Figures and figure supplements

Mapping cell type-specific transcriptional enhancers using high affinity, lineage-specific Ep300 bioChIP-seq

Pingzhu Zhou et al
Figure 1. Generation and characterization of Ep300\textsuperscript{flbio} allele. (A) Experimental strategy for high affinity Ep300 pull down. A flag-bio epitope was knocked onto C-terminus of the endogenous Ep300 gene. The bio peptide sequence is biotinylated by BirA, widely expressed from the Rosa26 locus. (B) Targeting strategy to knock flag-bio epitope into the C-terminus of Ep300. A targeting vector containing homology arms, flag-bio epitope, and Frt-neo-Frt cassette was used to insert the epitope tag into embryonic stem cells by homologous recombination. Chimeraic mice were mated with Act:Flpe mice to excise the Frt-neo-Frt cassette in the germline, yielding the Ep300\textsuperscript{fb} allele. (C) Biotinylated Ep300 is quantitatively pulled down by streptavidin beads. Protein extract from Ep300\textsuperscript{flbio/flbio}, Rosa26\textsuperscript{BirA} embryos was incubated with immobilized streptavidin. Input, bound, and unbound fractions were analyzed for Ep300 by immunoblotting. GAPDH was used as an internal control. (D) Ep300\textsuperscript{flbio/flbio} mice from heterozygous intercrosses survived normally to weaning. DOI: 10.7554/eLife.22039.002
Figure 1—figure supplement 1. Characterization of Ep300 fb allele. (A) Southern blot genotyping of Ep300 fb allele. Probes are marked in Figure 1B. (B) The expression level of tagged Ep300 protein is similar to that of wild-type Ep300 in neonatal mouse hearts. (C, D) The heart weight to body weight ratio of 10-week-old Ep300 fb/fb heart was not significantly different from littermates. (E) Echocardiograms of 10-week old Ep300 fb/fb mice show that their ventricular function was not significantly different from littermate controls. Sample sizes are shown in bars in D, E. Bars represent mean ± standard deviation. n.s., not significantly different (t-test).

DOI: 10.7554/eLife.22039.003
Figure 2. Comparison of Ep300 bioChiP-seq to antibody ChIP-seq for mapping Ep300 chromatin occupancy. (A–B) Comparison of biological duplicate antibody Ep300 ChIP-seq (Encode) and Ep300 bioChiP-seq (flbio). The Ep300 bioChiP-seq data had greater overlap between replicates and greater
Figure 2 continued

Intragroup correlation. Most antibody peaks were covered by the bioChIP-seq data. There were 3.3 times more Ep300 regions detected by Ep300 bioChIP-seq. 89.6% of Ep300 regions detected by antibody ChIP-seq were recovered by Ep300 bioChIP. Panel B shows Spearman correlation between samples over the peak regions. (C) Tag heatmap shows input-subtracted Ep300 antibody or bioChIP signal in the union of the Ep300-bound regions detected by each method. (D) Correlation plots show greater Ep300 bioChIP-seq signal compared to antibody bioChIP-seq.

DOI: 10.7554/eLife.22039.004
Figure 3. Tissue specific Ep300 bioChiP-seq. (A) Tag heatmap showing Ep300\textsuperscript{fb} signal in heart and forebrain. Each row represents a region that was bound by Ep300 in heart, forebrain, or both. A minority of Ep300-enriched regions were shared between tissues. (B) Ep300\textsuperscript{fb} pull down from E12.5 heart or forebrain. Ep300-bound regions identified by bioChiP-seq of Ep300\textsuperscript{fb/\#}; Rosa26\textsuperscript{BirA} tissues in biological duplicates are compared to regions identified by antibody-mediated Ep300 ChIP-seq (Visel et al., 2009). (C) GO biological process terms enriched for genes that neighbor the tissue-specific heart or forebrain peaks. Bars indicate statistical significance. DOI: 10.7554/eLife.22039.005
Figure 3—figure supplement 1. Heart and forebrain Ep300 bioChiP-seq. (A) Ep300 bioChiP-seq signals within called Ep300 peaks were highly correlated between biological replicates and discordant between heart and forebrain. (B) Genome browser views showing Ep300 bioChiP-seq signal in heart and brain at Srf (preferential cardiac expression) and Nes (preferential Brain expression). (C) Comparison of Ep300 bioChiP-seq and H3K27ac ChIP-seq signals in heart. Numbers indicate correlation coefficients. There was excellent correlation between biological replicates of Ep300 or H3K27ac. Ep300 and H3K27ac also showed highly significant genome-wide correlation. Similar results were observed in forebrain (data not shown). (D) Tag heatmap of H3K27ac-bound regions showing overlap between heart and brain H3K27ac enriched regions from each tissue.

