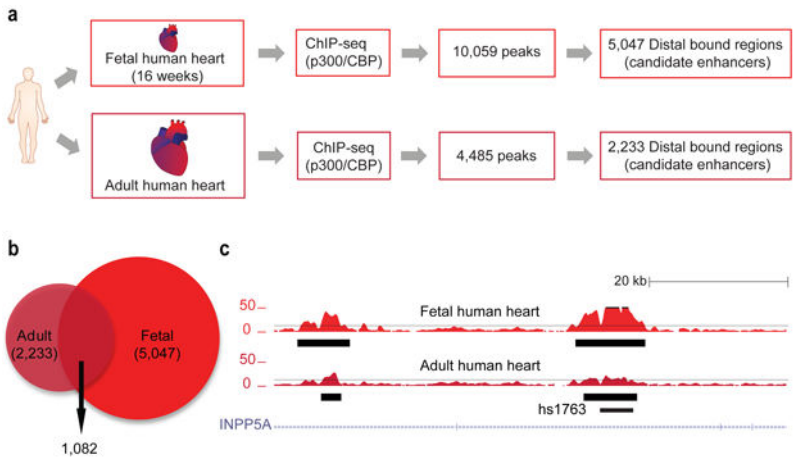

Figure 1. ChIP-seq identification of candidate enhancer regions from human fetal and adult heart

Human fetal heart was obtained at gestational week 16, adult heart tissue was obtained from the septum of an adult failing heart. a) Overview of strategy and results of ChIP-seq analysis. 5,047 regions from fetal heart and 2,233 from adult heart were significantly enriched in p300/CBP binding and considered as candidate human heart enhancers (distal: >2.5kb from the nearest transcript start site; peaks <2.5kb from the nearest transcript start site were considered proximal/promoter-associated). b) Overlap of candidate enhancers identified in both fetal and adult heart tissues. c) ChIP-seq profiles of p300/CBP in the genomic region of the tested element hs1763 (thin black bar). Thick black bars indicate two regions of significant enrichment in p300/CBP binding, in introns of the INPP5A gene. Thin grey line represents read depth of 10, maximum read depth shown is 50.

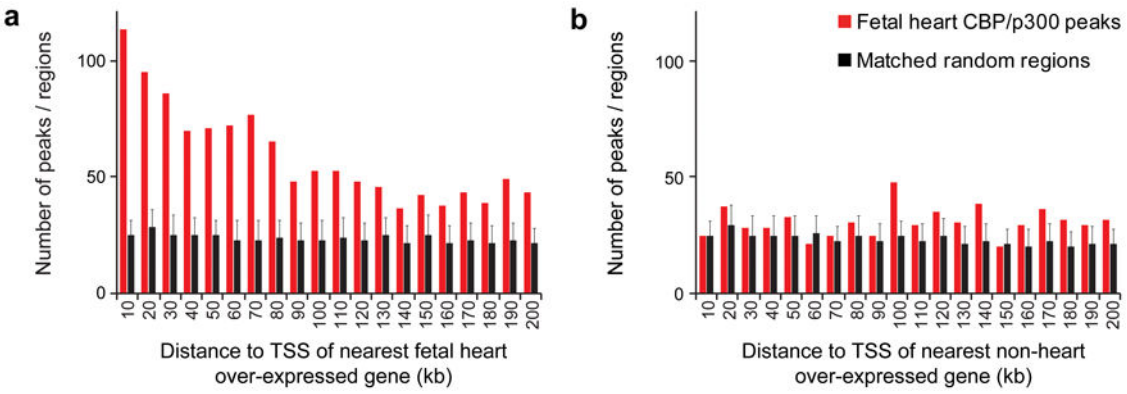

Figure 2. Human p300/CBP candidate enhancers are enriched near human heart-expressed genes

Frequency of human fetal heart candidate enhancers (red) compared to matched random regions (black), near genes that are (a) over-expressed and (b) under-expressed in fetal heart relative to other human tissues (see **Methods**). Errors bars indicate 95% confidence intervals.

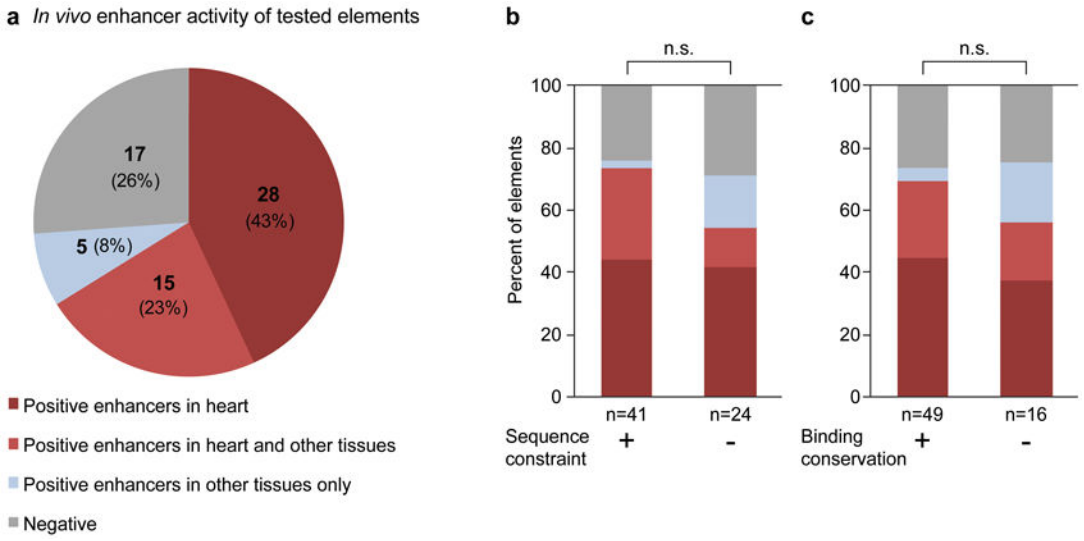

Figure 3. *In vivo* testing of predicted human heart enhancer activities in transgenic mice

a) *In vivo* enhancer activity of the 65 tested elements. b) Proportion of reproducible enhancers by extent of sequence constraint (+, phastCons > 350; -, phastCons ≤ 350). c) Proportion of reproducible enhancers by binding conservation to the mouse (+, p300/CBP binding significant or sub-significant but above background; -, p300/CBP binding at no more than genome-wide background levels or non-alignable peaks). n.s., not significant (pairwise comparison for each subcategory, two-tailed Fisher's Exact test, $P > 0.05$ in all cases).

