Tbx5 maintains atrial identity in postnatal cardiomyocytes by regulating an atrial-specific enhancer network

Mason E. Sweat [1,6], Yangpo Cao [1,2,6], Xiaoran Zhang [1],
Ozanna Burnicka-Turek[3], Carlos Perez-Cervantes [3], Arulsamy Kulandaisamy[1],
Fujian Lu[1], Erin M. Keating[1], Brynn N. Akerberg[1], Qing Ma[1], Hiroko Wakimoto[4],
Joshua M. Gorham [4], Lauren D. Hill[1], Mi Kyoung Song[1,5], Michael A. Trembley[1],
Peizhe Wang[1], Matteo Gianeselli[1], Maksymilian Prondzynski [1],
Raul H. Bortolin[1], Vassilios J. Bezzerides[1], Kaifu Chen[1], Jonathan G. Seidman [4],
Christine E. Seidman[4], Ivan P. Moskowitz[3] & William T. Pu [1]✉

Understanding how the atrial and ventricular heart chambers maintain distinct identities is a prerequisite for treating chamber-specific diseases. In this study, we selectively knocked out the transcription factor *Tbx5* in the atrial working myocardium to evaluate its requirement for atrial identity. Atrial *Tbx5* inactivation downregulated atrial cardiomyocyte (aCM)-selective gene expression. Using concurrent single-nucleus transcriptome and open chromatin profiling, genomic accessibility differences were identified between control and *Tbx5* knockout aCMs, revealing that 69% of the control-enriched ATAC regions were bound by TBX5. Genes associated with these regions were downregulated in knockout aCMs, suggesting that they function as TBX5-dependent enhancers. Comparing enhancer chromatin looping using H3K27ac HiChIP identified 510 chromatin loops sensitive to TBX5 dosage, and 74.8% of control-enriched loops contained anchors in control-enriched ATAC regions. Together, these data demonstrate that TBX5 maintains the atrial gene expression program by binding to and preserving the tissue-specific chromatin architecture of atrial enhancers.

The cardiac atria rapidly propagate the electrical impulse of the heartbeat and coordinately contract to efficiently move blood to the cardiac ventricles[1]. Disrupting these processes results in reduced cardiac pump function and abnormal heart rhythms, including atrial fibrillation (AF), a common and life-shortening arrhythmia[2]. Understanding the pathobiology of atrial diseases such as AF and developing improved treatments requires an in-depth understanding of the regulatory networks that support healthy atrial function.

Atrial and ventricular cardiomyocytes (aCMs and vCMs) are distinct cardiomyocyte subtypes specialized for distinct physiological functions[3–5]. The mechanisms that specify and maintain these differences are only beginning to be elucidated. Factors shown to promote ventricular identity and impede aCM gene expression include FGF, NKX2-5/2-7, IRX4, HEY2 and HRT2 (refs. [6–11]). In contrast, only a single transcription factor (TF), NR2F2, has been demonstrated to specify aCMs during early embryonic development[12]. Inactivation of NR2F2 later in heart

[1]Department of Cardiology, Boston Children's Hospital, Boston, MA, USA. [2]Department of Pharmacology, School of Medicine, Southern University of Science and Technology, Shenzhen, China. [3]Department of Pediatrics, Pathology, and Human Genetics, University of Chicago, Chicago, IL, USA. [4]Department of Genetics, Harvard Medical School, Boston, MA, USA. [5]Department of Pediatrics, Seoul National University College of Medicine, Seoul, Korea. [6]These authors contributed equally: Mason E. Sweat, Yangpo Cao. ✉e-mail: william.pu@cardio.chboston.org
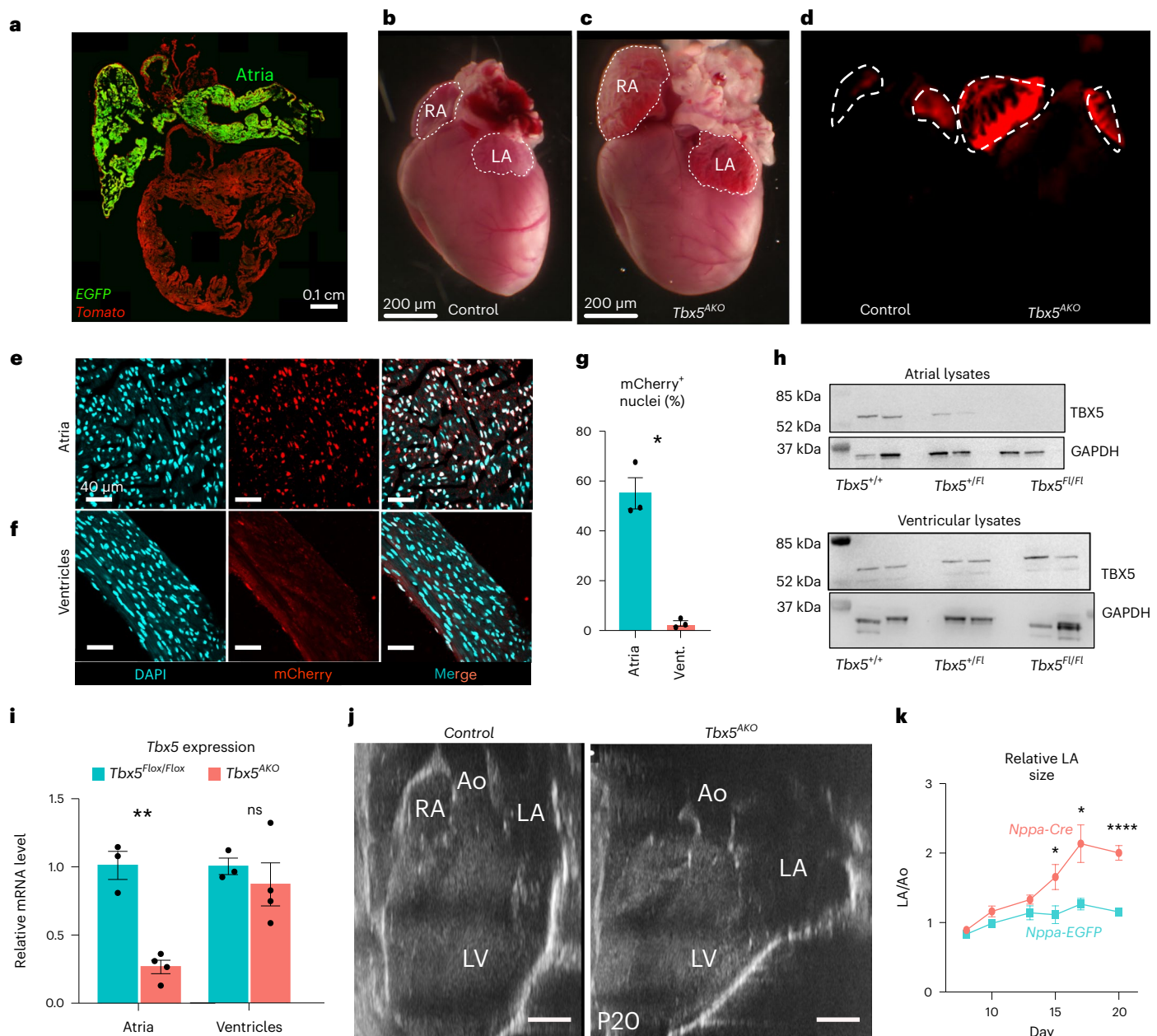
Fig. 1 | Inactivation of *Tbx5* in aCMs results in atrial remodeling. a–k, Mice were treated with 2×10^{11} VG/g AAV at P2. **a**, Rosa^{mTmG} Cre reporter mice were injected with AAV9:*Nppa-Cre*. Hearts were analyzed at P20. Cre-activated EGFP signal was restricted to the atria. **b–d**, Whole mount images of control and atrial-specific *Tbx5* KO (*Tbx5^{AKO}*) hearts. **b**,**c**, Bright-field. **d**, Red fluorescence channel. **e–g**, AAV9:*Nppa-Cre*-mediated recombination. H2B-mCh Cre reporter mice were treated with AAV9:*Nppa-Cre*. mCherry$^+$ nuclei in atria (**e**) and ventricles (**f**) were quantified (**g**) by confocal imaging of P20 myocardial sections. *$P = 0.012$; $n = 3$ hearts per group. **h**, TBX5 protein level. Atrial and ventricular lysates, prepared at P20, were analyzed by western blotting in biological duplicates. **i**, RT–qPCR for *Tbx5* mRNA from atria and ventricles. **$P = 0.0079$; NS, not significant ($P = 0.48$); $n = 3$ control or 4 *Tbx5^{AKO}* hearts per group. **j**,**k**, Echocardiographic assessment of atrial size. **j**, Representative echo images used to quantify atrial size, compared to aortic size. Bar, 1 mm. **k**, Relative size of control ($n = 5$) and *Tbx5^{AKO}* ($n = 4$) left atria, normalized to aortic diameter. P15, *$P = 0.04$; P17, *$P = 0.011$; P20, ****$P = 0.0001$. Unpaired, two-sided *t*-tests. Graphs show mean ± s.e.m. Ao, aorta; LA, left atrium; LV, left ventricle; RA, right atrium.

development did not alter atrial identity[12], demonstrating its selective role in atrial specification, but not maintenance, and highlighting that different factors may promote atrial specification versus maintenance. Retinoic acid has also been shown to be critical for atrial specification: disrupting the retinoic acid biosynthetic pathway in developing heart tubes caused hypoplastic development of the posterior heart tube (the primitive atria)[13], and retinoic acid has been used to promote aCM differentiation from induced pluripotent stem cells (iPSCs)[14].

The cardiac TF *TBX5* is enriched in postnatal aCMs and directly promotes the expression of several atrial specific genes, including

Nppa (atrial natriuretic factor), *Cx40* (connexin 40) and *Bmp10* (bone morphogenic protein 10) (ref. 15). *TBX5* mutations cause congenital cardiac malformations, including atrial septal defect and upper limb malformations (OMIM: 142900). A familial *TBX5* missense variant, *TBX5^{G125R}*, causes AF in patients and mice heterozygous for the point mutation[16]. Widespread postnatal *Tbx5* inactivation in mice using *Rosa26^{CreER}* also caused AF[17–19]. These data suggest that TBX5 is essential for atrial homeostasis, albeit TBX5's essential functions in vCMs, conduction system and nodal tissues potentially confound the interpretation of this widespread inactivation model[20,21].

Here we dissect TBX5's regulation of atrial gene regulatory networks. Using a model of postnatal, aCM-selective *Tbx5* inactivation, we demonstrate alteration of aCM structure and rapid development of AF. Concurrent single-nucleus transcriptome and open chromatin sequencing (snRNA-seq and snATAC-seq) coupled with H3K27ac HiChIP showed that aCMs lacking *Tbx5* lose accessibility and chromatin looping at many enhancers of aCM-selective genes.

Results

Nppa-Cre specifically inactivates Tbx5 in aCMs

To inactivate a loxP-flanked *Tbx5* allele (*Tbx5^Flox*) selectively in postnatal aCMs, we expressed Cre using the *Nppa* promoter and the cardiotropic adeno-associated virus serotype 9 (AAV9). AAV9:*Nppa-Cre* was previously shown to selectively recombine floxed alleles in atria[22,23]. In control experiments, we demonstrated that AAV9:Nppa-Cre is active in aCMs and not SAN or AVN (Extended Data Fig. 1), consistent with transgenic lines established with this promoter[24].

We characterized AAV9:*Nppa-Cre* inactivation of *Tbx5^Flox*. Inclusion of the Cre-inducible H2b-mCherry reporter line (*Rosa^H2B-mch*) allowed us to identify myocytes transduced by AAV9:*Nppa-Cre*. At postnatal day 2 (P2), *Tbx5^+/+*;*Rosa^H2B-mch*, *Tbx5^Flox/+*;*Rosa^H2B-mch* and *Tbx5^Flox/Flox*;*Rosa^H2B-mch* mice were treated with AAV9:*Nppa-Cre*. At P20, H2B-mCherry fluorescence was localized to the atria, and AAV9:*Nppa-Cre*-treated *Tbx5^Flox/Flox* atria were markedly enlarged (Fig. 1b–d). In histological sections, aCMs, but not vCMs, exhibited prominent nuclear mCherry signal (Fig. 1e–g). Immunoblots revealed loss of TBX5 protein in atrial lysates from *Tbx5^Flox/Flox* mice compared to control Tbx5^+/+ mice and no reduction in ventricular lysates (Fig. 1h and Source Data Fig. 1). TBX5 was partially reduced in Tbx5^Flox/+ mice, consistent with heterozygosity. RT–qPCR confirmed 80% reduction of *Tbx5* transcripts in *Tbx5^Flox/Flox* atria and no significant change in *Tbx5^Flox/Flox* ventricles (Fig. 1i). Taken together, these results demonstrate that AAV9:*Nppa-Cre* effectively and selectively inactivates *Tbx5* in aCMs. Hereafter, we refer to *Tbx5^Flox/Flox* mice treated with AAV9:*Nppa-Cre* as *Tbx5^AKO* (aCM knockout (KO)).

Atrial remodeling and fibrillation in Tbx5^AKO mice

To further examine the atrial phenotype of *Tbx5^AKO mice*, we performed serial echocardiography from P8 to P20 (Fig. 1j). After P15, the atria were significantly dilated in *Tbx5^AKO* compared to controls treated with AAV9:*Nppa-EGFP* (Fig. 1k). At P20, ventricular function did not differ significantly between control and *Tbx5^AKO* mice (Extended Data Fig. 2f). We assessed fibrosis by collagen staining (Masson's trichrome) and RT–qPCR for fibrosis genes *Col1a1* and *Postn* (Extended Data Fig. 2a,b).Unlike some other AF models that showed atrial fibrotic remodeling[22,23], we did not observe fibrosis in *Tbx5^AKO* at this early timepoint.

We assessed aCM morphology by immunostaining for the Z-line protein sarcomeric α-actinin (SAA) and the junctional sarcoplasmic reticulum (jSR) protein FSD2 (ref. 25). In control aCMs and aCMs with one functional copy of *Tbx5*, we observed the expected striated Z-line staining pattern (Extended Data Fig. 2c–c′,d–d′). However, in aCMs lacking both copies of *Tbx5*, this striated pattern was absent in many myocytes, as illustrated by plotting SAA signal intensity along the aCM long axis (Extended Data Fig. 2e).

Because widespread *Tbx5* inactivation caused AF[17–19], we obtained electrocardiograms (EKGs) on *Tbx5^AKO* mice at P21. *Tbx5^Flox/Flox* mice treated with AAV9:*Nppa-GFP* (control) were in normal sinus rhythm, as demonstrated by P-waves preceding each QRS complex and regular intervals between QRS complexes (RR intervals) (Extended Data Fig. 2g). In *Tbx5^AKO* mice, the EKG changed markedly and exhibited loss of P-waves and irregularly irregular RR intervals (Extended Data Fig. 2g), hallmarks of AF. The change in RR regularity was reflected by the dispersion of points in Poincaré plots of the RR interval of two successive beats in *Tbx5^AKO* compared to control (Extended Data Fig. 2h). We quantified the irregularity by calculating the standard deviation of the RR interval (SDRR) at P21. Strikingly, the mean SDRR was five times greater in *Tbx5^AKO* animals compared to controls (Extended Data Fig. 2i). Serial measurements between P8 and P21 showed significant elevation of SDRR after P14 (Extended Data Fig. 2j), indicating that *Tbx5* is required in aCMs to maintain atrial rhythm even in young mice.

To confirm the diagnosis of AF and exclude other potential diagnoses, such as sinus node dysfunction, we recorded the surface EKG and right atrial electrogram in *Tbx5^Flox/+* and *Tbx5^Flox/Flox* mice treated with AAV9:*Nppa-Cre*. Recordings from control mice showed normal, synchronous atrial-ventricular rhythm (Extended Data Fig. 2k). In contrast, *Tbx5^AKO* mice exhibited nearly continuous low-amplitude atrial activity and irregular ventricular rhythm, consistent with AF (Extended Data Fig. 2I).

Tbx5 is required to maintain the atrial gene program in aCMs

Recent work aimed at identifying important regulators of chamber identity uncovered the TBOX motif as an enriched motif in atrial-specific enhancers and demonstrated 11-fold higher *Tbx5* transcripts in aCMs compared to vCMs[3]. Other studies demonstrated that TBX5 is required for expression of atrial-specific genes, including *Nppa* and *Bmp10* (ref. 26). To further evaluate the requirement of *Tbx5* for expression of atrial genes, we stained for MYL7 and MYL4, atrial-specific myosin regulatory light chain proteins, in P21 heart sections from untreated (control) and AAV9:*Nppa-Cre*-treated *Tbx5^Flox/Flox*;*Rosa^H2B-mch* or *Tbx5^Flox/+* mice. MYL7 and MYL4 immunoreactivity similarly localized in Cre-treated *Tbx5^Flox/+* and untreated *Tbx5^Flox/Flox* aCM A-bands, demonstrating that *Tbx5* heterozygosity is sufficient to support aCM expression of MYL7 and MYL4 (Fig. 2a,b). In contrast, *Tbx5^AKO* aCMs lacking *Tbx5* (mCherry^+) lost MYL7 and MYL4 staining almost entirely (Fig. 2a,b, arrows). Within the same tissue, neighboring aCMs that did not undergo Cre recombination (mCherry^−) retained MYL7 or MYL4, indicating that this phenotype is cell autonomous and not secondary to tissue-level abnormalities, such as AF. In AAV9:*Nppa-Cre*-treated *Tbx5^+/+* heart sections, immunoreactivity for MYL2, the myosin light chain selectively expressed in vCMs, was robust in vCMs and not detected in aCMs (Fig. 2c). In contrast, MYL2 expression was detected in *Tbx5^AKO* aCMs (Fig. 2d). RT–qPCR confirmed downregulation of aCM-specific transcripts *Myl7*, *Myl4* and *Nppa* and upregulation of vCM-specific transcript *Myl2* in *Tbx5^AKO* aCMs (Fig. 2e). However, there was no significant change in the transcript *Nppb*, a gene adjacent to the *Nppa* locus that normally exhibits slightly biased atrial expression[3] (aCM/vCM ~1.6) (Fig. 2e).

To better understand how *Tbx5* inactivation alters atrial gene expression, we analyzed previously reported bulk RNA sequencing

---

Fig. 2 | Inactivating *Tbx5* alters the expression of aCM-selective genes. a,b, LA myocardial expression of aCM-selective myosin light chain MYL7 (**a**) or MYL4 (**b**). Arrows, nuclear mCherry indicative of cell transduction by AAV9:*Nppa-Cre*. White boxed regions are magnified in insets. Bar, 40 μm. **c,d**, Expression of vCM-selective myosin light chain MYL2. White boxed regions are enlarged to the right. Arrows, *Tbx5^AKO* atrial regions that express MYL2. Bar, 100 μm. **e**, Transcript levels in control and *Tbx5^AKO* atria. *Myl7*, *Myl4*, *Myl2*, *Nppa* and *Nppb* transcripts were measured from pooled left and right atrial samples of the indicated genotype. Graphs show mean ± s.e.m. Unpaired two-sided *t*-test. *Myl7*: $P = 2.4 \times 10^{-8}$; *Myl4*: $P = 1.3 \times 10^{-5}$; *Nppa*: $P = 2.5 \times 10^{-6}$; *Myl2*: $P = 0.0041$; *Nppb*: $P = 0.087$. $n = 9$ (*Tbx5^AKO*) and 7 (control), except for *Myl4* control $n = 8$. **f**, Genome-wide effect of *Tbx5* atrial KO on expression of LA genes. PCA was used to compare transcriptomes of isolated aCMs, isolated vCMs, control LA and global adult *Tbx5* KO LA (GSE129503). Inactivating *Tbx5* shifted aCM gene expression toward vCMs along PC1 (black arrow). **g**, Definition of aCM-selective (red) and vCM-selective (blue) transcripts based on P0 RNA-seq of aCMs and vCMs (from GSE215065; $\log_2 FC > 1.5$, $P_{adj} < 0.05$, Wald test followed by Benjamini–Hochberg correction). **h**, Enrichment of aCM-selective and vCM-selective genes defined in **g** among genes downregulated and upregulated, respectively, in *Tbx5^iKO* LA samples. Normalized enrichment score (NES) and FDR by GSEA.

