nature genetics

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[6,8], 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] & Axel Visel[1,2]

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 noncoding sequences highly enriched in human heart enhancers that is likely to facilitate downstream 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 is strongly influenced by genetic factors[1–5]. Genome-wide association studies indicate that variation in noncoding sequences, including distant-acting transcriptional enhancers, affects susceptibility to many types of human disease[6–10]. However, the possible role of enhancers in heart disease has been difficult to evaluate because of the lack of a catalog of human cardiac enhancers. Mapping of enhancer-associated epigenomic marks via chromatin immunoprecipitation coupled with 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 the enhancer-associated coactivator 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 tended 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 profiles of two enhancer-associated coactivator 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 coactivator 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) throughout the genome that were located at least 2.5 kb from the nearest transcription start site (TSS; **Fig. 1a–c**, **Supplementary Fig. 1**, **Supplementary Table 1** and Online Methods). In an equivalent analysis, we identified 2,233 regions from adult human heart. Nearly half of the adult human heart enhancer candidates (1,082; 48%) coincided with candidate enhancers found in fetal human heart. In addition, many peaks identified in one of the samples exhibited read densities above background but below the peak significance threshold in the other sample. In total, 4,257 of fetal peaks (84%) and 2,113 of adult peaks (95%) showed significantly or subsignificantly increased read densities in the adult or fetal data sets, 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** and **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 whether cardiac enhancers were localized in this manner, we examined the cardiac expression and function of genes located near the regions identified by ChIP-Seq. 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 cardiac p300/CBP-binding peaks within 2.5–10 kb of the TSSs of genes highly expressed in fetal human heart ($P < 1 \times 10^{-40}$, binomial

[1]Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA. [2]United States Department of Energy Joint Genome Institute, Walnut Creek, California, USA. [3]Department of Molecular and Cell Biology, California Institute of Quantitative Biosciences, University of California, Berkeley, Berkeley, California, USA. [4]School of Computer Science and Engineering, The Hebrew University, Jerusalem, Israel. [5]Cardiovascular Research Institute, University of California, San Francisco, San Francisco, California, USA. [6]Division of Cardiology, University of California, San Francisco, San Francisco, California, USA. [7]Cardiology Division, Virginia Medical Center, San Francisco, California, USA. [8]Current address: Division of Cardiology, University of North Carolina, Chapel Hill, North Carolina, USA. Correspondence should be addressed to L.A.P. (lapennacchio@lbl.gov) or A.V. (avisel@lbl.gov).

Received 13 April; accepted 20 October; published online 4 December 2011; doi:10.1038/ng.1006

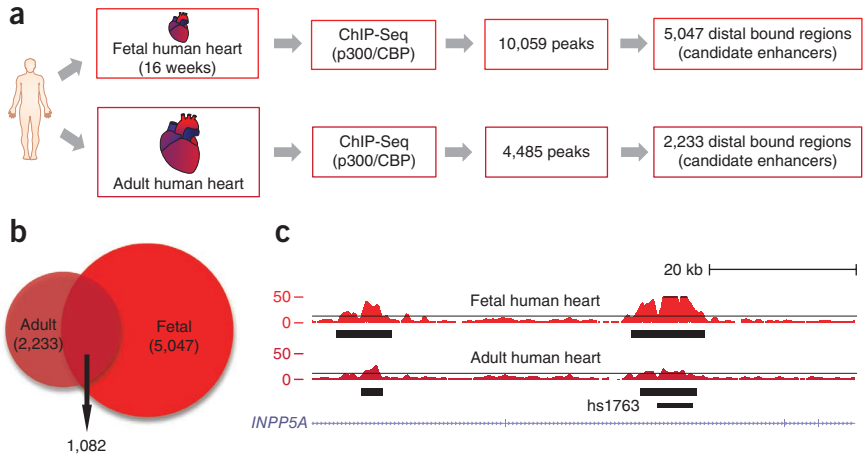

Figure 1 ChIP-Seq identification of candidate enhancer regions from human fetal and adult heart. Human fetal heart was obtained at gestational week 16, and adult heart tissue was obtained from the septum of an adult failing heart. (**a**) Overview of strategy and results of ChIP-Seq analysis. In total, 5,047 regions from fetal heart and 2,233 from adult heart were significantly enriched in p300/CBP-binding sites and were considered as candidate human heart enhancers (distal: ≥2.5 kb from the nearest transcript start site; proximal or promoter associated: <2.5 kb from the nearest TSS). (**b**) Overlap of candidate enhancers identified in fetal and adult heart tissues. (**c**) ChIP-Seq profiles of p300/CBP in the genomic region of the tested hs1763 element (thin black bar). Thick black bars indicate two regions significantly enriched for p300/CBP binding in introns of the *INPP5A* gene. The thin black line represents a read depth of 10; maximum read depth shown is 50.