DOI: 10.7554/eLife.22039.006
**Figure 3—figure supplement 2.** Comparison of heart Ep300\(^{fb}\) regions to a compendium of heart enhancers assembled by Dickel *et al.* (2016). The Dickel heart enhancer compendium (*Dickel et al.*, 2016) is based on H3K27ac ChIP-seq and Ep300 antibody ChIP-seq. Most compendium regions had low enhancer scores and have low validation frequency in transient transgenic assays. Compendium regions were compared to heart Ep300\(^{fb}\) regions using the LiftOver tool to map human genome coordinates to mouse. (A) Venn diagram showing that 13% of prenatal heart compendium regions overlap with heart Ep300\(^{fb}\) regions. On the other hand, 47% of Ep300\(^{fb}\) heart regions overlap with heart compendium regions, and 53% do not. (B) Analysis of the fraction of compendium regions that overlap the Ep300\(^{fb}\) heart regions, as a function of enhancer score. Regions with higher enhancer score (and higher likelihood of in vivo validation) corresponded well with heart Ep300\(^{fb}\) regions.

DOI: 10.7554/eLife.22039.007
Machine learning with weighted KNN method
ROC curve calculated using 10x cross-validation

<table>
<thead>
<tr>
<th>Chromatin Feature</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>H3K27ac</td>
<td>0.747</td>
</tr>
<tr>
<td>H3K27me3</td>
<td>0.608</td>
</tr>
<tr>
<td>H3K36me3</td>
<td>0.579</td>
</tr>
<tr>
<td>H3K4me1</td>
<td>0.589</td>
</tr>
<tr>
<td>H3K4me2</td>
<td>0.646</td>
</tr>
<tr>
<td>H3K4me3</td>
<td>0.637</td>
</tr>
<tr>
<td>H3K9ac</td>
<td>0.540</td>
</tr>
<tr>
<td>Ep300 bioChip-seq</td>
<td>0.805</td>
</tr>
<tr>
<td>ATAC-seq</td>
<td>0.749</td>
</tr>
<tr>
<td>P300 + ATAC</td>
<td>0.866</td>
</tr>
<tr>
<td>All</td>
<td>0.862</td>
</tr>
</tbody>
</table>

Figure 3—figure supplement 3. Comparison of enhancer prediction based on indicated chromatin features of E12.5 heart. Data are from ENCODE except for ATAC-seq (Pu lab, unpublished), and Ep300<sup>b</sup> (this manuscript). The KNN classifier was trained on a subset of the VISTA enhancer database and validated on the remaining data, using a 10x cross-validation design. Ep300 was the most accurate single predictor, based on the area under the receiver operating characteristic curve (AUC, indicated in parentheses). All of the features together (All), or just Ep300 plus ATAC-seq together (not shown), gave the best prediction.

DOI: 10.7554/eLife.22039.008
Figure 4. Cre-directed, lineage-selective Ep300 bioChIP-seq. (A) Experimental strategy. Lineage-specific Cre recombinase activates expression of BirA (HA-tagged) from Rosa26-flox-stop-BirA (Rosa26\textsuperscript{fsBirA}). This results in Ep300 biotinylation in the progeny of Cre-expressing cells. (B) Cre-dependent Figure 4 continued on next page