(RNA-seq) of left atrial (LA) tissue with and without adult-induced, ubiquitous *Tbx5* KO (LA-*Tbx5*[iKO])[17]. Principal component analysis (PCA) plots of aCM, vCM, LA-control and LA-*Tbx5*[iKO] RNA-seq datasets showed

that *Tbx5*-ablated atrial gene expression shifted away from aCMs and toward vCMs on principal component 1 (PC1) (Fig. 2f). Next, we defined aCM-selective and vCM-selective gene sets based on bulk RNA-seq of
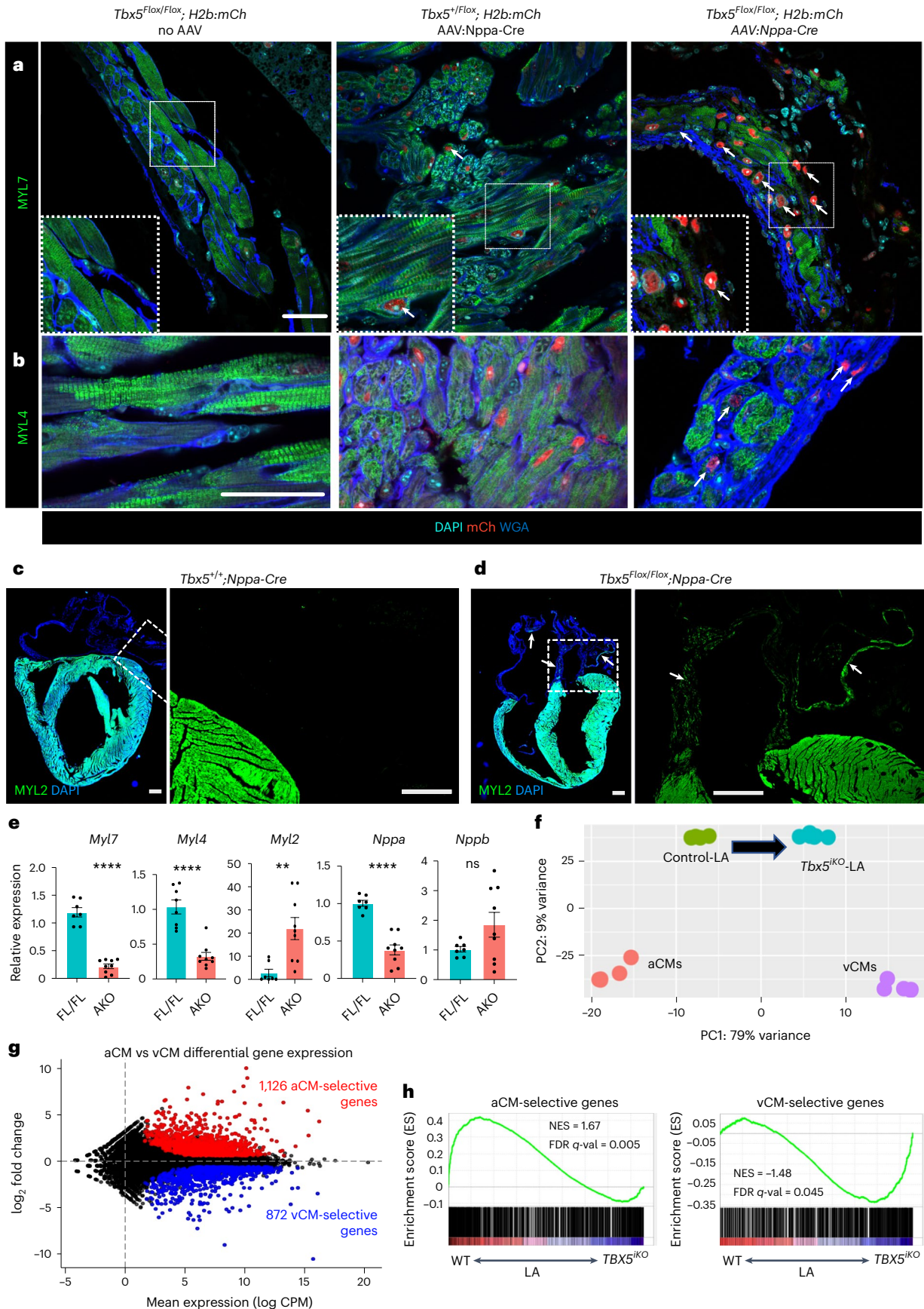
Fig. 2 | ...

P0 aCMs and vCM[3]. Differential gene expression analysis ($P_{adj} < 0.5$, $\log_2$ fold change (FC) > 0.58) identified 1,126 aCM-selective and 872 vCM-selective genes (Fig. 2g). Genes downregulated by *Tbx5* KO were enriched for aCM-selective genes, whereas genes upregulated by *Tbx5* KO were enriched for vCM-selective genes (Fig. 2h). Together, these data demonstrate that *Tbx5* is required to maintain expression of a subset of atrial-specific genes and downregulates the expression of a subset of ventricular genes.

Tbx5 overexpression promotes atrial gene expression in vCMs

To determine if increased *Tbx5* is sufficient to promote expression of atrial genes, we overexpressed *Tbx5* in vCMs using a Cre-dependent *Tbx5* transgene[27] and cardiomyocyte-specific Myh6-Cre (*Tbx5-OE*) (Extended Data Fig. 3a). Bulk RNA-seq of *Tbx5-OE* and control left ventricular myocardium at 6 weeks of age revealed thousands of significantly altered transcripts (Extended Data Fig. 3b,c and Supplementary Tables 1 and 2). There were 156 upregulated aCM-selective genes in ventricular myocardium compared to only 57 that were downregulated (Extended Data Fig. 3d; Fisher's exact test: $P < 1 \times 10^{-5}$). Furthermore, TBX5 overexpression downregulated 153 vCM-selective genes compared to only 63 that were upregulated (Extended Data Fig. 3e; Fisher's exact test: $P < 1 \times 10^{-5}$). These data demonstrate that *Tbx5* overexpression is sufficient to drive the transcriptome of established vCMs toward an aCM profile, activating 14% of all aCM-selective genes while repressing 18% of vCM-selective genes. Moreover, immunostaining of control and TBX5-OE hearts for atrial-specific myosin light chains MYL4 and MYL7 demonstrated ectopic expression of MYL4, but not MYL7, in TBX5-OE ventricles (Extended Data Fig. 3f,g). Expression of ventricle-specific MYL2 was unaltered (Extended Data Fig. 3h).

Multiomic analysis of Tbx5^{AKO} and control LA tissue

Human AF typically initiates in the left atrium. Therefore, we focused on the left atrium to gain insights into AF pathogenesis in *Tbx5^{AKO}* mice. We isolated P21 left atria of *Tbx5^{Flox/Flox}* animals treated with AAV9:*Nppa-GFP* (control) or AAV9:*Nppa-Cre* (*Tbx5^{AKO}*) and concurrently measured gene expression and accessible chromatin from single nuclei (concurrent snRNA-seq and snATAC-seq from each nucleus), in biological duplicate (Fig. 3a). In total, 14,583 nuclei passed rigorous quality and doublet filters, and an average of 1,926 genes were detected per nucleus (Extended Data Fig. 4a and Supplementary Table 1). Accessible chromatin from each of the samples demonstrated the expected enrichment at the transcription start site (TSS) of mouse genes (Extended Data Fig. 4b). Uniform manifold approximation and projection (UMAP) embeddings were created separately for the transcriptome and accessibility assays, and then the datasets were integrated using a weighted nearest neighbor (WNN) approach (Fig. 3b)[28–30]. Biological duplicate samples showed excellent overlap between replicates (Extended Data Fig. 4c), demonstrating reproducibility of the data. We identified clusters corresponding to 11 different cell states, which were annotated using established cardiac cell marker genes. Remarkably, aCMs from control and *Tbx5^{AKO}* samples largely clustered separately (Extended Data Fig. 4c, arrows).

Higher-resolution clustering of aCMs revealed four cell states derived from control samples and two from *Tbx5^{AKO}* samples (Fig. 4a). Trajectory analysis[31] showed that predominantly control (Myocyte_1 to Myocyte_4) and *Tbx5^{AKO}* (Myocyte_5 and Myocyte_6) clusters arranged into separate trajectories (Fig. 4a). Genes differentially expressed between either end of the pseudotime trajectories did not correspond to genes differentially expressed between P28 and P0 aCMs (Extended Data Fig. 5a–d), indicating that the trajectories do not correspond to chronological aCM maturation (Extended Data Fig. 5c–e). These genes were enriched for similar Gene Ontology (GO) terms, including muscle contraction and sarcomere organization, in the late pseudotime clusters (Extended Data Fig. 5f,g). Terms enriched in the early pseudotime clusters were not specific

for cardiomyocytes and included extracellular matrix organization, regulation of cell migration and vascular development. These data suggest that the clusters at the terminal end of the trajectories, Myocyte_1 and Myocyte_6, represent cardiomyocytes with the highest expression of cardiomyocyte functional components. Consistent with this conclusion, a functional index of the sarcomeric and calcium handling genes *Atp2a2*, *Ryr2*, *Myh6*, *Ttn*, *Trdn* and *Actn2* was highest in the late pseudotime clusters (Extended Data Fig. 5h). Unless stated otherwise, subsequent analyses are primarily focused on these more functional cardiomyocyte cell states, which we refer to as the 'control aCM' (Myocyte_1) and 'KO aCM' (Myocyte_6) clusters.

We identified gene expression differences between the control aCM and KO aCM clusters (Supplementary Table 3). Differentially expressed genes (DEGs) upregulated in the control aCM cluster were enriched for functional terms associated with heart contraction and cardiac conduction, including regulation of the atrial action potential, a known *Tbx5*-dependent feature[17] (Fig. 4b). Among these genes were ion channel genes *Scn5a*, *Atp2a2* and *Ryr2*—known direct TBX5 targets that are critical for action potential propagation and calcium handling[32,33] (Fig. 4b). *Sbk2*, a gene recently shown to be essential for atrial sarcomere integrity[34], was also more highly expressed in control compared to *Tbx5^{AKO}* aCMs, possibly contributing to the observed disorganization of *Tbx5^{AKO}* sarcomeres (Extended Data Fig. 2c,d). In the KO aCM cluster, there was an upregulation of structural and stress-induced genes, including *Col4a3*, *Col4a4* and *Ankrd1* (Fig. 4c). *Ank3*, previously linked with Brugada syndrome[35], was also upregulated in the KO aCM cluster.

To further investigate the role of *Tbx5* in regulation of the aCM gene program, we analyzed the distribution of aCM-selective and vCM-selective genes (Fig. 2e) among genes differentially expressed in the control and KO aCM clusters (Fig. 4d,i–iii). Strikingly, 57% of the genes enriched in the control aCM cluster were aCM selective compared to only 22% of genes in the KO aCM cluster (Fig. 4e,i; Fisher's exact test: $P < 1 \times 10^{-5}$). Conversely, 30% of the genes enriched in the KO aCM cluster were vCM selective versus 19% of genes more highly expressed in control aCMs (Fig. 4e,iii; Fisher's exact test: $P = 0.005$). Analysis of LA-control and LA-*Tbx5^{iKO}* bulk RNA-seq yielded similar results: aCM-selective genes were disproportionately upregulated in LA-control compared to LA-*Tbx5^{iKO}* (17% versus 8%) and vCM-selective genes in LA-*Tbx5^{iKO}* compared to LA-control (9% versus 5%; Fig. 4d,e,iv–vi). The decrease in aCM-selective and increase in vCM-selective expression after *Tbx5* inactivation suggest that aCMs lacking *Tbx5* lose features of the atrial gene program and acquire a more mixed phenotype.

To further test this hypothesis, we compared gene expression changes in *Tbx5^{AKO}* aCMs to those observed in LA tissue after global, adult-induced Tbx5 KO (*Tbx5^{iKO}* (ref. 17); Extended Data Fig. 6a). Considering methodological and biological differences between these experiments, the gene expression data showed excellent correlation (Pearson's correlation = 0.6; $P = 1 \times 10^{-105}$). Most DEGs showed concordant behavior between experiments, and aCM-selective genes were overrepresented in genes downregulated by *Tbx5* KO (quadrant I; Extended Data Fig. 6b) in both experiments. This result demonstrates that the inactivation of *Tbx5* in adults (*Tbx5^{iKO}* (ref. 17)) or in neonates (this study) leads to similar gene expression changes in aCMs.

We used a similar approach to compare the *Tbx5^{AKO}* aCM loss of function data to two datasets with *Tbx5* gain of function: (1) aCM gene expression changed caused by deletion of a TBX5 intronic regulatory element in mice (*Tbx5^{RE(int)KO}*), which modestly increased TBX5 level and elevated AF susceptibility[36]; and (2) ventricular tissue gene expression changed due to *Tbx5* overexpression (described in Extended Data Fig. 3). Comparison of aCM gene expression changes due to modest *Tbx5* overexpression in *Tbx5^{RE(int)KO}* to Tbx5 inactivation in *Tbx5^{AKO}* demonstrated strong negative correlation (Pearson's $r = -0.41$, $P = 1.5 \times 10^{-11}$), indicating consistent responses to altered Tbx5 levels between models (Extended Data Fig. 6c). Comparison of ventricular
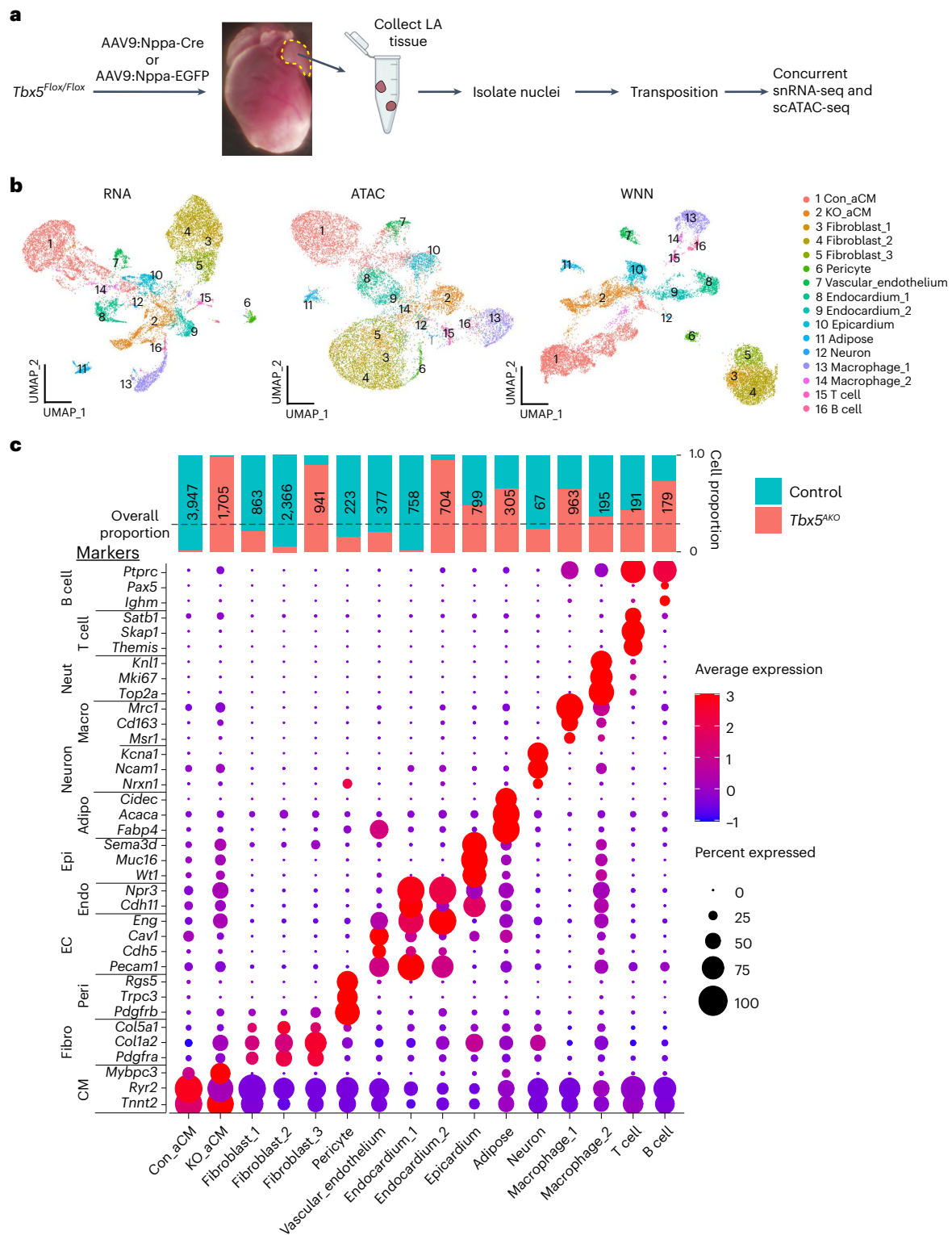
Fig. 3 | Concurrent scRNA-seq and scATAC-seq analysis of cell states in control and *Tbx5^AKO* atria. a, Experimental design. Neonatal *Tbx5^Flox/Flox* mice were treated with AAV9:*Nppa-EGFP* (control) or AAV9:*Nppa-Cre* (KO). At P20, isolated nuclei from two left atria per replicate were analyzed by concurrent scATAC-seq and scRNA-seq. Two replicates were prepared per group. **b**, Clustering of cell states based on scRNA-seq, scATAC-seq or both assays using WNN analysis. **c**, Cluster identities were established using cell-type-specific markers. The total number of nuclei and the contribution from control and *TBX5^AKO* samples are shown. The black dotted line indicates the overall relative contribution of KO and control nuclei to the assay. Other dataset metrics are available in Extended Data Fig. 4. Adipo, adipocyte; CM, cardiomyocyte; EC, endothelial cell; Endo, endocardium; Epi, epicardium; Fibro, fibroblast; Peri, pericyte; Macro, macrophage.