distribution), with significant enrichment up to 200 kb away from promoters (cumulative fold enrichment of 2.6; $P < 0.001$, binomial distribution; **Fig. 2a**). We also observed enrichment for human adult candidate enhancers when the binding of RNA polymerase II 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 whether candidate cardiac enhancers were enriched near genes with known cardiac functions, we performed an unsupervised statistical enrichment analysis of functional gene annotations[20]. Candidate cardiac enhancers were indeed associated with genes that have been linked to cardiovascular functions in mouse deletion studies and by Gene Ontology annotation[21] (**Table 1** and **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 the idea that they function as human heart enhancers.

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 the genetic diagnosis of heart diseases and found that candidate heart enhancers were significantly enriched near these genes ($P < 0.05$, binomial distribution; **Supplementary Fig. 5**). We identified 81 human fetal heart candidate enhancers that are located within 50 kb of 30 of these cardiac genes, an enrichment of 4.5-fold over what would be expected by chance ($P < 7 \times 10^{-15}$, hypergeometric distribution; see Online 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 SNPs within these 81 candidate enhancers (**Supplementary Table 3**). In addition, we identified genetic variants in 11 candidate enhancers that have been linked to changes in gene expression in human cells[22] (**Supplementary Table 4**). Although in-depth studies will be required to clarify whether 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 how noncoding enhancer variants contribute 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 identified in this study (86%) 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 re-examined the usefulness of mouse-derived ChIP-Seq data sets for the accurate annotation of human heart enhancers. We performed ChIP-Seq for p300/CBP-binding sites in mouse heart tissue at postnatal day 2, when the developmental progression and gene expression profile of the mouse heart is broadly similar to that of human fetal heart at gestational week 16 (refs. 25,26). Using identical methods as for human heart tissue, we identified 6,564 candidate enhancers (distal ChIP-Seq peaks; **Supplementary Table 1**)

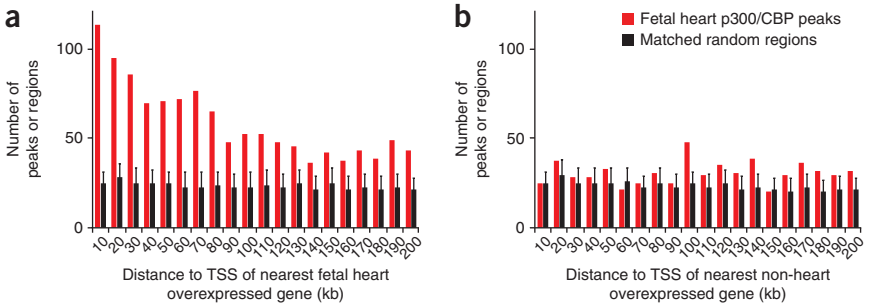

Figure 2 Human p300/CBP candidate enhancers are enriched near genes expressed in human heart. (**a**,**b**) Frequency of human fetal heart candidate enhancers (red) compared to matched random regions (black) near genes that are overexpressed (**a**) or underexpressed (**b**) in fetal heart relative to other human tissues (see Online Methods). Error bars indicate 95% confidence intervals.

Table 1 Top enriched annotations of putative target genes near the candidate human heart enhancers

Top enriched phenotypes	Binomial FDR *Q* value	Binomial fold enrichment
1. Abnormal cardiac muscle morphology	3.32×10^{-44}	2.1
2. Abnormal vitelline vasculature	1.13×10^{-31}	2.2
3. Abnormal epicardium morphology	5.23×10^{-27}	5.7
4. Thin ventricular wall	3.43×10^{-22}	2.6
5. Abnormal myocardial trabeculae morphology	7.74×10^{-22}	2.3
6. Decreased cardiac muscle contractility	2.79×10^{-21}	2.3
7. Abnormal renal glomerulus morphology[a]	3.34×10^{-21}	2.1
8. Anemia[a]	1.14×10^{-20}	2.0
9. Pericardial effusion	7.44×10^{-20}	2.9
10. Abnormal cardiac muscle contractility	4.35×10^{-19}	2.1
Top enriched GO terms	**Binomial FDR *Q* value**	**Binomial fold enrichment**
1. Blood vessel morphogenesis	2.37×10^{-30}	2.2
2. Heart looping	5.19×10^{-28}	7.5
3. Blood vessel development	7.73×10^{-27}	2.0
4. Vasculature development	2.30×10^{-26}	2.0
5. Regulation of heart contraction	3.76×10^{-25}	2.9