Zhou et al. eLife 2017;6:e22039. DOI: 10.7554/eLife.22039
Ep300 biotinylation using Rosa26$^{fsBirA}$ Protein extracts were prepared from E11.5 embryos with the indicated genotypes. R26$^{BirA/+}$ (ca) and R26$^{fsBirA/+}$ indicate the constitutively active and Cre-activated alleles, respectively. (C) Immunostaining demonstrating selective Tie2Cre-mediated expression of HA-tagged BirA in ECs of R26$^{fsBirA}$, Tie2Cre embryos. Arrows and arrowheads indicate endothelial and hematopoietic lineages, respectively. (D) Tissue-selective Ep300 bioChIP-seq. Tie2Cre (T; endothelial and hematopoietic lineages), Myf5Cre (M; skeletal muscle lineages), and EllaCre (E; germline activation) were used to drive tissue-selective Ep300 bioChIP-seq. Correlation between ChIP-seq signals within peak regions are shown for triplicate biological repeats. Samples within groups were the most closely correlated. (E) Ep300$^{fb}$ bioChIP-seq signal at Gata2 (EC/blood specific) and Myod (muscle specific). Enhancers validated by transient transgenic assay are indicated along with the citation’s Pubmed identifier (PMID). (F) Biological process GO terms illustrate distinct functional groups of genes that neighbor Ep300 bioChIP-seq driven by lineage-specific Cre alleles. The heatmap contains the top 10 terms enriched for genes neighboring each of the three lineage-selective Ep300 regions.

DOI: 10.7554/eLife.22039.009
Figure 4—figure supplement 1. Cre-dependent Ep300 bioChIP-seq. (A) Cre-dependent expression of BirA. Heart with or without cardiac-specific TNTCre was immunoblotted for expression of BirA from the Rosa26-fsBirA/ allele. (B) Distribution of Ep300 fb bioChIP-seq signal when driven by Tie2Cre or Myf5Cre. Regions with Ep300 peaks called in samples from any of the three Cre drivers were scored for their EC- or muscle- enriched signal by calculating the ratio of the region’s signal in Tie2Cre or Myf5Cre bioChIP-seq to its signal with ubiquitous Ep300 bioChIP-seq (EIIaCre). The Tie2Cre-enrichment score showed a bimodal distribution with a small peak at values greater than 1.5. A value of greater than 1.5 similarly was used to define Myf5Cre-enriched regions. (C) Relationship of Tie2Cre-enriched (Ep300-T-fb) and Myf5Cre-enriched (Ep300-M-fb) regions to peaks called from Tie2Cre or Myf5Cre samples alone.

DOI: 10.7554/eLife.22039.010
Figure 5. Functional validation of enhancer activity of Ep300-T-fb-bound regions. (A) Expression of genes neighboring regions bound by Ep300 in different Cre-marked lineages. Translating ribosome affinity purification (TRAP) was used to enrich for RNAs from the Tie2Cre lineage. Input or Tie2Cre-enriched RNAs were profiled by RNA sequencing. The expression of the nearest gene neighboring Ep300 regions in ECs (Ep300-T-fb regions), but not skeletal muscle (Ep300-M-fb regions), was higher than that of genes neighboring regions bound by Ep300 across the whole embryo (EIIaCre). Box and whiskers show quartiles and 1.5 times the interquartile range. Groups were compared to EIIaCre using the Mann-Whitney U-test. (B) Transient transgenesis assay to measure in vivo activity of Ep300-T-fb regions. Test regions were positioned upstream of an hsp68 minimal promoter and lacZ. Embryos were assayed at E11.5. (C) Summary of transient transgenic assay results. Out of 20 regions tested, nine showed activity in ECs or blood in three or more embryos, and two more showed activity in two embryos. See also Table 2. (D) Representative whole mount Xgal-stained embryos. Enhancers that directed LacZ expression in an EC or blood pattern in two or more embryos are shown. Numbers indicate embryos with LacZ distribution similar to shown image, compared to the total number of PCR positive embryos. (E) Sections of Xgal-stained embryos showing examples of enhancers active in arteries, veins, and endocardium, or selectively active in arteries or veins. AS: aortic sac; CV: cardinal vein; DA: dorsal aorta; EC: endocardial cushion; HV: head vein; LA: left atrium; LV: left ventricle; RV: right ventricle. Scale bars, 100 μm. See also Figure 5—figure supplements 1 and 2 and Table 2.