DEGs in Tbx5-OE to Tbx5 inactivation *Tbx5^AKO* revealed a slight but significant anticorrelation ($r = -0.11$, $P = 0.009$). Thirty-five percent of the DEGs found in quadrant I, representing wild-type (WT) aCMs and vCMs overexpressing TBX5, were aCM selective, demonstrating the sufficiency of TBX5 to activate aCM gene expression across different chambers.
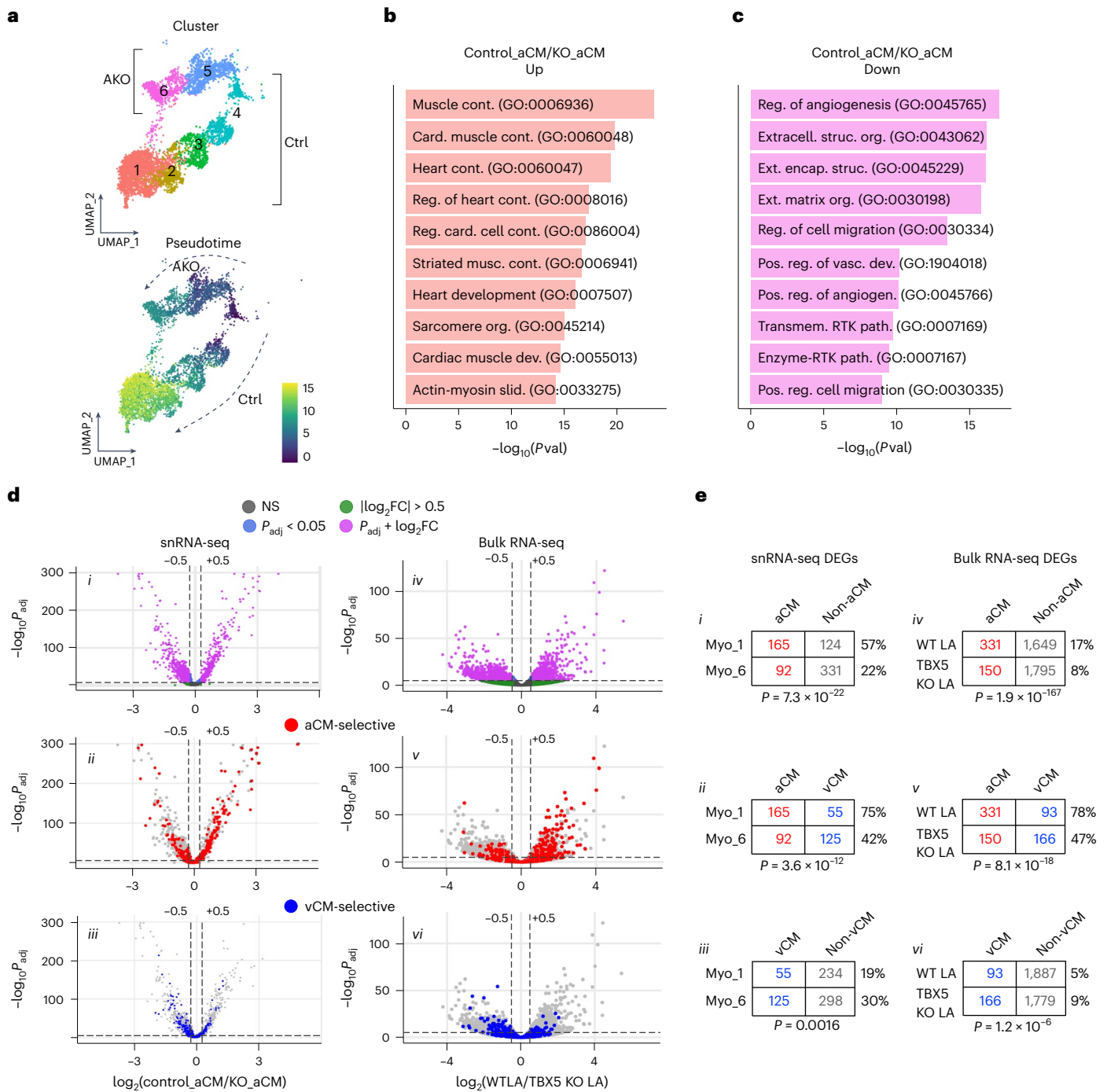
Fig. 4 | TBX5 is required to promote the expression of aCM genes.
a, Pseudotime trajectory of myocyte clusters. Trajectory analysis was performed with Monocle3. Clusters containing predominantly control (Ctrl) and knockout (AKO) myocytes participated in two separate inferred trajectories. **b,c**, Biological process GO terms enriched for genes with significantly greater expression in control (pink, Myocyte_1) or KO (magenta, Myocyte_6) clusters. The top 10 terms are shown. Fisher's exact test with Benjamini–Hochberg multiple testing correction. **d**, Distribution of chamber-selective genes among genes differentially expressed between control (Myo_1) or KO (Myo_6) myocyte clusters (left column) and from bulk RNA-seq experiment comparing WT and *Tbx5*[iKO] left atria (GSE129503; right column). aCM-selective genes (middle) were disproportionately found in genes with significant upregulation in control aCMs and WT left atria, and vCM-selective genes (bottom) were enriched in genes upregulated in KO aCMs or left atria. snRNA-seq statistics, Wilcoxon rank-sum test and Bonferroni correction. Bulk RNA-seq analysis, Wald test and Benjamini–Hochberg correction. **e**, Statistical analysis of distribution of aCM-selective or vCM-selective genes for the control versus *Tbx5*[AKO] aCMs (left) and control versus *Tbx5*[iKO] left atria (right) datasets. Fisher's exact test (two-tailed).

We further performed an orthogonal analysis of these four differential expression datasets to test the enrichment of aCM or vCM genes among genes differentially expressed by perturbation of *Tbx5*. For the 1,126 and 873 aCM-selective and vCM-selective genes, respectively, we calculated gene expression FC between the condition with higher *Tbx5* to that with lower *Tbx5* and ordered the genes by ascending FC in the *Tbx5*[iKO] dataset[17] (Extended Data Fig. 6g). This analysis highlighted the overall consistency of expression changes of these chamber-selective genes to altered *Tbx5* levels, particularly among LA datasets. Statistical analysis confirmed that aCM-selective genes are
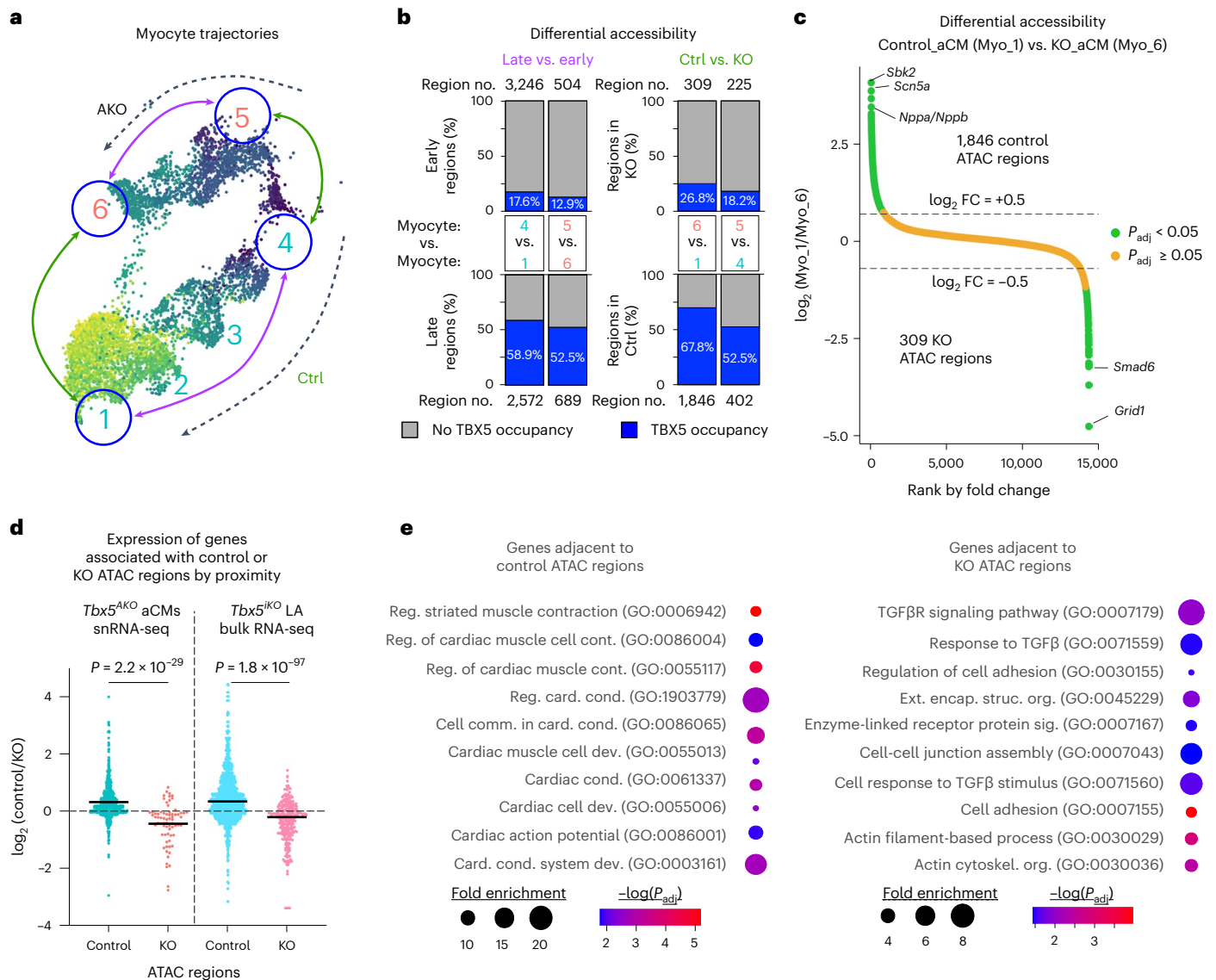
Fig. 5 | Multiomics reveals a set of *Tbx5*-dependent CREs that promote aCM gene expression. a, Myocyte trajectories and comparisons performed to identify differentially accessible genomic regions. Arrows indicate the four pairwise comparisons between the clusters at the start and end of the control and *Tbx5^{AKO}* trajectories. **b**, Differentially accessible regions were analyzed for the four of the pairwise comparisons illustrated in **a**. The number of differentially accessible regions (region number) is shown at the end of each bar. The bar indicates the fraction of these regions that are occupied by TBX5 in WT aCMs. **c**, Accessibility change between the control and TBX5 KO aCM clusters. Plotting the average $\log_2 FC$ accessibility of regions in the control aCM cluster compared to the KO aCM cluster revealed 1,846 and 309 significant control and KO regions, respectively (logistic regression model followed by comparison with a null model using a likelihood ratio test, Bonferroni corrected $P_{adj} < 0.05$). Selected genes nearest to top control and KO regions are listed. **d**, Expression of genes nearest to control or KO regions, from *Tbx5^{AKO}* aCM snRNA-seq cluster Myocyte_1 versus Myocyte_6 (left) and *Tbx5^{iKO}* LA bulk RNA-seq (right, *R26^{CreERt2}* ($n = 4$) versus *Tbx5^{Flox/Flox}*;*R26^{CreERt2}* ($n = 6$)) datasets. Two-sided Mann–Whitney test comparing control and KO ATAC regions. **e**, GO terms for genes nearest to control or KO regions. Genes next to control regions tended to be associated with cardiomyocyte function, whereas genes adjoining KO regions were associated with TGFβ signaling and remodeling terms (Fisher's exact test with Benjamini–Hochberg multiple testing correction).

enriched among *Tbx5*-dependent atrial genes and genes upregulated in ventricles by *Tbx5* overexpression (Extended Data Fig. 6g). On the other hand, vCM-selective genes were enriched among atrial genes upregulated in the condition with lower *Tbx5* expression. These results demonstrate that *Tbx5* activates expression of aCM-selective genes.

TBX5 maintains accessibility of atrial CREs

The interaction of TFs with distal *cis*-regulatory elements (CREs) regulates tissue-specific gene expression programs[37]. We hypothesized that TBX5 exerts its control over atrial-selective gene expression by binding and modulating the activity of atrial enhancers. To test this hypothesis, we compared chromatin accessibility between control and *Tbx5^{AKO}* aCM clusters using the multiome's snATAC-seq data (Fig. 5a). We quantitatively compared accessibility signals between clusters at the termini of the control and *Tbx5^{AKO}* trajectories to define differentially accessible regions ($P_{adj} < 0.05$; Fig. 5a,b and Supplementary Table 4). In the control trajectory, 5,818 regions were differentially accessible compared to only 1,193 regions that changed accessibility in the *Tbx5^{AKO}* trajectory. This result suggests that cell state changes in both control and KO involve substantial changes in chromatin accessibility and CRE usage, with many dependent on *Tbx5*.

Comparison between control and *Tbx5^{AKO}* clusters at early (Myocyte_4, control versus Myocyte_5, AKO) or late (Myocyte_1, control versus Myocyte_6, AKO) pseudotimes identified *Tbx5*-dependent

chromatin accessibility (Fig. 5b). We integrated these data with TBX5 genomic occupancy in WT aCMs, based on high-affinity streptavidin pulldown of an endogenously biotinylated *Tbx5* allele[3]. Of 1,846 regions that were more accessible in the control versus KO aCM clusters (Fig. 5b), 68% were occupied by TBX5 in control aCMs. In contrast, a minority of regions (27%; Fig. 5b) with greater accessibility in the KO cluster were TBX5-bound in control aCMs. Similarly, in control versus KO early pseudotime clusters (Myocyte_4 versus Myocyte_6; Fig. 5b), 402 regions with greater accessibility in control were predominantly TBX5-bound (53%), whereas regions with greater accessibility in KO were infrequently TBX5-bound (18%). Analysis of patterns of differential accessibility showed that the regions grouped into those with and those without greatest accessibility in the control aCM cluster. Those with greatest accessibility in the control aCM cluster had functional terms related to cardiac cell development and striated muscle contraction, and most of these regions were bound by TBX5 (Extended Data Fig. 7). Together, these data indicate that TBX5 binding is required to maintain accessibility of 1,462 regions in control atrial myocytes.

In subsequent analyses, we focused on the 1,846 and 309 regions with greater accessibility in control ('control ATAC regions') or *Tbx5^{AKO}* ('KO ATAC regions') late pseudotime clusters (Fig. 5c and Supplementary Table 5), respectively. We analyzed the contribution of these differentially accessible regions to gene expression. One method of associating regions to genes is by proximity—that is, pairing each region with its nearest gene. Some of the top differentially accessible control ATAC regions were nearest to aCM-selective genes *Sbk2, Nppa, Nppb, Scn5a, Myl7, Myl4* and *Bmp10* (Fig. 5c and Supplementary Table 5). Top differentially accessible KO ATAC regions neighbored *Grid1*, a glutamate receptor-related gene that is upregulated in aCMs of humans and dogs with AF[38], and *Smad6*, a regulator of TGF and BMP signaling that was upregulated in KO aCMs (Fig. 5c). Regions were also linked to genes by a region-to-gene linkage score based on the correlation between gene expression and region accessibility across nuclei (Extended Data Fig. 8a)[39]. Using either region-to-gene association method, genes associated with control ATAC regions were more highly expressed in control rather than KO aCMs, whereas genes near KO ATAC regions were more highly expressed in *Tbx5^{AKO}* aCMs than control (Fig. 5d and Extended Data Fig. 8b). These results indicate that control and KO ATAC regions act as CREs that govern expression of neighboring genes.

We next investigated the biological significance of the genes associated with control or KO ATAC regions. GO terms enriched for genes associated with control ATAC region involved the regulation of cardiac contraction, development and conduction (Fig. 5e). Furthermore, genes associated with control ATAC regions were highly enriched for aCM-selective genes, as 747 (40.4%) control ATAC regions were related to 404 (35.8%) aCM-selective genes (Extended Data Fig. 8c). In comparison, only 243 (13.1%) control ATAC regions were related to 42 (4.8%) vCM-selective genes. GO terms enriched for genes associated with KO ATAC regions involved TGFβ and structural terms (Fig. 5e). These results are congruent with the atrial remodeling that we observed in *Tbx5^{AKO}* mice, and elevated levels of TGFβ have been shown to result in atrial remodeling and fibrosis[40]. KO ATAC regions interacted infrequently with aCM-selective (3.6%) or vCM-selective (5.2%) genes, respectively (Extended Data Fig. 8d). These data suggest that TBX5-dependent accessible regions act as CREs that regulate aCM functional properties, including many aCM-selective genes.

To further characterize the differentially accessible chromatin regions, we measured their occupancy by active enhancer marker H3K27ac (Extended Data Fig. 9a,b). In WT aCMs, H3K27ac overlapped control ATAC regions, and signal was ~50% lower in *Tbx5^{AKO}*. In contrast, WT aCMs had relatively weak H3K27ac signal at KO ATAC regions, and signal at these regions was ~30% greater in *Tbx5^{AKO}* aCMs. These data suggest that at least a subset of control and KO ATAC regions act as enhancers in control and *Tbx5^{AKO}*, respectively.

We scanned control and KO ATAC regions for known TF binding motifs. Consistent with high TBX5 occupancy, the TBOX motif was the most significantly enriched motif in control ATAC regions, closely followed by the MEF2 motif (Extended Data Fig. 9c). Further analysis of ATAC signal at these motifs demonstrated increased accessibility around the motif and sharp loss of accessibility at the motif (Extended Data Fig. 9d). These TF footprints provide additional support for TF occupancy of these motifs. Control ATAC regions were also enriched for other previously identified cardiac TF motifs (Extended Data Fig. 9c). In contrast, KO ATAC regions were enriched for TF motifs not classically associated with cardiomyocytes, including the NFYA binding site, which was recently implicated in cardiomyocyte regeneration after injury (Extended Data Fig. 9e)[41].

We further evaluated cardiac TF binding to control and KO ATAC regions using aCM and vCM TF occupancy data based on bioChIP-seq for GATA4, MEF2A, MEF2C, NKX2-5, SRF and TEAD1, in addition to TBX5 and the co-activator P300 (Extended Data Fig. 9f). At aCM control ATAC regions, TBX5 was the most enriched factor, followed by NKX2-5, which co-binds DNA with TBX5 and is implicated in AF by genome-wide association study (GWAS)[42]. Although the MEF2 motif was the second-most enriched motif, its occupancy signal was relatively low compared to the other TFs, with MEF2A being consistently stronger than MEF2C. At control ATAC regions in vCMs, TF binding was also observed, although the occupancy signal of all factors was lower than in aCMs. At KO ATAC regions in both aCMs and vCMs, the TF occupancy signal (measured in WT aCMs) was lower than in control ATAC regions; TBX5 did not have the strongest occupancy signal; and no single factor predominated. These data suggest that TBX5 co-occupies control ATAC regions in aCMs with other key cardiac TFs, most notably NKX2-5.

Together, these data show that TBX5 binding positively regulates the accessibility of a large number of atrial enhancers, including enhancers that regulate genes important for aCM-selective gene expression and function.

Control ATAC regions participate in TBX5-dependent looping

Chromatin looping bring enhancers into physical proximity of promoters[43]. To evaluate the participation of control and KO ATAC regions in chromatin loops, we performed H3K27ac HiChIP to identify loops with this active enhancer mark[44]. HiChIP was performed in biological duplicate, each from 10 pairs of AAV9:*Nppa-EGFP Tbx5^{Flox/Flox}* or AAV9:*Nppa-Cre Tbx5^{AKO}* atria[45,46]. An average of 375 million read pairs per sample identified 516,995 loops, with 384,293 enhancer–enhancer loops, 122,034 enhancer–promoter loops and 10,668 promoter–promoter loops (Fig. 6a and Supplementary Table 1). At false discovery rate (FDR) < 0.1, we identified 263 loops that were stronger in control ('control loops') and 247 that were stronger in *Tbx5^{AKO}* ('KO loops') (Fig. 6b and Supplementary Table 6). These loops corresponded to 427 and 128 unique control and KO loop anchors, respectively. Strikingly, 178 of 427 (41.7%) unique control anchors, and 196 of 262 (74.8%) control loops, overlapped control ATAC regions (Fig. 6c). In contrast, only two KO loop anchors overlapped control ATAC regions, and only eight of 128 (6.3%) KO loop anchors, and five of 247 (2.01%) KO loops, overlapped KO ATAC regions.