Unsupervised enrichment analysis[20] of annotated genes in the proximity of p300/CBP-binding distal regions. Top, the top ten enriched Mouse Genome Informatics phenotype ontology terms[29] showing highly significant enrichment of genes implicated in cardiovascular system–related phenotypes. Bottom, the top five terms in biological process Gene Ontology (GO) showing highly significant enrichment of cardiovascular system–related terms. Only terms that showed significant enrichment and had a binomial fold enrichment of ≥2 were considered.
[a]The only non-cardiovascular terms retrieved by this analysis.

and compared them to the human fetal candidate enhancer sets. We found that only 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 Note**, **Supplementary Fig. 7** and Online Methods). Although 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 human heart enhancers predicted by ChIP-Seq, we used a transgenic mouse enhancer assay that was previously shown to be an effective approach to define the patterns of *in vivo* activity 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). Candidate enhancers were selected without knowledge of the nearby genes, focusing on sequences with high-confidence binding to p300/CBP 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 tested sequences (66%) drove reproducible expression in the heart or in the vasculature, either exclusively in these tissues ($N = 28$; 43%) or as part of a reproducible compound pattern that included the heart ($N = 15$; 23%) (**Fig. 3**, selected examples in **Fig. 4** and **Supplementary Table 7**). The *in vivo* validation rate of the tested human sequences was significantly higher than the frequency of *in vivo* heart enhancers identified on the basis of conserved sequences near heart-expressed genes or by predictions derived from human cell lines ($P < 0.001$, χ^2 test; **Supplementary Figs. 8** and **9** and **Supplementary Table 8**) 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**). Although the tested human sequences in the present study were on average longer than the previously tested mouse sequences (3,719 bp versus 1,648 bp; **Supplementary Table 9**), the predicted human candidate heart enhancers had overall robust cardiac activity, and potential human-mouse species incompatibilities had only minor effects on the reproducibility of reporter patterns (**Supplementary Table 10**). Of note, we did not observe significant differences in the reproducible activity of the tested elements between the different classes of conservation (**Fig. 3b**; $P = 0.9$, χ^2 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 Fig. 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 noncoding 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 active enhancers in the adult human heart in transgenic mice at postnatal day 28 (P28). In total, 41 of the 65 candidate enhancers tested at E11.5 (63%) 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 4-week-old transgenic mice. All eight enhancers showed reproducible activity in hearts from these animals, regardless of their sequence and binding conservation (**Supplementary Table 11**). To explore the spatial activity of these

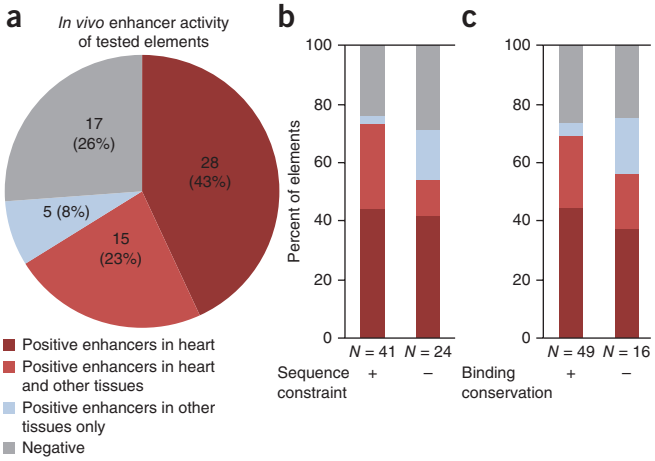

Figure 3 *In vivo* testing of predicted human heart enhancer activities in transgenic mice. (**a**) *In vivo* enhancer activity of the 65 tested elements. (**b**) Proportion of reproducible enhancers by extent of sequence constraint (+, phastCons > 350; –, phastCons ≤ 350). (**c**) Proportion of reproducible enhancers by binding conservation to the mouse (+, p300/CBP binding significant or subsignificant but above background; –, p300/CBP binding not above background or in non-alignable peaks). Pairwise comparison for each subcategory was calculated with two-tailed Fisher's exact test; $P > 0.05$ in all cases.