Figure 5 continued on next page
Figure 5 continued

DOI: 10.7554/eLife.22039.011
Figure 5—figure supplement 1. Relationship of EC gene expression to Ep300 regions in ECs, skeletal muscle, and whole embryo. (A) EC enrichment of actively translating genes that neighbored Ep300 regions in skeletal muscle (Ep300-M-fb), whole embryo (Ep300-E-fb), and endothelial cell/blood (Ep300-T-fb), using several different thresholds for the maximum distance allowed to associate genes to regions. The relationship between Ep300-T-fb regions and gene expression was not dependent upon the threshold used. The rightmost panel (no limit to maximum distance) is the same as in Figure 5A. Mann-Whitney U-test. (B) EC expression of actively translating genes that neighbored Ep300-T-fb, compared to those that did not, using several different thresholds for the maximum distance used associate genes to regions. Ep300-T-fb-associated genes were more highly expressed regardless of the distance threshold used. Ep300-T-fb genes were significantly more highly expressed (Mann-Whitney U-test). (C) Fraction of genes expressed or not expressed, for Ep300-T-fb-associated genes compared to all genes. Ep300-T-fb-associated genes were more highly expressed than non-associated genes. However, not all genes associated with Ep300-T-fb regions were expressed, and some genes without Ep300-T-fb regions were expressed. This may reflect additional mechanisms of gene regulation as well as limitations of the gene-to-region association rule.
Figure 5—figure supplement 2. Transient transgenic assays to measure in vivo activity of candidate endothelial cell/blood enhancers. For each tested enhancer with positive endothelial cell or blood activity, we show the candidate enhancer’s Ep300 signal in skeletal muscle (Ep300-M-fb), whole embryo.
Figure 5—figure supplement 2 continued

(Ep300-E-fb), and endothelial cell/blood (Ep300-T-fb), with respect to the gene both. The whole mount images show each of the positive Xgal-stained embryos.

DOI: 10.7554/eLife.22039.013
**Figure 5—figure supplement 3.** Histological sections of transient transgenic embryos. Histological sections of embryos containing the indicated enhancer-driven lacZ transgene. AS: aortic sac; CV: cardinal vein; DA: dorsal aorta; EC: endocardial cushion; HV: head vein; PA: pulmonary artery; RA: right atrium; RV: right ventricle. Bar, 100 μm.

DOI: 10.7554/eLife.22039.014
**Figure 6.** Motifs enriched in Ep300-T-fb and Ep300-M-fb regions. (A) Motifs enriched in Ep300-T-fb or Ep300-M-fb regions. 1445 motifs were tested for enrichment in Ep300 bound regions compared to randomly permuted control regions. Significantly enriched motifs (neg ln p-value>15) were clustered and the displayed non-redundant motifs were manually selected. Heatmaps show statistical enrichment (left), fraction of regions that contain the motif (center), and GO terms associated with genes neighboring motif-containing, Ep300-bound regions (fraction of the top 20 GO biological process terms). Grey indicates that the motif was not significantly enriched (neg ln p-value≤15). (B) Conservation of sequences matching Tie2Cre- or Myf5Cre-enriched motifs within 100 bp of the summit of Ep300-T-fb or Ep300-M-fb regions, compared to randomly selected 12 bp sequences from the same regions. PhastCons conservation scores across 30 vertebrate species were used. ****p<0.0001, Kolmogorov–Smirnov test. (C) Luciferase assay of activity of enhancers containing indicated motifs. Three repeats of 20–30 bp regions from Ep300-bound enhancers linked to the indicated gene and centered on the indicated motif were cloned upstream of a minimal promoter and luciferase. The constructs were transfected into human umbilical vein endothelial cells. Luciferase activity was expressed as fold activation above that driven by the enhancerless, minimal promoter-luciferase construct. *p<0.05 compared to MeI2c-enhancer with mutated ETS:FOX2 motif (MeI2c mut). n = 3. (D) Luciferase assay of indicated motifs repeated three times within a consistent DNA context. Assay was performed as in C. *p<0.05 compared to negative control sequencing lacking predicted motif. n = 3. Error bars in C and D indicate standard error of the mean.