We investigated the relationship of chromatin looping to gene expression. By proximity, genes were classified as associated with a control loop with control ATAC region, a control loop without control ATAC region, a control ATAC region that was not part of a control loop or without association. Genes associated with control loops involving a control ATAC region had the greatest overlap with aCM-selective genes, followed by genes associated with a control ATAC region not part of a control loop (Fig. 6d). Moreover, the ratio of gene expression between the control and *Tbx5^{AKO}* aCM clusters was significantly higher in genes associated with control loops and control ATAC regions, followed by genes associated with control ATAC regions but not loops (Fig. 6e). These data indicate that TBX5 is required to maintain enhancer
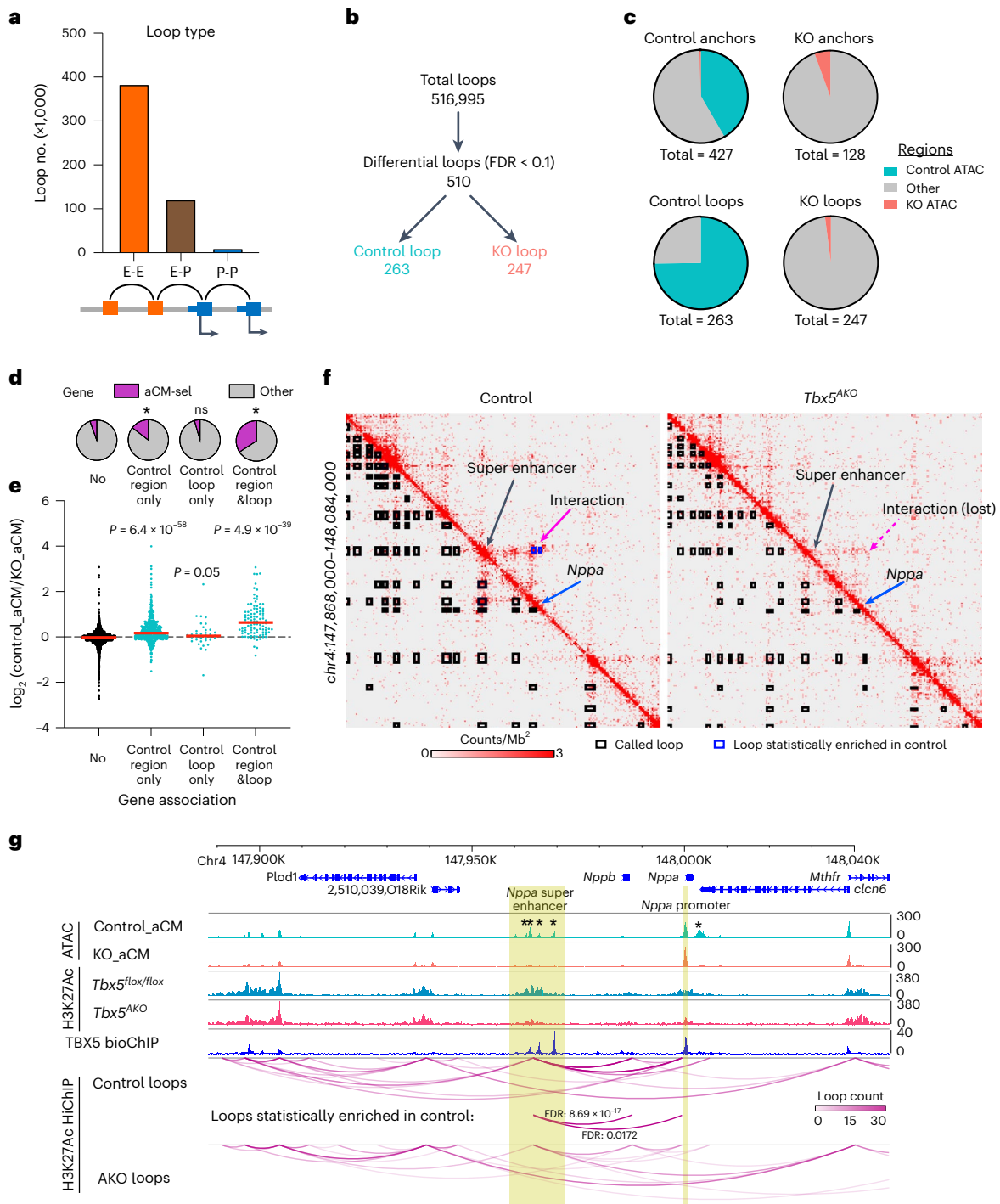
Fig. 6 | TBX5 maintains local chromatin structure to regulate gene expression. a, Loops identified by H3K27ac HiChIP. Loops were classified as enhancer–enhancer (E–E), enhancer–promoter (E–P) and promoter–promoter (P–P). **b**, Differential loops between control and *Tbx5^AKO* aCMs. Statistical analysis identified 510 differential loops with FDR < 0.1, 263 with greater loop score in control ('control loops') and 247 with greater loop score in *Tbx5^AKO* ('KO loops'). **c**, Overlap of differential loop anchors with control and KO ATAC regions. Top row (anchors) considers each loop anchor individually. Bottom row (loops) considers overlap of at least one anchor from each loop. **d**,**e**, Comparison of genes associated with control ATAC regions only, control loops only or both

control ATAC regions and control loops. **d**, Fraction of genes that are aCM selective. Proportion test versus 'no' association group. **e**, Gene expression ratio in control_aCMs (Myo_1) versus KO_aCMs (Myo_6). Kruskal–Wallis versus 'no' association group. **f**, Interaction maps for control and *Tbx5^AKO* samples near the *Nppa* locus. Significant loops are indicated by boxes on the lower-left part of the plot. Interactions significantly enriched in control samples are indicated by blue boxes in the upper right-hand region of the plot. **g**, Genome browser views of indicated chromatin features. TBX5 bioChIP-seq from aCMs is from GSE215065. Asterisks mark regions with significantly enriched accessibility in control_aCM (Myo_1) compared to KO_aCM (Myo_6).

accessibility and local three-dimensional (3D) chromatin conformation to promote expression of aCM genes, including aCM-selective genes. In agreement with this model, most genes near control loops were more

highly expressed in control compared to *Tbx5^AKO* aCM clusters, and many had aCM-selective expression (Extended Data Fig. 10b). Genes near KO loops were not enriched for chamber-selective expression and
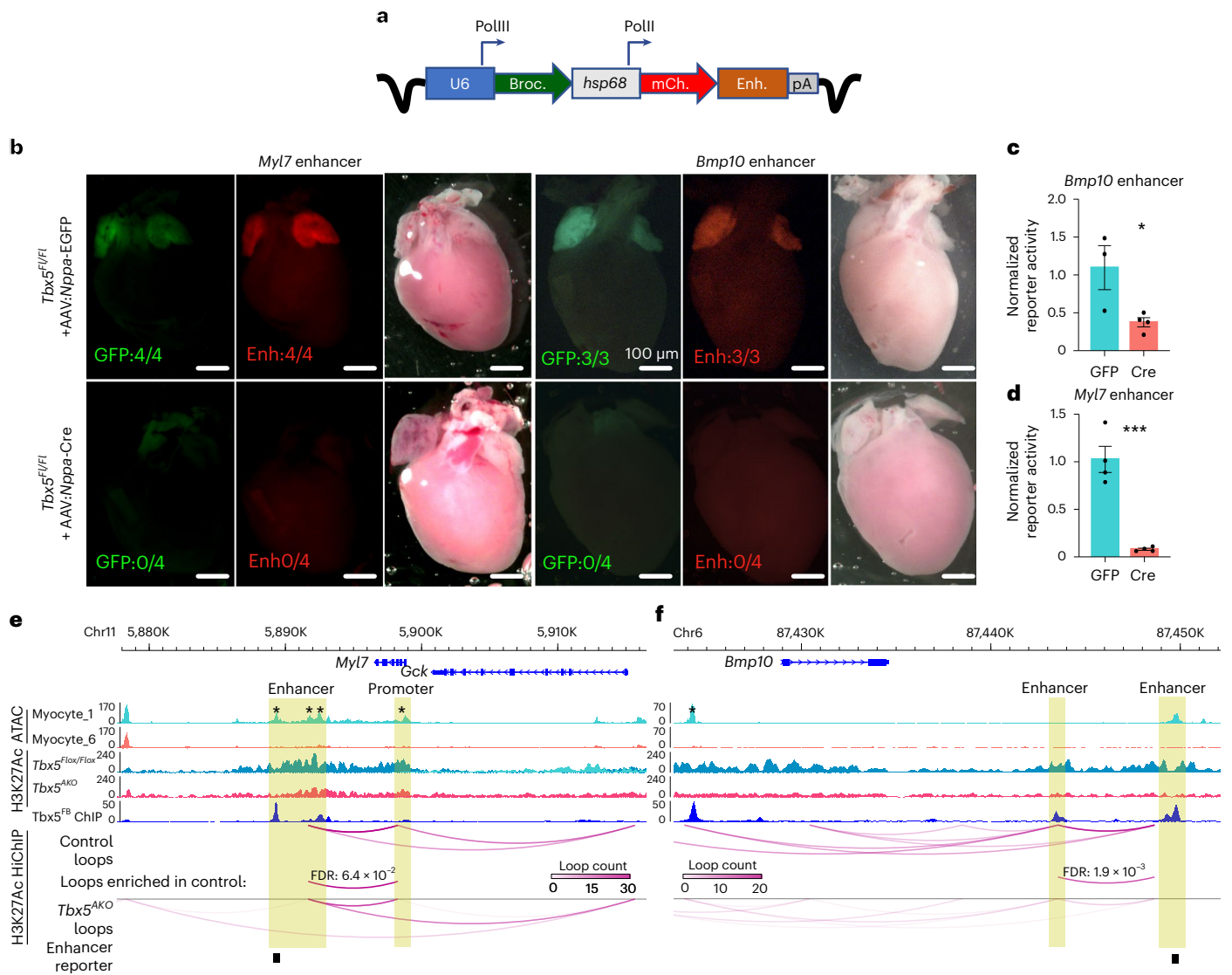
Fig. 7 | TBX5 regulates enhancer activity, accessibility and looping at aCM-selective genes *Myl7* and *Bmp10*. a, AAV enhancer reporter design. Enhancers of *Myl7* or *Bmp10* (highlighted and labeled in **e** and **f**) were cloned within the 3′ UTR of the mCherry reporter gene, downstream of a minimal *hsp68* promoter. The RNA polymerase III U6 promoter drove the expression of a *Broccoli* non-coding RNA, which was used to normalize for AAV transduction efficiency. **b**, Activity of *Myl7* and *Bmp10* enhancers require *Tbx5*. AAV9 containing either the *Myl7* or *Bmp10* enhancer was co-injected into P2 *Tbx5^Flox/Flox^* mice with either AAV9:*Nppa-EGFP* (top) or AAV9:*Nppa-Cre* (bottom). Hearts were visualized at P8.

Number of samples with positive signal and total number of samples is shown. **c,d**, Quantification of mCherry transcripts. *mCherry* and *Broccoli* transcript levels were measured by RT−qPCR from LA RNA. mCherry was normalized to Broccoli. Graphs show mean ± s.e.m. Unpaired two-sided *t*-test: *P = 0.0359 and ***P = 0.0005. *Bmp10*, n = 3 control and 4 *Tbx5^AKO^*; *Myl7*, n = 4 control and 4 *Tbx5^AKO^*. **e,f**, Genome browser views of the *Myl7* and *Bmp10*. aCM TBX5 bioChip-seq data are from GSE215065. Statistically significant differences in looping are diagrammed. Asterisks mark regions with significantly reduced accessibility in *Tbx5^AKO^* aCMs. Pol, polymerase; UTR, untranslated region.

tended to be more highly expressed in KO compared to control aCM clusters (Extended Data Fig. 10c).

We visualized looping differences at the *Nppa* locus (Fig. 6f,g), an aCM-selective gene that was downregulated in *Tbx5^AKO^* atria (Fig. 2e) and neighbors several control ATAC regions (Fig. 6g, asterisk). Chromatin contact maps[47] revealed similar overall organization of this locus in control and *Tbx5^AKO^* (Fig. 6f). However, *Tbx5^AKO^* had significantly less interaction of the *Nppa* and *Nppb* promoters with a neighboring, previously characterized super enhancer (SE) that regulates these genes[48] (Fig. 6f,g). Four control-enriched ATAC regions fell within the *Nppa/Nppb* SE and were bound by TBX5. In *Tbx5^AKO^*, these regions lost accessibility and participated in fewer chromatin loops. These data indicate that TBX5 regulates *Nppa* expression by binding the SE and promoting its accessibility and looping to the *Nppa* promoter. *Tbx5* ablation partially inactivated the SE in association with *Nppa* downregulation.

Although the connection of the SE to *Nppb* is weaker than the connection to *Nppa* in control aCMs, in KO aCMs, contact with the *Nppb* promoter is relatively stronger, consistent with maintenance of *Nppb* expression and *Nppa* downregulation (Fig. 6). This may be due to the shorter distance separating *Nppb* and the SE or by direct TBX5 binding to the *Nppa* but not the *Nppb* promoter, which could make *Nppa* more susceptible to *Tbx5* inactivation than *Nppb*.

To extend these observations to other atrial-selective genes, we tested TBX5's ability to regulate the atrial-specific enhancers of *Myl7* and *Bmp10* (ref. 3) (Fig. 7). These enhancers were sufficient to drive atrial-selective mCherry reporter gene expression within the context of AAV episomes. We tested these AAV enhancer reporters for their dependence on *Tbx5* by co-injecting them with either AAV9:*Nppa-EGFP* or AAV9:*Nppa-Cre* into neonatal *Tbx5^Flox/Flox^* mice (Fig. 7a). At P8, hearts were collected and visualized under a fluorescence dissecting

microscope. AAV9:*Nppa-EGFP*-treated hearts had EGFP⁺ atria (Fig. 7b). The *Myl7* and *Bmp10* enhancers drove selective *mCherry* expression in the atria. Although endogenous *Bmp10* is normally restricted to right atrium in postnatal mice[49], the *Bmp10* enhancer drove mCherry in both atria, demonstrating that additional regulatory information at the endogenous *Bmp10* locus is required to restrict its expression to only the right atrium. Ablation of *Tbx5* by co-injection with AAV9:*Nppa-Cre* resulted in loss of atrial mCherry signal, indicating that the *Myl7* and *Bmp10* enhancers require *Tbx5* for activity. We confirmed this finding by measuring mCherry transcripts in LA RNA by RT–qPCR, normalizing for transduction efficiency using a *Broccoli* non-coding RNA driven from U6, an RNA polymerase III promoter (Fig. 7c,d).

Chromatin features at these loci illustrated key elements of aCM-selective enhancers. Both of the tested *Myl7* and *Bmp10* enhancers were bound by TBX5 in aCMs (Fig. 7e,f). The *Myl7* enhancer's accessibility was TBX5 dependent (Fig. 7e, asterisk), and it contacted the *Myl7* promoter via a TBX5-dependent loop (Fig. 7e, 'Loops enriched in control', and Extended Data Fig. 10a). The *Bmp10* enhancer also had greater accessibility in the control compared to the KO aCM cluster, although this did not reach statistical significance (Fig. 7f). Furthermore, it contacted a second, TBX5-dependent accessible region through a chromatin loop (Fig. 7f, asterisk). The *Bmp10* enhancer also made a TBX5-dependent contact with an additional TBX5-bound region and with the *Bmp10* gene body (Fig. 7f and Extended Data Fig. 10a). These data show that TBX5 facilitates promoter–enhancer contacts of atrial genes by binding to and maintaining the accessibility of atrial enhancer elements.

TBX5 is required to maintain aCM identity

Our analysis indicates that *Tbx5* is required to maintain expression of aCM-selective genes, leading us to hypothesize that *Tbx5* is required to maintain aCM identity. At the transcriptome level, 'cell identity genes' have been defined in several different ways, including: (1) selective expression in a cell type; (2) characteristic chromatin features[50,51]; and (3) association with cell-type-selective SEs[37]. To test the hypothesis that *Tbx5* is required to maintain aCM identity, we defined 'aCM identity genes' using each of these approaches and evaluated the effect of aCM-specific *Tbx5* inactivation.

RNA-seq at P0 defined the aCM-selective and vCM-selective gene sets (Fig. 2g). We intersected these gene sets with RNA-seq from P28 aCMs (this study) and vCMs[52] to define 683 aCM-selective and 410 vCM-selective genes that are independent of developmental timepoint (Fig. 8a). When we used these chamber-selective gene lists to re-analyze the *Tbx5^{AKO}* and LA-*Tbx5^{iKO}* RNA-seq datasets, we again found that stage-independent aCM genes were enriched in control samples (Fig. 8b). However, stage-independent vCM genes were no longer enriched in KO populations compared to WT.

Machine learning approaches have been used to define cell identity genes (CIGs) based on their distinct chromatin features and gene expression[50]. We used an implementation of this strategy, Computational Epigenetic Framework for Cell Identity Gene Discovery (CEFCIG), to integrate aCM and vCM H3K27ac ChIP-seq (ref. 3) and RNA-seq (ref. 3) datasets and identify genes that are critical for aCM and vCM identity (Fig. 8c). CEFCIG aCM and vCM CIGs (FDR < 0.05) were compared to find 147 and 236 genes unique for atrial identity or ventricular identity, respectively. These included many of the chamber-selective

genes examined in previous experiments, such as *Myl4*, *Myl7*, *Bmp10* and *Tbx5* (aCMs) and *Myl2* and *Irx4* (vCMs). We analyzed the effect of aCM-specific *Tbx5* inactivation on expression of these atrial or ventricular identity genes. Atrial identity genes were highly and statistically enriched in control samples compared to KO samples (Fig. 8d). In contrast, ventricular identity genes were enriched in the KO Myocyte_6 population from *Tbx5^{AKO}* but not from the bulk RNA-seq of LA-*Tbx5^{iKO}*.

SEs have been shown to be important for cell lineage specification[37]. H3K27ac ChIP data[3] were used to identify SEs from aCMs and vCMs (Fig. 8e). The results revealed 954 SEs that were shared by vCMs and aCMs (shared SEs), 453 SEs specific to aCMs (atrial SEs) and 375 SEs specific to vCMs (ventricular SEs). Atrial SEs neighbored aCM-selective genes more frequently than vCM-selective genes (98 (22%) versus 12 (3%); Fisher's $P < 1 \times 10^{-5}$), and ventricular SEs neighbored vCM-selective genes more frequently than aCM-selective genes (72 (12%) versus 11 (3%); Fisher's $P < 1 \times 10^{-5}$). Shared SEs neighbored a similar number of aCM-selective and vCM-selective genes (90 aCMs, or 8% of aCM-selective genes; 52 vCMs, or 6% of vCM-selective genes). Therefore, each category of SE was enriched for the expected class of chamber-selective gene.

We next compared the overlap of SEs with the control and KO ATAC regions identified in Fig. 5c. Remarkably, 59% of the aCM SEs and 59% of the shared SEs overlapped with *Tbx5*-dependent control regions, whereas only 34% of vCM SEs overlapped with control regions, demonstrating the dependence of the atrial SE and shared SE networks on *Tbx5* (Fig. 8f). KO regions non-specifically overlapped with all three SE categories by approximately 13%. Genes regulated by atrial SEs were expressed at higher levels in control samples compared to KO samples in both the *Tbx5^{AKO}* and *Tbx5^{iKO}* datasets, whereas genes controlled by shared SEs and ventricular SEs were not (Fig. 8g).

Collectively, these analyses demonstrate that *Tbx5* promotes the expression of genes important for aCM identity.

Discussion

Differences between aCM and vCM gene expression programs imbue each cell type with distinct functional properties. Here we demonstrate that TBX5 is necessary to maintain the identity of postnatal aCMs. Postnatal TBX5 inactivation disrupted atrial SEs and downregulated genes important for atrial identity, morphology and function.