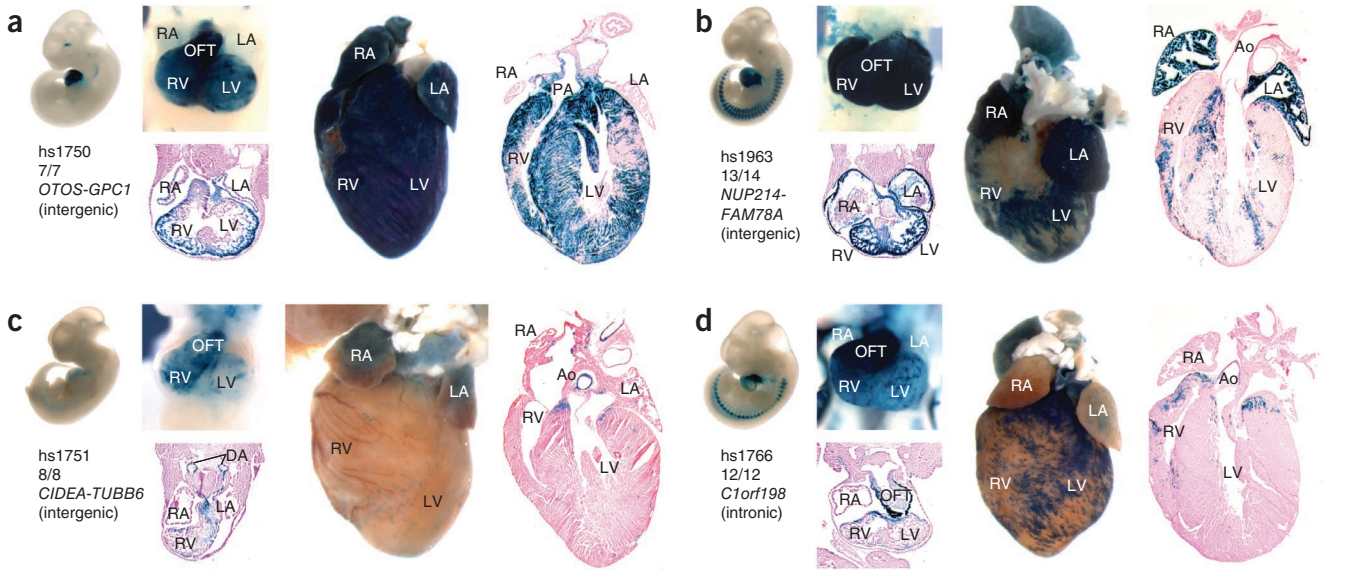

Figure 4 *In vivo* activity of human cardiac enhancers in embryonic and 4-week-old transgenic mice. (**a**–**d**) From left to right: whole-mount stained E11.5 embryo, close-up and histological section of heart at E11.5, whole-mount stained heart at P28 and longitudinal section of heart at P28. All specimens were stained for *LacZ* enhancer reporter activity (dark blue). 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.

enhancers, we examined transverse sections from representative whole-mount embryos and longitudinal sections from P28 hearts, annotating reproducible reporter staining patterns. Although different enhancers drive expression in distinct subregions of the E11.5 and P28 heart, we observed a remarkable overall concordance between the regions of enhancer activity in the embryonic and P28 heart for seven of the eight sequences examined (**Fig. 4** and **Supplementary Fig. 12**). These results further support the idea that the ChIP-Seq data sets obtained in this study correctly predict the genomic location of human noncoding sequences that are active *in vivo* enhancers in the developing and adult heart.

Deciphering the gene-regulatory architecture required for development and function of the human heart represents a pressing challenge. In the present study, we 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 showed that the majority of these sequences are authentic *in vivo* heart enhancers. Notably, comparison of these human-derived epigenomic data sets with ChIP-Seq data from mouse heart tissue showed that many heart enhancers are neither evolutionarily nor functionally conserved between the human and the mouse. In our studies, even human heart enhancers with sequences not conserved in the mouse had 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.