DOI: 10.7554/eLife.22039.017
[Figure 6—figure supplement 1. Motifs with enriched Ep300 occupancy in Tie2Cre or Myf5Cre embryo samples. (A) De novo motif discovery. The top 10 discovered motifs are shown, with the significance P-value, percent of regions containing the motif, and the best matching known motif class. (B) Motifs with enrichment in Myf5Cre vs. Tie2Cre samples.](Zhou et al. eLife 2017;6:e22039). DOI: 10.7554/eLife.22039}
Conservation of Tie2Cre-enriched motifs (dashed lines) or Myf5Cre-enriched motifs (solid lines) within 100 bp of Ep300-T-fb or Ep300-M-fb peak summits. The distribution of conservation scores for each indicated motif is shown as a cumulative frequency plot. Random indicates randomly sampled 12 nt regions from within the Ep300-T-fb or Ep300-M-fb regions.

DOI: 10.7554/eLife.22039.018
Figure 7. Enhancers in adult heart and lung ECs. VEcad-CreERT2 and neonatal tamoxifen pulse was used to drive BirA expression in ECs. Ep300 regions in adult heart or lung ECs were identified by bioChIP-seq in biological triplicate. Ep300 regions were ranked into deciles by the ratio of the

DOI: 10.7554/eLife.22039

Zhou et al. eLife 2017;6:e22039. DOI: 10.7554/eLife.22039
Figure 7 continued

Ep300-VE-fb signal in heart to lung. (A) Tag heatmap shows that the top and bottom deciles have selective Ep300 occupancy in heart and lung ECs, respectively. (B) Expression of genes in heart ECs (red) or lung ECs (blue) neighboring Ep300 regions, divided into deciles by Ep300-VE-fb signal ratio in heart and lung. ***p<0.0001, Wilcoxon test. Expression values were obtained from (Nolan et al., 2013). Box plots indicate quartiles, and whiskers indicate 1.5 times the interquartile range. (C) Genes with selective expression in heart or lung ECs were identified by K-means clustering (see Materials and methods and Figure 7—figure supplement 1). The number of Ep300-VE-fb regions in heart (red) or lung (blue) neighboring these genes was determined, stratified by decile. ***p<0.0001, Fisher’s exact test. (D) Enrichment of selected motifs in Ep300-VE-fb regions from heart (decile 1) compared to lung (decile 10) or vice versa. Grey indicates no significant enrichment. Displayed non-redundant motifs were selected from all significant motifs by manual curation of clustered motifs.

DOI: 10.7554/eLife.22039.019
Figure 7—figure supplement 1. Organ-specific EC gene expression. (A) Genes with at least four-fold differential mean expression between highest and lowest EC samples were identified from microarray data (Nolan et al., 2013). These genes were grouped by k-means clustering. Genes in clusters containing selective heart or lung expression were analyzed further. Expression of these heart or lung EC genes is displayed in panel (B) using hierarchical clustering.

DOI: 10.7554/eLife.22039.020
Figure 7—figure supplement 2. VEcad-CreERT2 activation of Ep300 biotinylation. (A) VEcad-CreERT2 activated Ep300 biotinylation by Rosa26fsBirA, permitting its pulldown on streptavidin (SA) beads. (B) EC-selective activation of the Rosa26mTmG Cre reporter allele by VEcad-CreERT2. Tamoxifen was administered at P1 and P2, and tissues were analyzed at six weeks of age.

DOI: 10.7554/eLife.22039.021
Figure 7—figure supplement 3. Ep300 bioChIP-seq signal in whole organ or ECs. Ep300 bioChIP-seq signal in whole organ ('Con'; EIIaCre) or ECs ('EC'; VEcad-CreERT2) at genes with EC or non-EC expression in lung or heart. DOI: 10.7554/eLife.22039.022
Figure 7—figure supplement 4. Gene functional terms enriched in decile 1 or decile 10 of Ep300-VE-fb regions. GO biological process (top) or disease ontology (bottom) terms enriched in decile 1 or decile 10 of Ep300-VE-fb regions. Grey indicates no significant enrichment. The top five biological process terms in deciles 1 and 10 and selected relevant disease ontology terms were displayed.

DOI: 10.7554/eLife.22039.023