Transcriptional mechanisms maintain cell identity, and dysregulation of cell identity has been suggested to contribute to diverse pathological processes, such as cancer and fibrosis[53,54]. We used three separate approaches to define aCM identity genes and probe their regulation by *Tbx5*: (1) selective expression in aCMs compared to vCMs; (2) a validated machine learning model[50] that identifies identity genes based on H3K27ac chromatin features; and (3) genes linked to aCM-selective SEs, regulatory features that drive cell identity[37]. By all three approaches, *Tbx5* was required to maintain expression of aCM identity genes. On the other hand, *Tbx5* inactivation did not clearly lead to gain of vCM identity, although it did upregulate some vCM genes. This shows that loss of aCM identity does not necessarily lead to increased vCM identity.
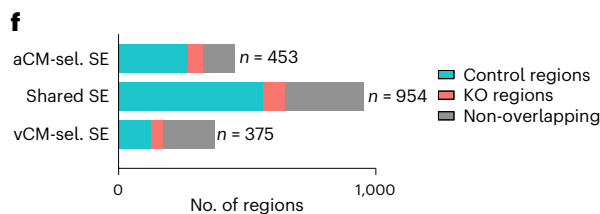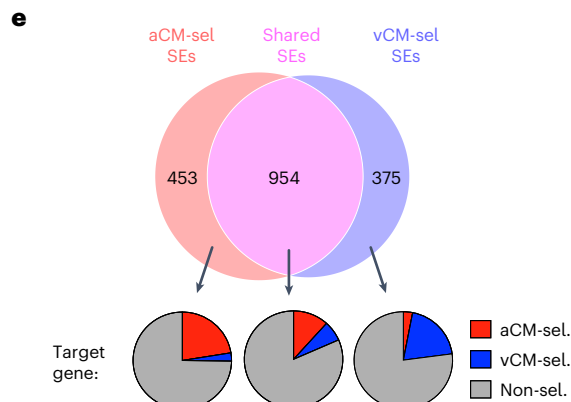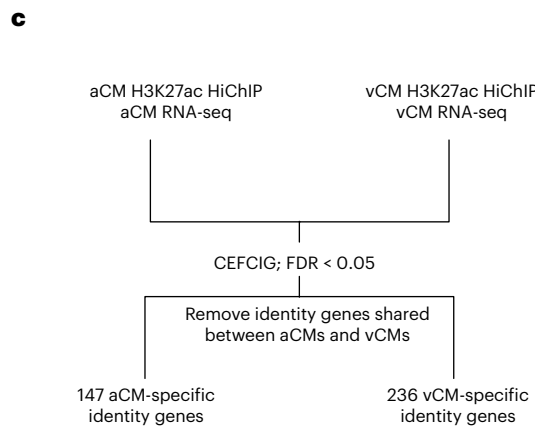
Our study identified candidate regulatory regions that require TBX5 to maintain accessibility of thousands of regions in control aCMs. These regions were decorated by the active enhancer mark H3K27ac, enriched for motifs of TBX5, MEF2 and other cardiac TFs, co-bound by

Fig. 8 | TBX5 promotes atrial identity. a, Identification of aCM-selective and vCM-selective genes shared between P0 and P28. RNA-seq from P0 and P28 aCMs and vCMs were compared to identify genes that were chamber selective at both stages. **b**, Enrichment of stage-independent aCM genes among genes significantly upregulated in control or KO in the *Tbx5^{AKO}* aCMs (left) and *Tbx5^{iKO}* left atria (right). **c**, aCM and vCM identity genes were identified by machine learning. Genes overlapping both cell types were removed, leaving a total of 147 atrial identity (AI) genes and 236 ventricular identity (VI) genes. **d**, Enrichment

of AI and VI genes among genes significantly upregulated in control or KO in the *Tbx5^{AKO}* aCMs (left) and *Tbx5^{iKO}* left atria (right). **e**, SEs were identified by H3K27ac signal in aCMs or vCMs. Chamber-selective SEs were enriched for expression of chamber-selective genes. **f**, Overlap of different classes of SEs with regions with greater accessibility in control (control regions) or *Tbx5^{AKO}* (KO regions) aCMs. **g**, Enrichment of genes neighboring aCM-selective and vCM-selective SEs among genes significantly upregulated in control or KO in the *Tbx5^{AKO}* aCMs (left) and *Tbx5^{iKO}* left atria (right). **b,d,g**: Fisher's exact test, two-tailed.

these TFs and linked to genes enriched for aCM-selective expression with important functional roles for cardiomyocyte development, structure and activity. In other words, these represent TBX5-regulated

enhancers through which TBX5 influences aCM identity and maintenance of normal cardiac rhythm. This inventory of TBX5-dependent aCM enhancers will be invaluable for future studies of atrial gene

regulation in homeostasis and disease. Consistent with our results, a recent study demonstrated that introduction into mice of a human disease-causing *Tbx5* variant, *Tbx5^G125R^*, that increases DNA binding to TBX5 motifs[16], increased accessibility of thousands of regions occupied by WT TBX5, with associated changes in gene expression and atrial electrophysiology. One mechanism by which TBX5 may maintain accessibility is by recruiting chromatin remodeling complexes. TBX5 directly interacts with CHD4 and SMARCD3, subunits of chromatin remodeling complexes, to facilitate heart development[55,56]. In postnatal aCMs, TBX5 recruitment of these chromatin remodeling complexes may also be required to maintain chromatin accessibility and aCM cardiac gene expression. Further experiments are required to test this hypothesis.

Enhancers form loops to contact promoters and drive gene expression[57]. H3K27ac HiChIP (ref. 44) measurements revealed changes in enhancer–promoter looping that occurred after *Tbx5* inactivation in aCMs. Remarkably, the large majority of loops that were weakened by *Tbx5* inactivation were anchored on regions that required TBX5 to maintain accessibility. This result suggests that, in addition to maintaining accessibility, TBX5 also maintains the 3D genome structure of a subset of its target enhancers. Because the affected enhancers are enriched for aCM-selective genes, including *Nppa*, *Myl7* and *Bmp10*, this result further implies that TBX5-dependent maintenance of 3D genome structure at key loci is integral to maintenance of aCM identity.

Tbx5 is essential for maintenance of atrial rhythm[17–19]. Consistent with previous studies of global postnatal *Tbx5* inactivation and with GWASs that implicate *TBX5* variants in human AF[58], our data show that *Tbx5* inactivation selectively in aCMs rapidly results in AF. Previous studies focused on *Tbx5* regulation of Ca^{2+} handling genes as a mechanism underlying AF in *Tbx5*-depleted hearts[18,19], which we also observed in our study. In addition to these Ca^{2+} handling genes, we observed dysregulation of many aCM-selective genes, suggesting that altered expression of these signature atrial genes may contribute to development of AF. Indeed, *Tbx5* inactivation led to both AF and loss of aCM identity, suggesting that impaired maintenance of aCM transcriptional identity participates in AF pathogenesis.

Although this study demonstrates an essential role for TBX5 in maintenance of atrial identity of postnatal aCMs, it does not discount roles of TBX5 in other contexts. Germline inactivation of *Tbx5* disrupts development of both atrial and ventricular chambers[15], and *Tbx5* is an essential component of the cocktail of TFs that reprogram fibroblasts into cardiomyocyte-like cells[59]. In these contexts, TBX5 is required to promote cardiomyocyte differentiation and proliferation. TBX5 is also expressed in the postnatal vCMs, albeit at lower levels than in aCMs. Overexpression of TBX5 in vCMs was sufficient to promote the expression of aCM genes, suggesting that TBX5 chamber-selective function is dose dependent. This dosage sensitivity may contribute to human phenotypes observed with TBX5 variants that increase or decrease expression or activity[16,20,60]. Further studies are required to elucidate the molecular mechanisms underlying these dosage-dependent and context-dependent effects of TBX5.

Methods

Mice
All research conducted using animals was done following protocols that were approved by institutional animal care and use committees at Boston Children's Hospital, Harvard Medical School or the University of Chicago. *Tbx5^Flox^* (ref. 15), Tbx5-OE (ref. 27), HCN4-CreERT2 (ref. 61), Rosa26^mTmG^ (ref. 62), Rosa26-H2B-mCh (ref. 63) and Myh6-Cre (ref. 64) mice were described previously. AAVs were injected subcutaneously at the indicated stage. Mice strains were maintained on a mixed background. Approximately equal numbers of male and female mice were used in each experiment. The number of mice used in each experiment is reported in the figure legend.

Echocardiography
Echocardiography was performed using a Vevo 3100 imaging system (Visual Sonics) with awake animals in a standard hand grip. Cardiac function was recorded from the parasternal short-axis view. Atrial size was determined as the ratio of the left atrium to aortic diameter. The echocardiographer was blinded to the genotype and treatment of each of the experimental mice.

Surface and intracardiac EKGs
Mice were anesthetized with 3% isoflurane and placed on a heating pad (37 °C) for the remainder of the experiment. Mice were positioned in dorsal recumbency, and platinum electrodes were placed under the skin of each of the four limbs. Three minutes of recording were collected using the iWorx IX-ECG6 EKG recorder and LabScribe acquisition and analysis software. In total, 1,500 beats per mouse were analyzed to calculate the SDRR interval or to generate a Poincaré plot. Recordings were performed using LabScribe 3 software, and RR intervals were exported using the same software. SDRR was calculated in Excel. Poincaré plots were generated in R using ggplot2. EKGs were acquired blinded to treatment groups. For intracardiac EKGs, we performed a right jugular vein cutdown and inserted a 1.1 Fr octapolar catheter (ADInstruments) so that leads were present in the right atrium and right ventricle. Surface echocardiography and intracardiac signals were digitized at 4 kS s^{-1} and recorded (PowerLab software, ADInstruments).

AAV purification
AAVs were purified and quantified following a standard published protocol[65]. In brief, AAV was produced in HEK293T cells (American Type Culture Collection, CRL-3216) using AAV9 Rep/Cap. Viruses were purified by ionoxidal density gradient ultracentrifugation as described previously[65]. AAV concentration was measured by qPCR. In experiments with HCN4-CreERT2 mice, AAV was injected into the mediastinum via a subxiphoid approach[66]. For all other experiments, AAV (2×10^{11} viral genomes per gram bodyweight (VG/g)) was injected subcutaneously to neonatal mice.

Immunostaining
Tissues were fixed overnight at 4 °C with rotation in 4% paraformaldehyde (PFA). For cryosections, tissues were then placed in 30% sucrose solution until they sank (usually about 4 h), positioned in O.C.T. and frozen. Then, 5-µm sections were attached to Superfrost Plus slides. Sections were permeabilized for 20 min using 0.1% Triton X, washed and blocked with 20% donkey serum for 1 h. Primary antibodies diluted in PBS were incubated on top of tissue sections overnight at 4 °C. Slides were washed 3× with PBS and counterstained with the appropriate Alexa Fluor-conjugated secondary antibody and Alexa Fluor 647-conjugated wheat germ agglutinin (WGA) (1:500) for 1 h. Slides were washed 3×, mounted with coverslip solution containing DAPI and sealed with nail polish. Information about the antibodies used in the study is available in Supplementary Table 7.

Microscopy
Confocal images were acquired using an Olympus FV3000RS confocal microscope using a ×60 oil immersion lens (numerical aperture (NA) = 1.4). To measure fibrosis, mouse hearts were excised and fixed in 4% PFA overnight, dehydrated through an ethanol gradient, embedded in paraffin and sectioned at 7 µm. Sections were dewaxed, rehydrated, stained and subjected to Masson trichrome staining for fibrosis visualization. Sections were imaged by a wide-field microscope (Keyence) at ×10 magnification or a laser scanning confocal microscope (Olympus FV3000RS).

RT–qPCR
Total RNA was isolated using TRIzol extraction followed by purification using the RNA Clean & Concentrator Kit (R1017) and an on-column

genomic DNA digest step (Qiagen, 79254). Reverse transcription was performed on 500 ng of total RNA using Takara PrimeScript (RR037B). RT–qPCR was performed using a Bio-Rad CFX96 or CFX384 real-time PCR instrument with SYBR Green detection chemistry. Gene expression values were normalized to GAPDH expression. PCR primer sequences: *Myl2-f* 5′ CCCTAGGACGAGTGAACGTG, *Myl2-R* 5′ TCCCGGACATAGTCAGCCTT, *Nppa-F* 5′ GCTTCCAGGCCATATTGGAG, *Nppa-R* 5′ GGGGGCATGACCTCATCTT, *Nppb-F* 5′ GAGGTCACTCCTATCCTCTGG, *Nppb-R* 5′ GCCATTTCCTCCGACTTTTCTC, *Myl7-F* 5′ GGCACAACGTGGCTCTTCTAA, *Myl7-R* 5′ TGCAGATGATCCCATCCCTGT, *mCherry-F* 5′ ACGGCCACGAGTTTGAGATT, *mCherry-R* 5′ CAAGTAGTCGGGGATGTCGG, *Broccoli-F* 5′ TATTCGTATCTGTCGAGTAGAGT, *Broccoli-R* 5′ GATCATCAGAGTATGTGGGAG, *GAPDH-F* 5′ AGGTCGGTGTGAACGGATTTG and *GAPDH-R* 5′ TGTAGACCATGTAGTTGAGGTCA. *Tbx5-F* 5′ GGCATGGAAGGAATCAAGGT, *Tbx5-R* 5′ CTAGGAAACATTCTCCTCCCTGC, *Myl4-F* 5′ ATGCCTCCCAAGAAACCCG and *Myl4-R* 5′ GTAGGTGATCTTCATCTCTCCCG.

Western blotting

Atrial and ventricular lysates were prepared by homogenizing samples in ice-cold RIPA buffer containing protease inhibitor. Protein concentrations for the different samples were determined using Pierce BCA Protein Assay Kit from Thermo Fisher Scientific. An equal amount of protein (20 μg) was loaded into each well of an SDS-PAGE gel. Gels were transferred to PVDF membranes, which were blocked with 5% milk and incubated with primary antibodies overnight at 4 °C (Supplementary Table 7). Secondary antibodies conjugated to HRP were incubated the following day for 1 h. The membranes were then developed and imaged using the ImageQuant LAS 4000 Luminescent Image Analyzer.

Concurrent snRNA-seq and snATAC-seq

LA samples were flash frozen in liquid nitrogen and preserved at −80 °C before the experiment. To minimize experimental differences, all samples were processed on the same day, and Gel Bead-In EMulsions were encapsulated in the same run. Nuclei were isolated following a published protocol with minor modifications[67]. In brief, whole frozen atria were resuspended in homogenization buffer (250 mM sucrose, 25 mM KCl, 5 mM MgCl₂, 10 mM Tris, pH 8, 1 μM DTT, with added protease inhibitor, 0.4 U μl⁻¹ RNaseIn, 0.2 U μl⁻¹ SUPERase-In and 0.1% Triton X-100). Samples were then homogenized in a Qiagen TissueLyser II with a 1-mm ball bearing set at 25 Hz for 90 s. Samples were filtered using a pluriSelect Mini Strainer (40 μm), centrifuged for 5 min at 500g and resuspended in storage buffer (4% BSA in PBS, 0.2 U μl⁻¹ RNaseIn). Samples were stained using 7-aminoactinomycin D (7-AAD) at a final concentration of 1 μg ml⁻¹ and FACS sorted for purification. For each multiome sample, 300,000 nuclei were sorted into storage buffer. Nuclei were then gently permeabilized using multiome lysis buffer (Tris-HCl, pH 7.4, 10 mM, NaCl 10 mM, MgCl₂ 3 mM, Tween-20 0.1%, NP40 0.1%, digitonin 0.01%, BSA 1%, DTT 1 mM, RNaseIN 1 U μl⁻¹), washed 2× with wash buffer (Tris-HCl, pH 7.4, 10 mM, NaCl 10 mM, MgCl₂ 3 mM, BSA 1%, Tween-20 0.1%, DTT 1 mM, RNaseIN 1 U μl⁻¹) and resuspended in 1× nuclei buffer (10x Genomics). Nuclei were quantified using the Countess 3 system and then processed using the 10x Genomics Next GEM Single Cell Multiome ATAC + Gene Expression Reagents Kit. The resulting joint snRNA and snATAC libraries were analyzed by TapeStation (Agilent) and then sequenced on an Illumina NovaSeq instrument.

Multiome analysis

Outputs from the 10x CellRanger-ARC software package were processed using the Seurat[30] and Signac[28] packages. In brief, CellRanger-ARC was run for each sample, and the output files were used to construct Seurat objects in R. A shared peak set derived from the integration of peaks identified for each sample was used to analyze the snATAC dataset, according to the merging Signac vignette. Each Seurat

object was quality controlled using the same metrics: 250 < nCount_ATAC < 100,000; 250 < nCount_RNA < 35,000; nucleosome_signal < 2; TSS.enrichment > 1; and percent mitochondrial reads < 25. Next, doublets were identified and removed using Doubletfinder[68] on the RNA portion of the assay. Finally, the objects were merged together. Pre-processing and dimensional reduction were performed on the RNA and ATAC assays independently, using standard approaches for RNA and ATAC-seq data[69], as follows. Gene expression unique molecular identifier (UMI) count data were normalized using SCTransform, and the 20 k-nearest neighbors (KNNs, $k = 20$) for each nucleus were found using the FindNeighbors function. Latent semantic indexing (LSI) was used to perform dimension reduction on the DNA accessibility assay dataset, and graph-based clustering on the LSI components 2:15 was performed by computing a nearest neighbor graph using LSI low-dimensional space ($k = 20$ neighbors) and then applying the Smart Local Moving algorithm for community detection. A WNN graph was generated using the weighted combination of the RNA and ATAC modalities. All downstream analysis were performed using features of Seurat/Signac or using Seurat wrappers for other packages. Trajectory analysis was performed using the Seurat wrapper for Monocle3 (ref. 31). Differential gene expression and differential genomic accessibility analyses were performed comparing clusters using the FindMarkers function on either the RNA or the ATAC assays. Differential accessibility was performed by logistic regression using Signac as described[28], with thresholds of |log₂FC| > 0.5 and P_{adj} < 0.05. The same thresholds were used to identify DEGs. Differential gene expression analysis was performed using the Wilcoxon test. Genes neighboring enhancer regions were identified using the ClosestFeature command in Signac. To generate bigWig files from the ATAC data for individual clusters, fragment files for each of the samples were split using the list of cell barcodes composing the cluster of interest using the sinto package. The bamcoverage tool was used to create reads per genome coverage (RPGC) normalized bigWig files.

Gene set analysis

We defined aCM-selective and vCM-selective genes based on bulk RNA-seq of P0 aCMs versus P0 vCMs (ref. 3). Each set of selective genes was defined by log₂FC > 0.58, P_{adj} < 0.05 and transcripts per million (TPM) ≥ 5. Skewed distribution of a set of genes among a list of genes rank ordered by relative expression in two conditions was analyzed using gene set enrichment analysis (GSEA)[70]. For *Tbx5-OE* and *Tbx5iKO* analyses, statistical significance was assessed by randomly permuting sample labels. For *Tbx5AKO* (snRNA-seq) and *Tbx5RE(int)KO*, GSEA-preranked was used with signed *P* value as the ranking metric, and permutation was performed on gene sets. Enrichment of a list of genes in predefined gene pathways ('GO term enrichment analysis') was performed using the DE and enrichR pathway visualization tool through Seurat, with default settings[71]. Enrichment of genes linked to regions in predefined gene pathways was performed using GREAT[72].

Bulk RNA-seq

Paired-end reads were aligned to the mm10 genome with STAR[73] version 2.5.3b using the parameter −quantMode GeneCounts to retrieve counts at exons. Genes were tested for differential expression with DESeq2 with default settings. Genes differentially expressed between Tbx5-OE and control ventricles were those with P_{adj} < 0.05 and |log₂FC| > 0.5.

H3K27ac HiChIP

HiChIP for H3K27ac was performed using the Arima-HiC+ Kit for HiChIP, and the library was prepared using the Swift Biosciences Accel-NGS 2S Plus DNA Library Kit following the manufacturer's protocol. HiChIP samples (10 left and right atria per biological replicate) were collected and snap frozen in liquid nitrogen. Nuclei were isolated using the same approach as the multiomics experiment. After homogenization in the TissueLyser II, nuclei were pelleted and resuspended in PBS.

Nuclei from aCMs were enriched by purification by MACS for PCM1, as previously described[74]. Formaldehyde was added to the nuclei for a final concentration of 2%, and the nuclei were fixed with gentle rocking at room temperature for 15 min. The reaction was quenched with glycine, and the nuclei were pelleted and counted. In total, each HiChIP reaction required ~3 million nuclei. The fixed nuclei were pelleted and snap frozen and stored at −80 °C until used in the HiChIP protocol. Libraries resulting from the HiChIP protocol were sequenced to a depth of 300 million reads per library.