URLs. VISTA Enhancer Browser, http://enhancer.lbl.gov/; Grizzly Peak fitting algorithm, http://eisenlab.org/software/grizzly/; CardioGenomics (retrieved October 2010), http://cardiogenomics.med.harvard.edu/home; gene expression in human non-heart tissue, http://www.affymetrix.com/Auth/support/downloads/demo_data/HG-U133_Plus_2.tissue-mixture-data-set.apt-results.zip; GeneTests database, http://www.ncbi.nlm.nih.gov/sites/GeneTests/.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Accession numbers. All raw sequences and processed data from ChIP-Seq experiments have been deposited in the NCBI GEO database (GSE32587). Complete *in vivo* data sets are available from the VISTA Enhancer Browser under the accession numbers provided in **Supplementary Tables 7, 8** and **10**.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

The authors thank R. Hosseini and S. Phouanenavong for technical support, S. Deutsch for help in retrieving human genetic data and C. Attanasio and D. Dickel for critical comments on the manuscript. L.A.P. and A.V. were supported by a grant funded by the National Human Genome Research Institute (HG003988). B.L.B. was supported by grants from the National Heart, Lung, and Blood Institute (NHLBI, HL64658 and HL89707). D.M. and T.K. were supported by European Molecular Biology Organization (EMBO) long-term postdoctoral fellowships. B.C.J. was supported by the GlaxoSmithKline Research and Education Foundation for Cardiovascular Disease, University of California, San Francisco (UCSF) Foundation for Cardiac Research and a grant from the NHLBI (HL096836). 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 United States Department of Energy Joint Genome Institute (Department of Energy Contract DE-AC02-05CH11231, University of California).

AUTHOR CONTRIBUTIONS

D.M., E.M.R., J.B., L.A.P. and A.V. conceived of 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 and materials and performed data analysis. All authors contributed to the writing of the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturegenetics/.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Hoffman, J.I., Kaplan, S. & Liberthson, R.R. Prevalence of congenital heart disease. *Am. Heart J.* **147**, 425–439 (2004).
2. Lloyd-Jones, D. *et al.* Heart disease and stroke statistics—2010 update: a report from the American Heart Association. *Circulation* **121**, e46–e215 (2010).
3. Bentham, J. & Bhattacharya, S. Genetic mechanisms controlling cardiovascular development. *Ann. NY Acad. Sci.* **1123**, 10–19 (2008).
4. Bruneau, B.G. The developmental genetics of congenital heart disease. *Nature* **451**, 943–948 (2008).
5. Pierpont, M.E. *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* **115**, 3015–3038 (2007).
6. Helgadottir, A. *et al.* A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science* **316**, 1491–1493 (2007).
7. McPherson, R. *et al.* A common allele on chromosome 9 associated with coronary heart disease. *Science* **316**, 1488–1491 (2007).
8. Schunkert, H. *et al.* Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat. Genet.* **43**, 333–338 (2011).
9. Visel, A., Rubin, E.M. & Pennacchio, L.A. Genomic views of distant-acting enhancers. *Nature* **461**, 199–205 (2009).
10. Visel, A. *et al.* Targeted deletion of the 9p21 non-coding coronary artery disease risk interval in mice. *Nature* **464**, 409–412 (2010).
11. Chen, X. *et al.* Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**, 1106–1117 (2008).
12. Heintzman, N.D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* **39**, 311–318 (2007).
13. Visel, A. *et al.* ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**, 854–858 (2009).
14. Xi, H. *et al.* Identification and characterization of cell type–specific and ubiquitous chromatin regulatory structures in the human genome. *PLoS Genet.* **3**, e136 (2007).
15. Blow, M.J. *et al.* ChIP-Seq identification of weakly conserved heart enhancers. *Nat. Genet.* **42**, 806–810 (2010).
16. Arany, Z., Sellers, W.R., Livingston, D.M. & Eckner, R. E1A-associated p300 and CREB-associated CBP belong to a conserved family of coactivators. *Cell* **77**, 799–800 (1994).
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.* **8**, 869–884 (1994).
18. Ogryzko, V.V., Schiltz, R.L., Russanova, V., Howard, B.H. & Nakatani, Y. The transcriptional coactivators p300 and CBP are histone acetyltransferases. *Cell* **87**, 953–959 (1996).
19. Capaldi, A.P. *et al.* Structure and function of a transcriptional network activated by the MAPK Hog1. *Nat. Genet.* **40**, 1300–1306 (2008).
20. McLean, C.Y. *et al.* GREAT improves functional interpretation of *cis*-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).
21. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).
22. Dimas, A.S. *et al.* Common regulatory variation impacts gene expression in a cell type–dependent manner. *Science* **325**, 1246–1250 (2009).
23. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005).
24. Narlikar, L. *et al.* Genome-wide discovery of human heart enhancers. *Genome Res.* **20**, 381–392 (2010).
25. Henderson, D.J. & Anderson, R.H. The development and structure of the ventricles in the human heart. *Pediatr. Cardiol.* **30**, 588–596 (2009).
26. Coppen, S.R. *et al.* Comparison of connexin expression patterns in the developing mouse heart and human foetal heart. *Mol. Cell. Biochem.* **242**, 121–127 (2003).
27. Kothary, R. *et al.* A transgene containing *lacZ* inserted into the *dystonia* locus is expressed in neural tube. *Nature* **335**, 435–437 (1988).
28. Pennacchio, L.A. *et al. In vivo* enhancer analysis of human conserved non-coding sequences. *Nature* **444**, 499–502 (2006).
29. Blake, J.A., Bult, C.J., Eppig, J.T., Kadin, J.A. & Richardson, J.E. The Mouse Genome Database genotypes:phenotypes. *Nucleic Acids Res.* **37**, D712–D719 (2009).