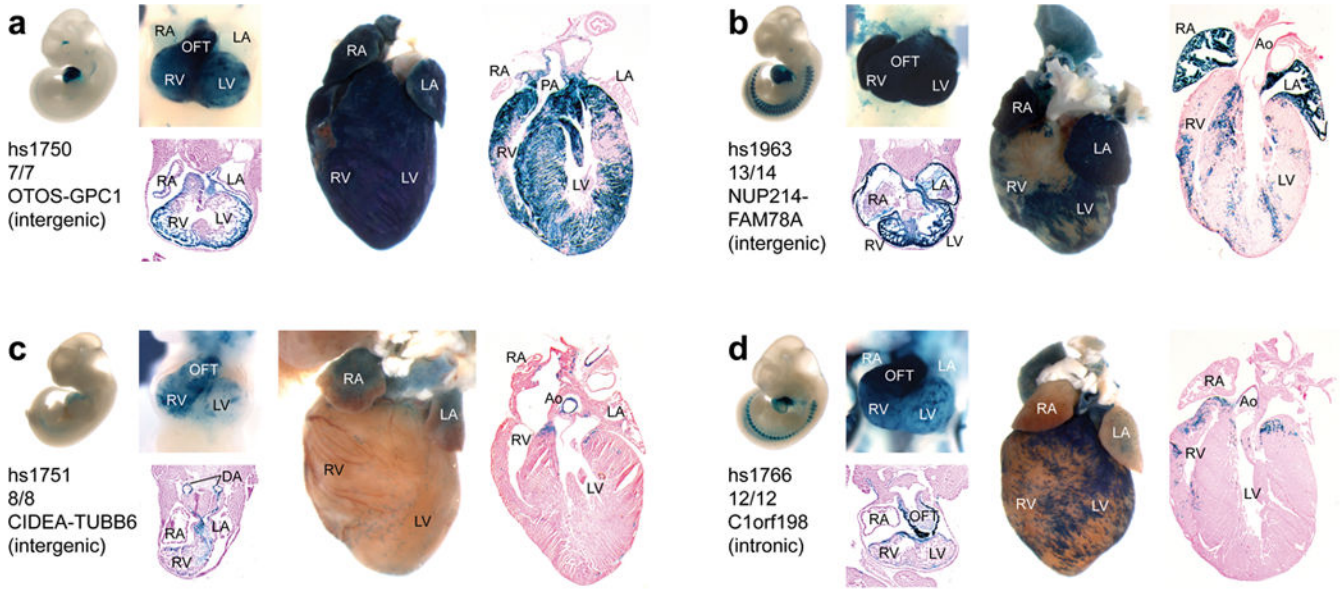

Figure 4. Activity of human cardiac enhancers in embryonic and adult transgenic mice

Examples of *in vivo* enhancer activities. In each panel from left to right: Whole-mount stained e11.5 embryo, close-up and histological section of heart at e11.5, whole-mount stained heart at postnatal day 28, longitudinal section of heart at postnatal day 28. All specimens were stained for LacZ enhancer reporter activity (dark blue). In each panel, element ID, reproducibility in e11.5 embryos and flanking genes are indicated. LV-left ventricle; RV-right ventricle; LA-left atrium; RA-right atrium; OFT-outflow tract; PA-pulmonary artery; Ao-Aorta.

Table 1

Top enriched annotations of putative target genes near the candidate human heart enhancers

Unsupervised enrichment analysis [20] of annotated genes in the proximity of p300/CBP distal regions. a) Top 10 enriched Mouse Genome Informatics phenotype ontology terms [29], showing highly significant enrichment of genes implicated in cardiovascular-system related phenotypes. b) Top 5 terms in the biological process gene ontology (GO), showing highly significant enrichment of cardiovascular-related terms. Only terms that showed significant enrichment and had a binomial fold-enrichment of at least 2 were considered.

Top Enriched Phenotypes	Binomial FDR Q-Val	Binomial Fold Enrichment
1. abnormal cardiac muscle morphology	3.32E-44	2.1
2. abnormal vitelline vasculature	1.13E-31	2.2
3. abnormal epicardium morphology	5.23E-27	5.7
4. thin ventricular wall	3.43E-22	2.6
5. abnormal myocardial trabeculae morphology	7.74E-22	2.3
6. decreased cardiac muscle contractility	2.79E-21	2.3
7. abnormal renal glomerulus morphology	3.34E-21	2.1
8. anemia	1.14E-20	2.0
9. pericardial effusion	7.44E-20	2.9
10. abnormal cardiac muscle contractility	4.35E-19	2.1

Top Enriched GO Terms	Binomial FDR Q-Val	Binomial Fold Enrichment
1. blood vessel morphogenesis	2.37E-30	2.2
2. heart looping	5.19E-28	7.5
3. blood vessel development	7.73E-27	2.0
4. vasculature development	2.30E-26	2.0
5. regulation of heart contraction	3.76E-25	2.9