H3K27ac ChIP signal was obtained by ChIP-seq analysis of the HiChIP data.

Loop calling and differential loop identification

FASTQ files from the HiChIP experiment were used to generate normalized contact matrices with HiC-PRO[45]. Then, the command line tool hichipper[46], a restriction enzyme site aware package that reduces the bias caused by the digest step of HiChIP, was used to identify loop anchors and to call loops, producing .mango interaction files that contain information on all significant contacts and anchor regions[75]. To determine if loops were significantly enriched in control or KO samples, interaction .mango files for each biological replicate were loaded into the R package diffloop[76]. To determine significance, the default edgeR statistical test used by diffloop was applied, where counts are modeled with negative binomial distribution, and the empirical Bayes procedure is used to mediate overdispersion. Loops that were called in one biological replicate more than five times but not the other, and interactions <5 kb in length, were excluded. Significant loop counts were multiplied by a size factor to normalize for sequencing depth, and mean values less than 5 were filtered out. Interaction types were determined by annotating the genome. Regions (±1 kb) from the TSS of mm10 genes were labeled as promoters, and other peaks called by MACS2 for each of the samples immunoprecipitated for the enhancer mark H3K27ac were defined as enhancer regions.

Generating contact maps of HiChIP data

To construct contact maps and bigWig files, FASTQ reads were aligned to the mm10 genome using BWA[77], and ligation junctions were identified with pairtools[78]. PCR duplicates were removed and .bam and .pairs files were then generated with pairtools. The resulting mapped pairs file was used as the input to generate a .hic file for each sample with juicer tools. Juicebox was used to visualize contact matrices[79]. Loops were identified on the resulting contact matrices by loading the .mango file output from hichipper from a given sample into Juicebox. Resulting .bam files were used as input with the bamCoverage tool to generate bigWig files, which were normalized using 'RPGC' to the effective mm10 genome size. Heat maps and profiles were generated from .bam files of control and KO ATAC regions using deeptools.

Systematic identification of super enhancers using the DANPOS algorithm

The peaks corresponding to H3K27ac in control and KO samples were identified using the DANPOS dregion algorithm[80]. Subsequently, for peaks associated with protein-coding regions, we computed their width by subtracting the start position from the end position. These width values were then used to rank the peaks, and the top 2,000 peaks were designated as SEs. SEs were identified from H3K27ac data from aCMs and vCMs[3], and SEs common to three biological replicates, identified by BEDTools intersect, were maintained for further analysis. To identify shared SEs, atrial SEs and ventricular SEs, the lists of SEs obtained from aCMs and vCMs were compared with each other.

Identification of cell identity genes using the CEFCIG algorithm

The CEFCIG algorithm includes the CIGdiscover module, designed to identify CIGs by integrating histone modification and RNA expression levels[50]. Leveraging this finding, we employed H3K27ac features, such as peak width within the gene body and peak kurtosis in the TSS region, as inputs for the CIGdiscover algorithm. The pipeline involved quantile normalization of the aligned '.bam' file of H3K27ac, followed by conversion into a '.wig' file using the DANPOS pipeline. The CIGdiscover module was used to identify CIGs in aCMs and vCMs by H3K27ac signatures and RNA expression levels. The significant CIGs for aCMs and vCMs were extracted by FDR cutoff of <0.05. CIG lists were subsetted to contain only unique aCM and vCM identity genes.

Statistics and reproducibility

Results are expressed as mean ± s.e.m. Statistical tests, indicated in figure legends, were performed using GraphPad Prism 9 or R. $P < 0.05$ was used as the statistical threshold for significance. Micrographs are representative of three independent experiments (unless indicated otherwise). Data points in graphs represent unique biological replicates.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

High-throughput data used in this manuscript are available from the Gene Expression Omnibus, accession number GSE222970. Other datasets used were obtained from GSE215065 (ref. 3), GSE129503 (ref. 17) and GSE195905 (ref. 52). All other data supporting the findings in this study are included in the main article and associated files. Source data are provided with this manuscript.

Code availability

Analysis was performed using standard R packages. No custom analysis software was created to perform these analyses. Analyses were performed using published R packages that are cited within the text and the Methods section.

References

1. Mulder, B. J. M. & van der Wall, E. E. Size and function of the atria. *Int. J. Cardiovasc. Imaging* **24**, 713–716 (2008).
2. Chung Mina, K. et al. Atrial fibrillation. *J. Am. Coll. Cardiol.* **75**, 1689–1713 (2020).
3. Cao, Y. et al. In vivo dissection of chamber selective enhancers reveals estrogen-related receptor as a regulator of ventricular cardiomyocyte identity. *Circulation.* **147**, 881–896 (2023).
4. Funakoshi, S. et al. Generation of mature compact ventricular cardiomyocytes from human pluripotent stem cells. *Nat. Commun.* **12**, 3155 (2021).
5. Devalla, H. D. et al. Atrial-like cardiomyocytes from human pluripotent stem cells are a robust preclinical model for assessing atrial-selective pharmacology. *EMBO Mol. Med.* **7**, 394–410 (2015).
6. Pradhan, A. et al. FGF signaling enforces cardiac chamber identity in the developing ventricle. *Development* **144**, 1328–1338 (2017).
7. Targoff, K. L. et al. Nkx genes are essential for maintenance of ventricular identity. *Development* **140**, 4203–4213 (2013).
8. Bao, Z. Z., Bruneau, B. G., Seidman, J. G., Seidman, C. E. & Cepko, C. L. Regulation of chamber-specific gene expression in the developing heart by *Irx4*. *Science* **283**, 1161–1164 (1999).
9. Bruneau, B. G. et al. Cardiomyopathy in *Irx4*-deficient mice is preceded by abnormal ventricular gene expression. *Mol. Cell. Biol.* **21**, 1730–1736 (2001).
10. Koibuchi, N. & Chin, M. T. CHF1/Hey2 plays a pivotal role in left ventricular maturation through suppression of ectopic atrial gene expression. *Circ. Res.* **100**, 850–855 (2007).

11. Xin, M. et al. Essential roles of the bHLH transcription factor Hrt2 in repression of atrial gene expression and maintenance of postnatal cardiac function. *Proc. Natl. Acad. Sci. USA.* **104**, 7975–7980 (2007).

12. Wu, S.-P. et al. Atrial identity is determined by a COUP-TFII regulatory network. *Dev. Cell* **25**, 417–426 (2013).

13. Niederreither, K. et al. Embryonic retinoic acid synthesis is essential for heart morphogenesis in the mouse. *Development* **128**, 1019–1031 (2001).

14. Lee, J. H., Protze, S. I., Laksman, Z., Backx, P. H. & Keller, G. M. Human pluripotent stem cell-derived atrial and ventricular cardiomyocytes develop from distinct mesoderm populations. *Cell Stem Cell* **21**, 179–194 (2017).

15. Bruneau, B. G. et al. A murine model of Holt–Oram syndrome defines roles of the T-box transcription factor Tbx5 in cardiogenesis and disease. *Cell* **106**, 709–721 (2001).

16. van Ouwerkerk, A. F. et al. Patient-specific TBX5-G125R variant induces profound transcriptional deregulation and atrial dysfunction. *Circulation* **145**, 606–619 (2022).

17. Nadadur, R. D. et al. *Pitx2* modulates a Tbx5-dependent gene regulatory network to maintain atrial rhythm. *Sci. Transl. Med.* **8**, 354ra115 (2016).

18. Dai, W. et al. A calcium transport mechanism for atrial fibrillation in Tbx5-mutant mice. *eLife* **8**, e41814 (2019).

19. Laforest, B. et al. Atrial fibrillation risk loci interact to modulate Ca^{2+}-dependent atrial rhythm homeostasis. *J. Clin. Invest.* **129**, 4937–4950 (2019).

20. Kathiriya, I. S. et al. Modeling human TBX5 haploinsufficiency predicts regulatory networks for congenital heart disease. *Dev. Cell* **56**, 292–309 (2021).

21. Steimle, J. D. & Moskowitz, I. P. TBX5: a key regulator of heart development. In *Current Topics in Developmental Biology* (ed Frasch, M.) Vol. 122, 195–221 (Academic Press, 2017).

22. Ni, L. et al. Atrial-specific gene delivery using an adeno-associated viral vector. *Circ. Res.* **124**, 256–262 (2019).

23. Hulsurkar, M. M. et al. Atrial-specific LKB1 knockdown represents a novel mouse model of atrial cardiomyopathy with spontaneous atrial fibrillation. *Circulation* **144**, 909–912 (2021).

24. Horsthuis, T. et al. Distinct regulation of developmental and heart disease-induced atrial natriuretic factor expression by two separate distal sequences. *Circ. Res.* **102**, 849–859 (2008).

25. Lu, F. et al. CMYA5 establishes cardiac dyad architecture and positioning. *Nat. Commun.* **13**, 2185 (2022).

26. Takeuchi, J. K. et al. Chromatin remodelling complex dosage modulates transcription factor function in heart development. *Nat. Commun.* **2**, 187 (2011).

27. Georges, R., Nemer, G., Morin, M., Lefebvre, C. & Nemer, M. Distinct expression and function of alternatively spliced Tbx5 isoforms in cell growth and differentiation. *Mol. Cell. Biol.* **28**, 4052–4067 (2008).

28. Stuart, T., Srivastava, A., Madad, S., Lareau, C. A. & Satija, R. Single-cell chromatin state analysis with Signac. *Nat. Methods* **18**, 1333–1341 (2021).

29. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587 (2021).

30. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).

31. Trapnell, C. et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).

32. Yang, X. H. et al. Transcription-factor-dependent enhancer transcription defines a gene regulatory network for cardiac rhythm. *Elife* **6**, e31683 (2017).

33. Arnolds, D. E. et al. TBX5 drives *Scn5a* expression to regulate cardiac conduction system function. *J. Clin. Invest.* **122**, 2509–2518 (2012).

34. van Gorp, P. R. R. et al. Sbk2, a newly discovered atrium-enriched regulator of sarcomere integrity. *Circ. Res.* **131**, 24–41 (2022).

35. Ackerman, M. J. & Mohler, P. J. Defining a new paradigm for human arrhythmia syndromes: phenotypic manifestations of gene mutations in ion channel- and transporter-associated proteins. *Circ. Res.* **107**, 457–465 (2010).

36. Bosada, F. M. et al. An atrial fibrillation-associated regulatory region modulates cardiac Tbx5 levels and arrhythmia susceptibility. *eLife* **12**, e80317 (2023).

37. Hnisz, D. et al. Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013).

38. Leblanc, F. J. A. et al. Transcriptomic profiling of canine atrial fibrillation models after one week of sustained arrhythmia. *Circ. Arrhythm. Electrophysiol.* **14**, e009887 (2021).

39. Ma, S. et al. Chromatin potential identified by shared single-cell profiling of RNA and chromatin. *Cell* **183**, 1103–1116 (2020).

40. Nakajima, H. et al. Atrial but not ventricular fibrosis in mice expressing a mutant transforming growth factor-β_1 transgene in the heart. *Circ. Res.* **86**, 571–579 (2000).

41. Cui, M. et al. Dynamic transcriptional responses to injury of regenerative and non-regenerative cardiomyocytes revealed by single-nucleus RNA sequencing. *Dev. Cell* **53**, 102–116 (2020).

42. Roselli, C. et al. Multi-ethnic genome-wide association study for atrial fibrillation. *Nat. Genet.* **50**, 1225–1233 (2018).

43. Schoenfelder, S. et al. The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. *Genome Res.* **25**, 582–597 (2015).

44. Mumbach, M. R. et al. HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat. Methods* **13**, 919–922 (2016).

45. Servant, N. et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259 (2015).

46. Lareau, C. A. & Aryee, M. J. hichipper: a preprocessing pipeline for calling DNA loops from HiChIP data. *Nat. Methods* **15**, 155–156 (2018).

47. Durand, N. C. et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst* **3**, 95–98 (2016).

48. Man, J. C. K. et al. Genetic dissection of a super enhancer controlling the *Nppa-Nppb* cluster in the heart. *Circ. Res.* **128**, 115–129 (2021).

49. Chen, H. et al. BMP10 is essential for maintaining cardiac growth during murine cardiogenesis. *Development* **131**, 2219–2231 (2004).

50. Xia, B. et al. Machine learning uncovers cell identity regulator by histone code. *Nat. Commun.* **11**, 2696 (2020).

51. Shim, W. J. et al. Conserved epigenetic regulatory logic infers genes governing cell identity. *Cell Syst* **11**, 625–639 (2020).

52. Zhou, P. et al. Dynamic changes in P300 enhancers and enhancer-promoter contacts control mouse cardiomyocyte maturation. *Dev. Cell* **58**, 898–914 (2023).

53. Yuan, S., Norgard, R. J. & Stanger, B. Z. Cellular plasticity in cancer. *Cancer Discov.* **9**, 837–851 (2019).

54. LeBleu, V. S. & Neilson, E. G. Origin and functional heterogeneity of fibroblasts. *FASEB J.* **34**, 3519–3536 (2020).

55. Lickert, H. et al. Baf60c is essential for function of BAF chromatin remodelling complexes in heart development. *Nature* **432**, 107–112 (2004).

56. Robbe, Z. L. et al. CHD4 is recruited by GATA4 and NKX2-5 to repress noncardiac gene programs in the developing heart. *Genes Dev.* **36**, 468–482 (2022).

57. Panigrahi, A. & O'Malley, B. W. Mechanisms of enhancer action: the known and the unknown. *Genome Biol.* **22**, 108 (2021).

58. Sinner, M. F. et al. Integrating genetic, transcriptional, and functional analyses to identify 5 novel genes for atrial fibrillation. *Circulation* **130**, 1225–1235 (2014).

59. Fu, J.-D. D. et al. Direct reprogramming of human fibroblasts toward a cardiomyocyte-like state. *Stem Cell Rep.* **1**, 235–247 (2013).

60. Holm, H. et al. Several common variants modulate heart rate, PR interval and QRS duration. *Nat. Genet.* **42**, 117–122 (2010).

61. Hoesl, E. et al. Tamoxifen-inducible gene deletion in the cardiac conduction system. *J. Mol. Cell. Cardiol.* **45**, 62–69 (2008).

62. Muzumdar, M. D., Tasic, B., Miyamichi, K., Li, L. & Luo, L. A global double-fluorescent Cre reporter mouse. *Genesis* **45**, 593–605 (2007).

63. Peron, S. P., Freeman, J., Iyer, V., Guo, C. & Svoboda, K. A cellular resolution map of barrel cortex activity during tactile behavior. *Neuron* **86**, 783–799 (2015).

64. Agah, R. et al. Gene recombination in postmitotic cells. Targeted expression of Cre recombinase provokes cardiac-restricted, site-specific rearrangement in adult ventricular muscle in vivo. *J. Clin. Invest.* **100**, 169–179 (1997).

65. Wang, S., Guo, Y. & Pu, W. T. AAV gene transfer to the heart. *Methods Mol. Biol.* **2158**, 269–280 (2021).

66. Jiang, J., Wakimoto, H., Seidman, J. G. & Seidman, C. E. Allele-specific silencing of mutant *Myh6*transcripts in mice suppresses hypertrophic cardiomyopathy. *Science* **342**, 111–114 (2013).

67. Nadelmann, E. R. et al. Isolation of nuclei from mammalian cells and tissues for single-nucleus molecular profiling. *Curr Protoc* **1**, e132 (2021).

68. McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst.* **8**, 329–337(2019).

69. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).

70. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA.* **102**, 15545–15550 (2005).

71. Chen, E. Y. et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128 (2013).

72. McLean, C. Y. et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).

73. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

74. Cui, M. & Olson, E. N. Protocol for single-nucleus transcriptomics of diploid and tetraploid cardiomyocytes in murine hearts. *STAR Protoc* **1**, 100049 (2020).

75. Phanstiel, D. H., Boyle, A. P., Heidari, N. & Snyder, M. P. Mango: a bias-correcting ChIA-PET analysis pipeline. *Bioinformatics* **31**, 3092–3098 (2015).

76. Lareau, C. A. & Aryee, M. J. diffloop: a computational framework for identifying and analyzing differential DNA loops from sequencing data. *Bioinformatics* **34**, 672–674 (2018).

77. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

78. Song, F., Xu, J., Dixon, J. & Yue, F. Analysis of Hi-C data for discovery of structural variations in cancer. In *Hi-C Data Analysis: Methods and Protocols* (eds Bicciato, S. & Ferrari, F.) 143–161 (Springer, 2022).

79. Durand, N. C. et al. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.* **3**, 99–101 (2016).

80. Chen, K. et al. DANPOS: dynamic analysis of nucleosome position and occupancy by sequencing. *Genome Res.* **23**, 341–351 (2013).

Acknowledgements

Author contributions

M.E.S. and Y.C. contributed equally to this study. M.E.S. and Y.C. conceived of the study; designed and performed the experiments; and analyzed the data. X.Z., C.P.-C. and K.A. contributed to the data analysis. O.B.-T., B.N.A., Q.M., E.K., H.W., J.M.G., L.H., M.K.S., M.A.T., P.W., F.L., M.G., M.P. and R.H.B. contributed data, reagents and analyses. V.B., K.C., J.G.S., C.E.S., I.P.M. and W.T.P. oversaw the project and provided resources. M.E.S. and W.T.P. wrote the manuscript, with input from Y.C. and the other authors.

Competing interests

The authors have no competing interests to declare.

Additional information

Extended Data Fig. 1 | Characterization of the cardiac expression domain of AAV9:*Nppa-EGFP*. a-b, *Hcn4^{CreERT2/+}; Rosa26^{LSL-Tomato}* pups were injected with 2×10^{11} viral genomes per gram bodyweight (VG/g) *AAV9:Nppa-EGFP* at P8 and injected with tamoxifen at P21 and P22 to activate tomato expression in the conduction system including the sinoatrial and atrioventricular nodes. Hearts were harvested in PBS and brightfield, GFP, and Tomato fluorescent images were acquired. Mutually exclusive GFP and tomato signals were observed in the atrium and the SA node, respectively. A similar result was observed with administration at P2. **c,** Hearts were sectioned and stained with WGA. Tomato signal marked the sinoatrial node (SAN) and atrioventricular node (AVN) but not the working myocardium. GFP was restricted to the working myocardium. A similar staining pattern was observed in 4 other littermates.

Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Phenotypic characterization of *Tbx5^{AKO}* mice. a, Trichrome staining of *Tbx5^{Flox/Flox}* and *Tbx5^{AKO}* hearts. **b**, RT-qPCR quantification of *Postn* (periostin) and *Col1a1* (collagen type 1 alpha 1 chain), two indicators of fibrosis (n = 3 control and 4 KO mice). Unpaired two-sided *t*-test. Error bars represent mean values +/-SEM. **c-d**, Immunostaining for sarcomeric α-actinin (SAA) or FSD2, markers of the Z line and the junctional Sarcoplasmic reticulum, respectively, in the left atrium of the indicated genotypes. Localization of both proteins is disrupted in *Tbx5^{AKO}* atria. **e**, Pattern of SAA signal intensity. SAA intensity along the long axis of cardiomyocytes demonstrated a periodic signal in control, consistent with regular position of sarcomere Z-lines, and loss of periodicity in *Tbx5^{AKO}*. **f**, Preserved ventricular function of *Tbx5^{AKO}* mice. Tbx5^{Flox/Flox} mice were treated with *AAV:Nppa-EGFP* (control) or *AAV:Nppa-Cre* (*Tbx5^{AKO}*) at P2. Echocardiography was performed at P20. LVID;d, left ventricular internal diameter at end diastole. EF, ejection fraction. n = 5 control and 4 KO mice. Graphs show mean ± SEM. **g**, Surface EKG recordings. Orange and black bars highlight successive RR intervals. **h**, Poincaré plots. The RR interval of greater than 1500 beats on EKG recordings is plotted versus the RR interval of the subsequent beat (RR[+ 1]). to visualize the dispersion of interbeat intervals in *Tbx5^{AKO}*, consistent with atrial fibrillation. **i**, Standard deviation of the RR interval, a measure of heart rate irregularity, was calculated for at least 1500 beats for each group at P21. n = 4 control and 9 KO mice. Unpaired two-sided *t*-test: ***, P = 0.005. Graphs show mean ± SEM. **j**, Time course of heart rate irregularity. Serial EKGs were acquired from control or *Tbx5^{AKO}* mice at the indicated time points. SDRR was measured and compared between groups at each time point using a two-way ANOVA with Sidek's multiple comparison test. **, P < 0.01. ***, P < 0.001. For P14 and P17 timepoints, n = 3 control and 3 KO mice. For P8 and P21 timepoints, n = 5 control and 8 KO mice. Graphs show mean ± SEM. **k**, Simultaneous intracardiac and surface ECG recordings demonstrate normal synchronous atrial-ventricular rhythm in *Tbx5^{Flox/+}* mice injected with *AAV-Nppa-Cre*. **l**, In contrast, animals with complete atrial ablation of *Tbx5* (*AAV-Nppa-Cre + Tbx5^{flox/flox}*) demonstratee nearly continuous low-amplitude atrial activity and irregularly irregular ventricular response consistent with atrial fibrillation.
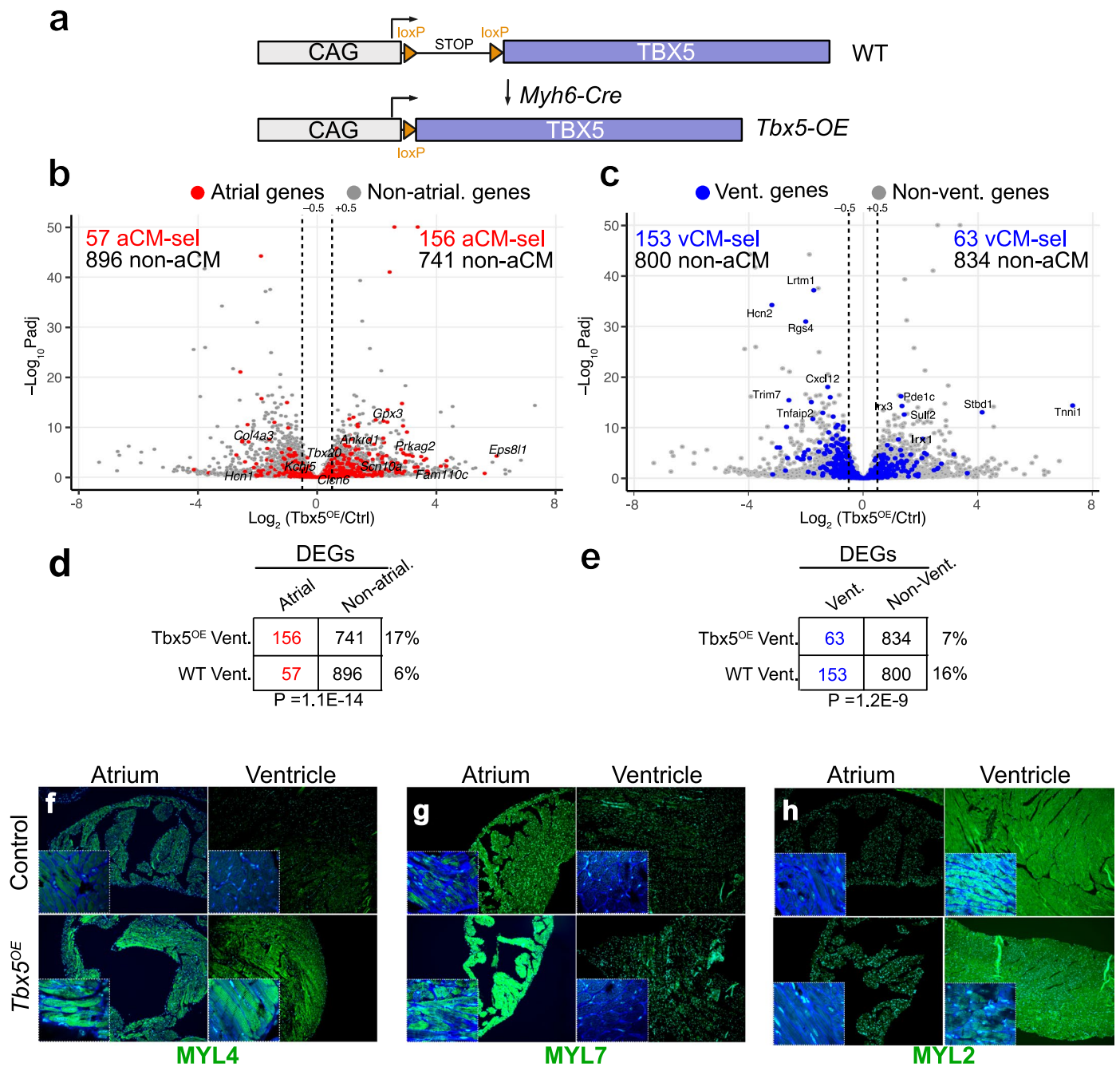
Extended Data Fig. 3 | Tbx5 overexpression atrializes ventricular myocytes.
a, Strategy to generate *TBX5-OE* ventricles. A full length *Tbx5* cDNA downstream of the ubiquitous CAG promoter was activated by the cardiomyocyte specific Myh6-Cre transgene. **b-c**, Volcano plot comparing the change in gene expression of mouse left ventricle overexpressing *Tbx5* compared to control LV. aCM genes are marked in red in (a) and vCM genes are denoted in blue in (b). b-c, Wald's test followed by Benjamini Hochberg correction. **d-e**, Fisher's exact test (two tailed) was performed to determine if changes in chamber selective gene expression downstream of *Tbx5* overexpression were significant. **f-h**, Staining for MYL4, MYL7, and MYL2 in atria and ventricles of control and *TBX5-OE* hearts. Staining for each marker was performed using 3 control hearts and 2 *TBX5-OE* hearts, and representative images are shown.

a

Sample	mean nCount_ATAC	mean nFeature_ATAC	mean nCount_RNA	mean nFeature_RNA
KO1	10779	7880	4893	1970
KO2	9810	7085	4760	1913
Con1	7560	5823	6432	2153
Con2	6717	5190	4504	1732

b

snATAC-seq enrichment at TSS

- Con_aCM
- KO_aCM
- Fibroblast_1
- Fibroblast_2
- Fibroblast_3
- Pericyte
- Vascular_endothelium
- Endocardium_1
- Endocardium_2
- Epicardium
- Adipose
- Neuron
- Macrophage_1
- Macrophage_2
- T-cell
- B-cell

c

Extended Data Fig. 4 | Single cell dataset metrics. a, Parameters from single nucleus datasets. Mean values per nucleus: nCount_ATAC, number of ATAC fragments; nFeature_ATAC, number of ATAC peaks with at least one read; nCount_RNA, number of RNA fragments; and nFeature_RNA, number of genes. **b**, Transcription start site (TSS) aggregation plots for the scATAC multiome datasets showing the expected enrichment of ATAC fragments. **c**, WNN UMAP of the dataset split by original sample. Each of the KO and control replicates have a high degree of overlap, demonstrating high reproducibility of cell state changes in *Tbx5^{AKO}* atria.

Extended Data Fig. 5 | Differentially expressed genes between myocyte clusters. a, Volcano plot of differentially expressed genes (DEGs) between early and late pseudotime clusters in the control trajectory (Myocyte_4 vs. Myocyte_1, left) and the KO trajectory (Myocyte_5 vs. Myocyte_6; right). Wilcoxon rank sum test, Bonferroni correction. **b**. MA plot of RNA-seq experiment comparing P0 and P28 aCMs. **c-d**. Volcano plots shown in (a) overlayed with genes enriched in P28 and P0 aCMs. Wilcoxon rank sum test, Bonferroni correction. **e**. The proportion of DEGs from comparisons in (a) that overlap genes selectively expressed in aCMs at P0 and P28. We did not observe enrichment of P0 or P28 selective aCM genes in early or late pseudotime clusters, respectively. This suggested that pseudotime trajectories did not correspond to chronological time. **f-g**. GO biological process terms enriched for DEGs for the comparisons shown in (a). The top 10 terms enriched by genes upregulated in the indicated cluster are shown. Functional terms related to cardiomyocyte function were enriched in the late pseudotime clusters, Myo_1 (control) and Myo_6 (AKO). Two-tailed Fisher's exact test, Bonferroni correction. **h**, WNN UMAP plot colored by a 'functional cardiac gene' index, which was calculated based on the aggregate expression of the six indicated genes, which are required for the efficient pumping function of aCMs.

a

***Tbx5* inactivation in atria: Comparison of**
***Tbx5^{AKO}* LA aCMs vs *Tbx5^{iKO}* LA tissue**

Colored by significance in each dataset

● Both sig.   ● *Tbx5^{AKO}*
● *Tbx5^{iKO}*   ● Not sig.

Both significant, colored by chamber selectivity

Non-sel.   aCM-sel   vCM-sel

b

	aCM	non-aCM	
Quad I	149	351	30%
Quad III	39	336	8%

P=1.4E-12

	vCM	non-vCM	
Quad I	29	471	6%
Quad III	48	327	13%

P=0.0004

c

***Tbx5* inactivation and overexpression in atria: Comparison of**
***Tbx5^{AKO}* LA aCMs to *Tbx5^{RE(int)KO}* LA**

Colored by significance in each dataset

● Both sig.   ● *Tbx5^{AKO}*
● *Tbx5^{RE(int)KO}*   ● Not sig.

Both significant, colored by chamber selectivity (DEGs)

Non-sel.   aCM-sel   vCM-sel

d

	aCM	non-aCM	
Quad II	21	229	8%
Quad IV	41	82	33%

P=5E-9

	vCM	non-vCM	
Quad II	43	207	6%
Quad IV	6	117	13%

P=0.0006

e

***Tbx5* inactivation in atria to overexpression in ventricles: Comparison of**
***Tbx5^{AKO}* in aCM to Tbx5-OE in vCM**

Colored by significance in each dataset

● Both sig.   ● *Tbx5^{AKO}*
● *Tbx5^{iKO}*   ● Not sig.

Both significant, colored by chamber selectivity (DEGs)

Non-sel.   aCM-sel   vCM-sel

f

	aCM	non-aCM	
Quad I	45	82	35%
Quad III	9	76	11%

P=4.4E-5

	vCM	non-vCM	
Quad I	5	122	4%
Quad III	17	68	20%

P=0.0003

g

	1126 aCM selective genes	NES	FDR	873 vCM selective genes	NES	FDR
WT LA / *Tbx5^{iKO}* LA		1.67	.005		-1.48	0.045
control_aCM / KO_aCM		1.79	0.0		-1.44	0.0
Tbx5^{Re(int)KO} LA / WT LA		1.02	0.41		-1.57	0.00
TBX5 OE LV / WT LV		1.37	0.00		-1.10	0.31

Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | A comparison of TBX5 RNA-seq datasets. We compared snRNAseq data from *Tbx5^AKO^* aCMs to three other bulk RNA-seq datasets involving *Tbx5* gain- or loss- of function. Control_aCM indicates Myocyte_1 and KO_aCM indicates Myocyte_6. **a**, **c**, **e**: Scatter plot of fold-change in *Tbx5^AKO^* aCMs compared to the other three datasets. Left, points are colored by significance in each dataset ($P_{adj} < 0.05$). Right, genes with significant differential expression in both datasets are colored by chamber selectivity. Pie charts summarize the proportion of each class of genes within each quadrant. **b**, **d**, **f**: Enrichment of aCM or vCM genes in the indicated quadrants of the scatter plots. Fisher's exact test, two-tailed. **a-b**. Comparison to LA tissue with ubiquitous, adult-induced inactivation of *Tbx5* (*Tbx5^iKO^*; Nadadur et al., 2016). **c-d**. Comparison to LA tissue with mild *Tbx5* upregulation due to deletion of an intronic regulatory element (*Tbx5^Re(int)KO^*; Bosada et al., 2023). **e-f**. Comparison to LV tissue with *Tbx5* overexpression in cardiomyocytes (*Tbx5-OE*; this study). **g**. Comparison of changes in expression of aCM- and vCM-selective genes across all four datasets. Fold-change was calculated between condition with higher *Tbx5* (numerator) to condition with lower *Tbx5* (denominator). Genes were ordered by ascending Log_2(WT/*TBX5^iKO^*). Genes not detected (ND) for a given experiment are colored white. Normalized enrichment score (NES) and false discovery rate (FDR) values from GSEA using the aCM-selective or vCM-selective gene lists are shown. Negative NES indicates enrichment in the genes upregulated in the lower *Tbx5* condition, whereas positive NES indicates enrichment in the genes upregulated in the higher *Tbx5* condition.
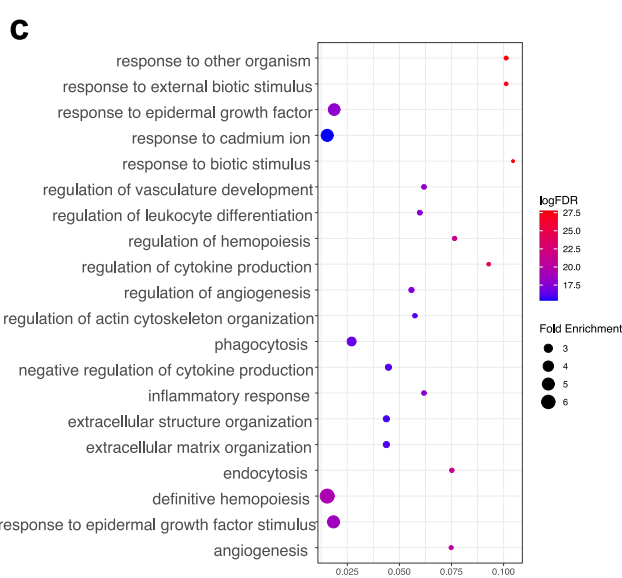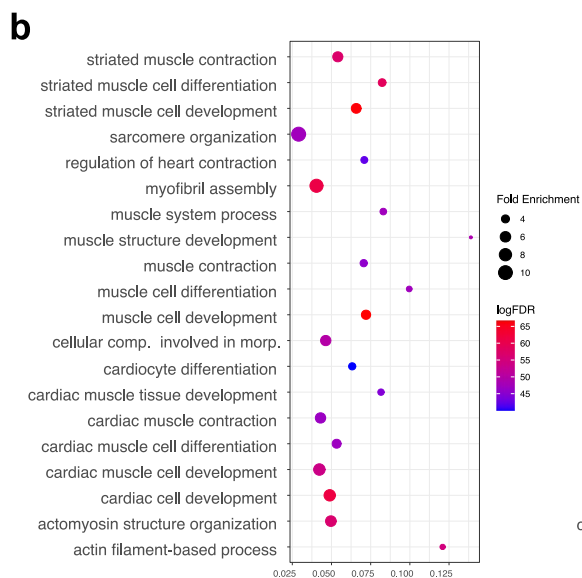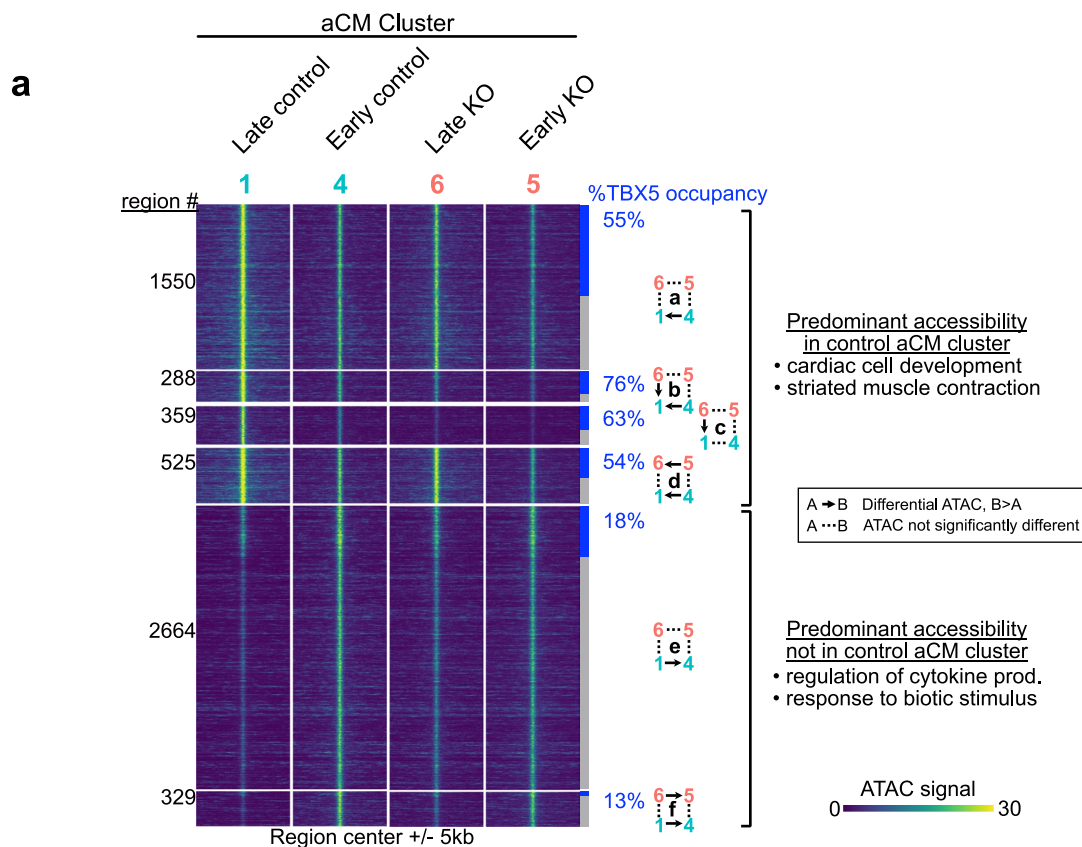
Extended Data Fig. 7 | Characterization of myocyte differentially accessible regions. Heatmap (**a**) shows the patterns of differential accessibility between myocyte clusters. The rows contain the union of regions with differential accessibility in the four pairwise comparisons shown in Fig. 5a. ATAC signal in each region is shown for myocyte clusters 1 (control_aCM cluster), 4, 6 (KO_aCM cluster), and 5. The regions are grouped (groups a-f) by their pattern of accessibility change in the four pairwise comparisons. Please also refer to Supp. Table 4. Arrows denote significant enrichment in one cluster compared to another. These six groups fit two predominant patterns: those with and those without predominant accessibility in the control aCM cluster (Myocyte_1). Most regions with predominant accessibility in the control aCM cluster were occupied by TBX5 and had GO terms related to cardiac cell development or striated muscle contraction (**b**). In contrast, a minority of regions without predominant accessibility in the control aCM cluster were occupied by TBX5 and had GO terms that were atypical for cardiomyocytes (**c**). Groups with less than 200 regions are not shown in the heatmap.
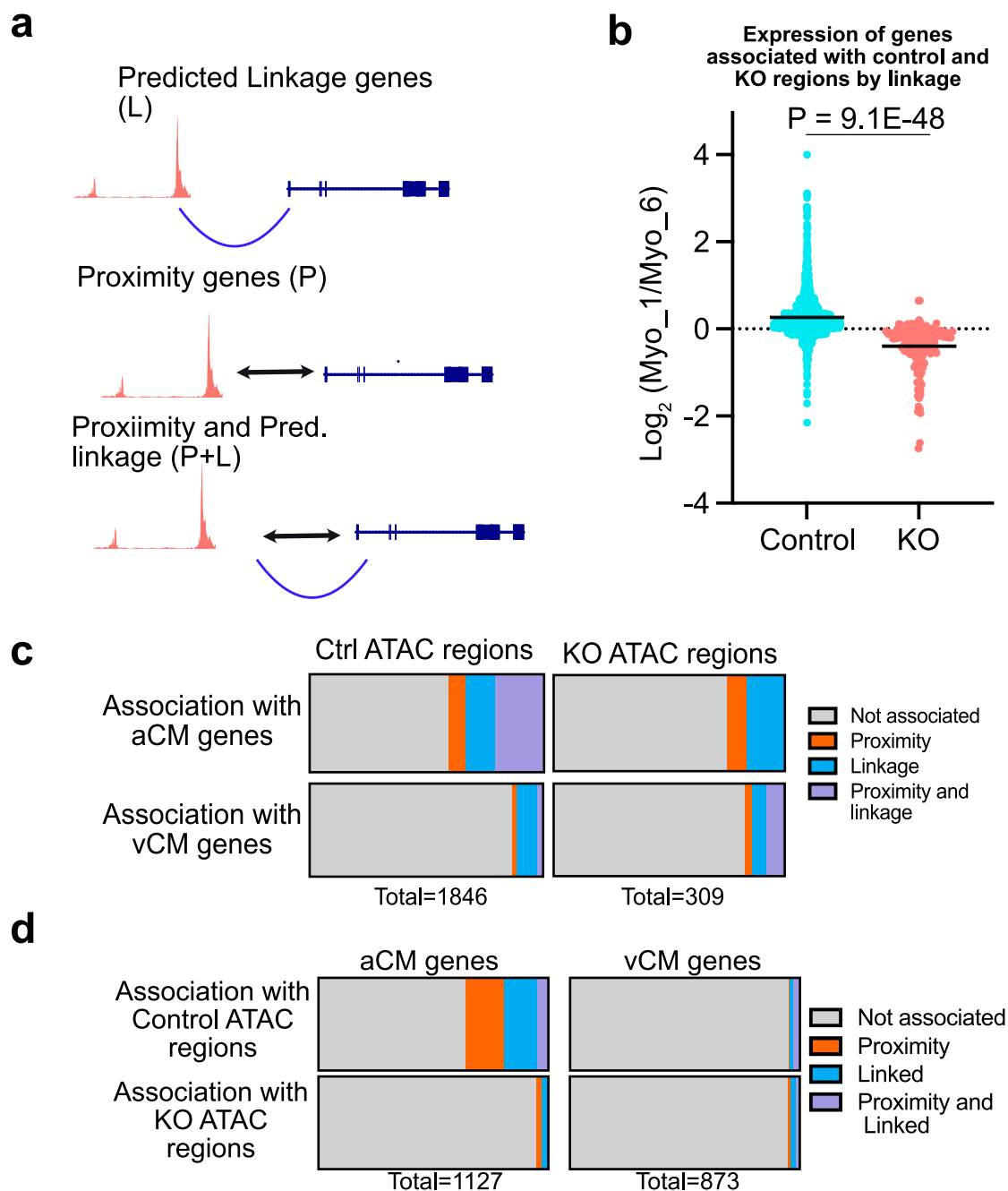
Extended Data Fig. 8 | Proximity and predicted gene linkages demonstrate the regulation of the atrial GRN by control peaks. a, The different types of association of a genomic region with a gene: (1) Linkage (L). Region-to-gene linkages are predicted based on co-variance of accessibility and expression on a nucleus-by-nucleus basis in the multiome data. (2) Proximity (P). Region-to-gene relationships are inferred by proximity of the region to the gene's transcriptional start site. (3) Proximity and linkage (P + L). A region-to-gene association can be supported by both proximity and linkage. **b**, Genes were associated with control and KO regions by linkage (n = 2 control and 2 KO multiome biological replicates). Ratio of gene expression between the control and KO aCM clusters was plotted and compared between groups by the two-sided Mann-Whitney test. P = 9.1E-48. **c**, Association of control and KO ATAC regions with aCM and vCM genes. Associations were made based on proximity, linkage, or both. **d**, aCM and vCM genes were interrogated for associations with control and KO ATAC regions by proximity, linkage, or both.