ONLINE METHODS

Human and mouse heart tissue collection and preparation. Adult human heart tissue (ischemic, failing; male; age 45 years; ejection fraction 31%) was obtained from a heart removed at the time of transplant at UCSF with the approval of the UCSF Committee for Human Research. Full informed consent was obtained from the transplant recipient before surgery. Cold cardioplegic solution was perfused antegrade before 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 human heart tissue from gestational week 16 was obtained from ABR, Inc. in compliance with applicable state and federal laws and with full informed consent. P2 mouse heart tissue was isolated and pooled from approximately 15 CD-1 mice. All procedures of this study involving human tissue samples were reviewed and approved by the Human Subjects Committee at Lawrence Berkeley National Laboratory. All animal work was performed in accordance with protocols reviewed and approved by the Lawrence Berkeley National Laboratory Animal Welfare and Research Committee.

Tissue samples were processed for ChIP and DNA sequencing as described previously[13,15], with a few modifications. After 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 µl of anti–acetyl-CBP/p300 antibody (rabbit polyclonal no. 4771, Cell Signaling Technology) that recognizes these proteins in their active, acetylated state[30,31] or 5 µg of anti–RNA polymerase II antibody (mouse monoclonal ab817, Abcam).

Processing of ChIP-Seq data. Sequence reads (36 bp) were aligned to the reference genomes (NCBI build 36, hg18 and NCBI build 37, mm9) using the Burrows-Wheeler Aligner (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 p300/CBP-bound regions were identified from the intersection of peaks identified by MACS[11] (version 1.4, with default settings except: bw = 300; $P = 1 \times 10^{-5}$; mfold = 10,30; local = 20,000; off-auto–shift size = 100) and a modified Grizzly Peak fitting algorithm (see URLs)[19]. Grizzly peaks with a separating gap of less than the length of either of the two peaks or less than 1 kb were merged into single regions. Peaks were scored and sorted on the basis of maximal occupancy, as called by Grizzly. Bound regions mapped to unassembled chromosomal contigs, centromeric regions, telomeric regions, segmental duplications or regions where >50% of the contributing reads originated in repeats were removed from further analyses as likely artifacts.

p300/CBP-bound regions were classified according to their proximity to the TSSs of known genes in the UCSC database[33]. Regions within 2.5 kb of the nearest TSS were defined as promoter peaks. The remaining peaks represent candidate distant-acting cardiac enhancers.

To assess the reproducibility of p300/CBP ChIP-Seq data, 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 nine human cell lines that were generated as part of the ENCODE project[34,35].

Enrichment of candidate enhancers near cardiovascular system–related genes. Enrichment of candidate human cardiac enhancers near genes expressed in the heart was determined by comparing the frequency of candidate human heart enhancers with the frequency of matched random regions near genes overexpressed or suppressed in the heart. Differentially expressed genes were identified using publicly available gene expression data from human fetal heart (GEO data set GSE1789, average expression across five normal fetal heart samples at ~20 weeks[36]), human adult heart (average expression across 32 ischemic human heart samples; CardioGenomics) and human non-heart tissues (average expression across ten non-heart tissues; see URLs). Genes that are differentially expressed in the heart were identified by comparison with the non-heart tissue data set. Genes with normalized MAS5.0 expression values of less than 100 in both heart and non-heart were excluded from further analysis. Genes overexpressed and suppressed in the heart were identified as the top 1,000 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 genes differentially expressed in the heart, matched randomized data sets were generated by moving each candidate enhancer to a random location on the same chromosome, excluding regions that failed 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 candidate p300/CBP human fetal heart enhancers was performed using the genomic regions enrichment of annotations tool (GREAT)[20]. Enrichment of candidate p300/CBP enhancers near RNA polymerase II proximal peaks was determined by comparing the frequency of candidate human heart enhancers with the frequency of matching random regions near RNA polymerase II proximal peaks (<2.5 kb away from nearest TSS) (RNA polymerase II peaks were called using the same method as for p300/CBP ChIP-Seq).

The GeneTests database (see URLs) 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* and *DMD*. The data set of candidate human heart enhancers was intersected with the flanking regions 50 kb downstream and upstream of heart disease–associated genes using the Galaxy server[37,38].

The list of SNPs and their locations was retrieved from dbSNP build 130 using the table function of the UCSC Genome Browser[33,39]. The list of functional SNPs affecting gene expression in primary human cells was retrieved[22], and 1,637 expression quantitative trait loci (eQTL) SNPs for three cell types of 75 individuals[22] were intersected with the 5,047 candidate fetal human heart 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 genomic sequences. Candidate enhancers were assigned the score of the highest-scoring phastCons element overlapping either the whole peak or the 1-kb genomic interval centered around 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 the mouse and placed the human candidate enhancers in one of four classes: class 1, human peaks mapped to the mouse genome ≤2 kb from a mouse p300/CBP peak; class 2, human peaks whose homologous regions in the mouse were not called as p300/CBP peaks but exceeded genome-wide background levels (subsignificant enrichment, 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 acetylated p300/CBP peaks from human and mouse heart tissues was performed as described[13,15].

Transgenic mouse enhancer assay. We selected regions for *in vivo* testing from the data set of 5,047 candidate fetal human heart enhancers on the basis of their Grizzly peak score. The 65 high-scoring *in vivo*–tested sequences also include three elements (hs1913, hs1948 and hs1959) that were initially called by the Grizzly Peak fitting algorithm but were subsequently removed from the final

doi:10.1038/ng.1006

genome-wide data set by an increased-stringency filtering process. Enhancer candidate regions consisting of ~3.7 kb of human genomic DNA flanking the p300/CBP peaks (**Supplementary Table 10**) were amplified by PCR (Clontech) from human genomic DNA and cloned into the *Hsp68-promoter-LacZ* reporter vector as described[27,41]. Genomic coordinates of amplified regions are reported in **Supplementary Table 12**. Transgenic mouse embryos were generated as 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 eight selected elements was tested at 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 candidate adult human heart 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 stained with X-Gal and sectioned using standard methods.

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.* **74**, 1059–1066 (2010).
31. Thompson, P.R. *et al.* Regulation of the p300 HAT domain via a novel activation loop. *Nat. Struct. Mol. Biol.* **11**, 308–315 (2004).
32. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
33. Kuhn, R.M. *et al.* The UCSC Genome Browser Database: update 2009. *Nucleic Acids Res.* **37**, D755–D761 (2009).
34. Birney, E. *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816 (2007).
35. Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43–49 (2011).
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* **8**, 268 (2007).
37. Blankenberg, D. *et al.* Galaxy: a web-based genome analysis tool for experimentalists. *Curr. Protoc. Mol. Biol.* Chapter 19, Unit 19.10.1–19.10.21 (Wiley, 2010).
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.* **11**, R86 (2010).
39. Sherry, S.T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29**, 308–311 (2001).
40. Hinrichs, A.S. *et al.* The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res.* **34**, D590–D598 (2006).
41. Nobrega, M.A., Ovcharenko, I., Afzal, V. & Rubin, E.M. Scanning human gene deserts for long-range enhancers. *Science* **302**, 413 (2003).
42. Visel, A., Minovitsky, S., Dubchak, I. & Pennacchio, L.A. VISTA Enhancer Browser—a database of tissue-specific human enhancers. *Nucleic Acids Res.* **35**, D88–D92 (2007).