a

Control ATAC regions

KO ATAC regions

b

c

motif	percent.observed	percent.background	fold.enrichment	pval_Adj	motif.name
MA0807.1	25.5	9.7	2.6	5.88E-51	TBX5
MA0497.1	36.6	25.1	1.5	5.09E-17	MEF2C
MA0783.1	23.7	14.5	1.6	5.07E-16	PKNOX2
MA0513.1	33.7	23.1	1.5	1.91E-15	SMAD2/3/4
MA0797.1	19.4	11.4	1.7	7.49E-15	TGIF2
MA0774.1	17.5	10.7	1.7	7.15E-12	MEIS2
MA0782.2	24.7	17.0	1.5	8.49E-11	PKNOX1
MA1114.1	26.6	19.2	1.4	2.27E-09	PBX3
MA1576.1	21.8	15.7	1.4	6.79E-08	THRB(var.3)
MA0595.1	14.9	10.0	1.5	3.03E-07	SREBF1
MA0093.3	17.6	12.4	1.4	3.60E-07	USF1
MA0831.2	12.4	8.0	1.6	4.30E-07	TFE3
MA1494.1	15.7	10.8	1.5	4.41E-07	HNF4A(var.2)
MA1485.1	17.0	11.8	1.4	4.45E-07	FERD3L
MA0526.3	17.0	12.0	1.4	1.37E-06	USF2
MA0730.1	18.3	13.2	1.4	1.49E-06	RARA(var.2)
MA1467.1	24.8	19.1	1.3	2.37E-06	ATOH1(var.2)
MA0596.1	15.3	10.7	1.4	3.18E-06	SREBF2
MA0504.1	26.7	21.0	1.3	5.23E-06	NR2C2
MA0499.2	22.5	17.3	1.3	7.35E-06	MYOD1
MA0672.1	13.6	9.7	1.4	1.75E-05	NKX2-3

d

e

motif	percent.observed	percent.background	fold.enrichment	pval_Adj	motif.name
MA0695.1	15.1	5.1	3.0	1.89E-20	ZBTB7C
MA0060.3	19.1	8.6	2.2	4.22E-16	NFYA
MA0525.2	12.9	4.9	2.6	1.12E-14	TP63
MA1570.1	14.4	6.3	2.3	1.02E-12	TFAP4(var.2)
MA0070.1	15.4	7.1	2.2	2.13E-12	PBX1
MA1643.1	32.0	21.3	1.5	6.79E-10	NFIB
MA1117.1	16.4	9.0	1.8	4.98E-09	RELB
MA0747.1	38.4	27.6	1.4	5.18E-08	SP8
MA0056.2	19.9	12.1	1.6	2.54E-08	MZF1
MA0073.1	55.6	44.8	1.2	5.38E-08	RREB1
MA1653.1	48.4	37.9	1.3	8.56E-08	ZNF148
MA1523.1	17.8	10.7	1.7	1.21E-07	MSANTD3
MA0741.1	43.9	33.8	1.3	1.49E-07	KLF16
MA1581.1	25.8	17.5	1.5	1.98E-07	ZBTB6
MA0143.4	16.8	10.2	1.6	5.50E-07	SOX2
MA1592.1	19.4	12.4	1.6	5.61E-07	ZNF274
MA1120.1	17.1	10.6	1.6	8.45E-07	SOX13
MA1522.1	37.7	28.8	1.3	1.43E-06	MAZ
MA1528.1	23.5	16.2	1.4	2.37E-06	NFIX(var.2)
MA0695.1	15.1	5.1	3.0	1.89E-20	ZBTB7C

f

Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Motif analysis of control and KO ATAC regions. a, Aggregation plots for H3K27Ac at control and KO regions. **b**, Heatmaps of the H3K27Ac signal at control and KO peaks. **c**, Top-enriched motifs identified in control regions. The TBOX motif was the most enriched, followed by MEF2. An extended table of non-redundant motifs showing the top 21 most enriched motifs in control regions. **d**, Transcription factor footprinting analysis demonstrates footprints at TBOX and MEF2 motifs in control clusters (Myocyte_1 and Myocyte_4) compared to KO clusters (Myocyte_5 and Myocyte_6). **e**, Top-enriched motifs identified in KO regions and extended table of the top non-redundant motifs. **f**. Occupancy of control and KO ATAC regions by cardiac TFs in aCMs. Occupancy data is from GEO GSE215065[3].
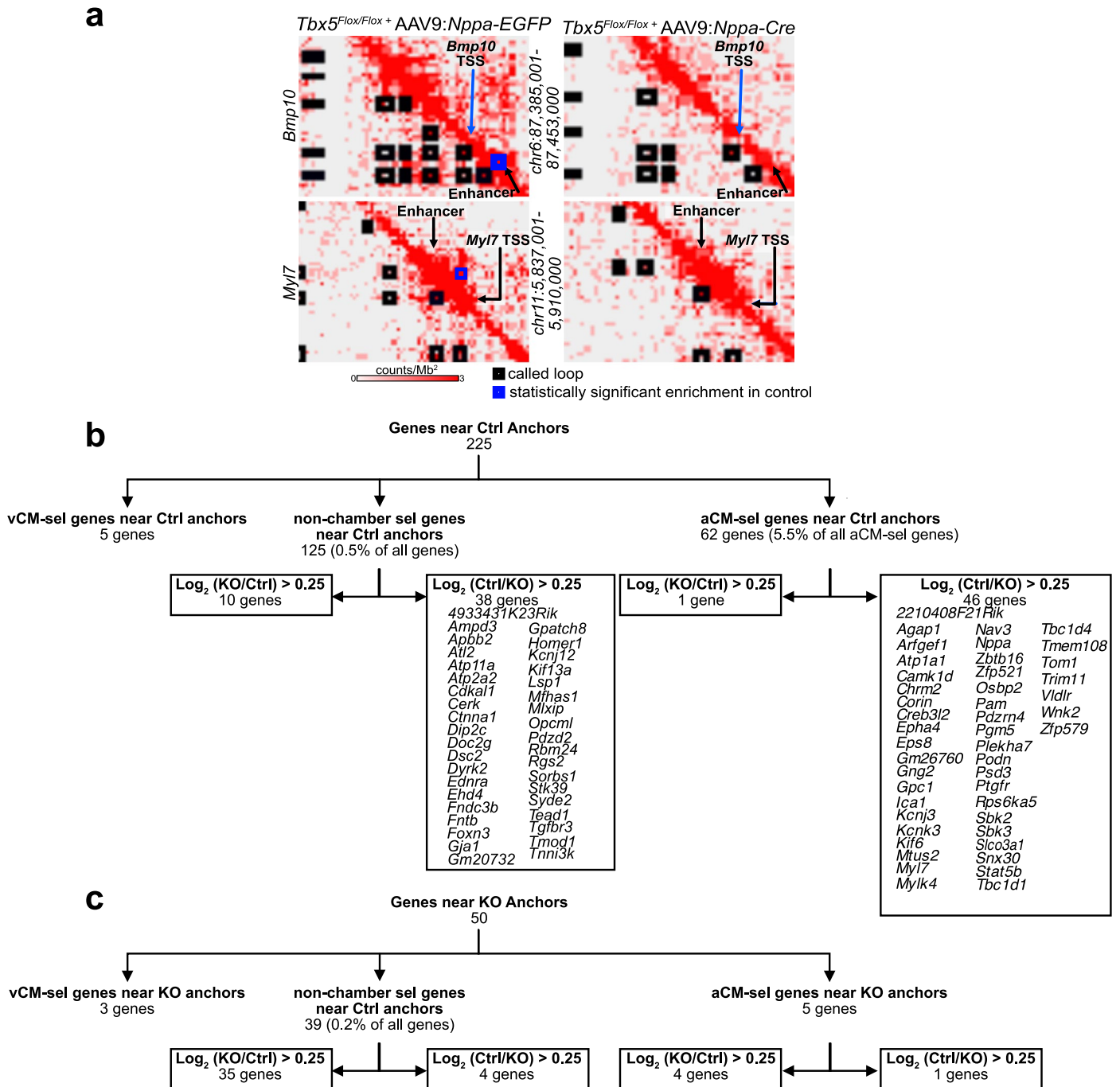
Extended Data Fig. 10 | Chromatin loops link TBX5 dependent enhancers with atrial genes. a, Contact maps of *Myl7* or *Bmp10*. Black boxed regions are loops called in each sample and blue boxed regions mark differential loops that are significantly stronger in control samples. **b**, Genes near control anchors grouped by adjacency to aCM-selective, vCM-selective, or non-chamber selective expression. Control anchors were present near 62 aCM-selective genes and of these, 46 were expressed at greater levels in control Myocyte_1 compared to KO

Myocyte_6. Notable genes from previous figures include *Nppa*, *Bmp10*, *Sbk2* and *Myl7*. 38 non-chamber selective genes were also upregulated in control samples and linked to enhancers by TBX5-dependent looping. These included *Gja1* and *Tead1*. Only 5 vCM-selective genes were found near control anchors. **c**, 50 genes neighbored KO anchors. These genes included 5 aCM-selective genes and 3 vCM-selective genes. Most of the differentially expressed genes near KO anchors were more highly expressed in Myocyte_6 (KO) compared to Myocyte_1 (Ctrl).

Corresponding author(s): William Pu

Last updated by author(s): Aug 15, 2023

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used for data collection.
Data analysis	Data analysis was performed using open source software packages, including R Version 4.1.2 and Seurat (v4.3.0) and Signac (v1.6.0). Overlapping peak analysis was performed with bedtools (v2.3.0). Hichip analysis was performed with pairtools (v0.3.0), HiCPro(v2.3.1) and HiChipper (v0.7.7). Bulk RNA-seq analysis was performed using DEseq2 (release 3.17). Interaction maps were performed with Juicebox (v1.11.08). Psuedobulk ATAC tracks were obtained using the Sinto package (v0.8.4). More information is available in the methods section. Custom code was not generated outside of the use of the predesigned packages.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

> All sequencing data was deposited in GEO under the superseries GSE222970. There are no applicable restrictions for the data.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender (identity/presentation), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	*Use the terms sex (biological attribute) and gender (shaped by social and cultural circumstances) carefully in order to avoid confusing both terms. Indicate if findings apply to only one sex or gender; describe whether sex and gender were considered in study design; whether sex and/or gender was determined based on self-reporting or assigned and methods used. Provide in the source data disaggregated sex and gender data, where this information has been collected, and if consent has been obtained for sharing of individual-level data; provide overall numbers in this Reporting Summary. Please state if this information has not been collected. Report sex- and gender-based analyses where performed, justify reasons for lack of sex- and gender-based analysis.*
Reporting on race, ethnicity, or other socially relevant groupings	*Please specify the socially constructed or socially relevant categorization variable(s) used in your manuscript and explain why they were used. Please note that such variables should not be used as proxies for other socially constructed/relevant variables (for example, race or ethnicity should not be used as a proxy for socioeconomic status). Provide clear definitions of the relevant terms used, how they were provided (by the participants/respondents, the researchers, or third parties), and the method(s) used to classify people into the different categories (e.g. self-report, census or administrative data, social media data, etc.) Please provide details about how you controlled for confounding variables in your analyses.*
Population characteristics	*Describe the covariate-relevant population characteristics of the human research participants (e.g. age, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."*
Recruitment	*Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.*
Ethics oversight	*Identify the organization(s) that approved the study protocol.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences   ☐ Behavioural & social sciences   ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](#)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample sizes are reported in each figure. For high throughput data, samples were performed at least in duplicate to demonstrate reproducibility. For assessment of AF, sample sizes were determined using power calculations based on historical effect sizes and variance. Other studies targeted n=3-5 as a minimum for statistical evaluation.
Data exclusions	No data were excluded from the study.
Replication	Mouse experiments were repeated in multiple litters. Staining was performed on samples obtained from multiple mice. Multiple batches of AAV were used in the experiment. All results were repeatable, and different experiments reached the same conclusions (for example, qPCR, bulk RNA-seq and snRNA-seq).
Randomization	Mice were randomly assigned to experimental and control groups.
Blinding	All EKG and heart function data was obtained in a blinded manner. The individual obtaining the measurements was unaware of the animal's inclusion in either the control or experimental groups. Images were analyzed in a blinded manner.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☐	☒ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	The information on all primary antibodies is provided in supp. table 7.
Validation	Where possible, antibodies were validated by using KO tissues. Where not possible, experiments using antibodies were cross-validated by measuring protein transcript levels by either NGS sequencing or RT-qPCR. SAA antibody (Sigma-Aldrich A7811) is specific for alpha actinin and labels Z lines. FSD2 antibody (Santa Cruz, sc-393072) is not validated but is highly specific to muscle tissue. MYL7 antibody (Proteintech, 60229) is not validated by is highly specific to mouse atria and marks A-bands. MYL2 antibody (abcam ab79935) is experimentally validated in KO cells and highly specific for the ventricle. TBX5 antibody (abcam 137833) was validated in our experiment by inactivating TBX5 in the atrium. GAPDH (invitrogen PA1-16777) was validated by a knockdown experiment. mCherry (Abcam ab167453) is a knockout-validated antibody.

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	HEK293T cells were purchased directly from ATCC (catologue # CRL-3216)
Authentication	Cells were not authenticated as they were purchased directly from ATCC.
Mycoplasma contamination	All cell lines tested negative for Mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	N/A

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	Mus musculus were used. Mice were maintained on a mixed background. Initial treatments were performed on P2 mice and experiments were performed on P21 mice, unless otherwise indicated by the figure.
Wild animals	N/A
Reporting on sex	Both male and females were used throughout the study in approximately equal numbers. We did not detect an effect of sex on the experimental results.
Field-collected samples	N/A
Ethics oversight	Boston Children's Hospital IACUC approved of all animal protocols used throughout the study.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	All data was deposited under accession number GSE222970.
Files in database submission	Both unprocessed FASTQ files and downstream analyses were deposited in the GEO submission. These include the final output of HiChipper, .mango files, which identified loops in each sample and can be used for statistical comparison
Genome browser session (e.g. UCSC)	Washu epigenome browser was used to create browser images, as described in the methods.

Methodology

Replicates	Biological duplicates consisting of 10 pairs of atria from control and TBX5 AKO mice were used for H3K27ac HichIP
Sequencing depth	Reads were paired-end. These data are summarized in Supp. Table 1 (Summary of high-throughput data used in the study).
Antibodies	The H3K27ac antibody used in the HiChIP assay was from active motif (39133)
Peak calling parameters	Peaks were called using HiCPro and then combined for use in Hichipper. The bowtie2 command is as follows: /programs/x86_64-linux/hicpro/3.1.0/bin/bowtie2 --very-sensitive -L 30 --score-min L,-0.6,-0.2 --end-to-end --reorder --un bowtie_results/bwt2_global/Con1/control_1_hichip_R1_mm10.bwt2glob.unmap.fastq --rg-id BMG --rg SM:control_1_hichip_R1 -p 1 -x /lab-share/Cardio-Pu-e2/Public/ch220446_Mason2/bowtie2/mm10/mm10 -U rawdata/Con1/control_1_hichip_R1.fastq.gz 2>> logs/Con1/control_1_hichip_R1_bowtie2.log| /programs/x86_64-linux/hicpro/3.1.0/bin/samtools view -F 4 -bS - > bowtie_results/bwt2_global/Con1/control_1_hichip_R1_mm10.bwt2glob.bam
Data quality	Data quality of a HiChIP experiment involves examining how reads are distributed among anchors, the % of HQ Unique mapped reads, the % Long Range Interaction, and the % in loops. These values are summarized in Supp. Table 1 for each replicate and are typical of a successful HiChIP experiment. All data metrics are reported in the supplemental data file.
Software	The data used to process the HiChIP experiment used the following pipeline: HiCPro -> Hichipper -> Diffloop. Contact matrices were produced using Juicebox software. A detailed description of the bioinformatics pipeline is found in the methods.