HHS Public Access

Author manuscript

Nat Cell Biol. Author manuscript; available in PMC 2021 June 30.

Published in final edited form as:

Nat Cell Biol. 2021 May ; 23(5): 467–475. doi:10.1038/s41556-021-00668-z.

The histone reader PHF7 cooperates with the SWI/SNF complex at cardiac super enhancers to promote direct reprogramming

Glynnis A. Garry[1,2,3], **Svetlana Bezprozvannaya**[1,2,3], **Kenian Chen**[4], **Huanyu Zhou**[1,2,3], **Hisayuki Hashimoto**[1,2,3], **Maria Gabriela Morales**[1,2,3], **Ning Liu**[1,2,3], **Rhonda Bassel-Duby**[1,2,3], **Eric N. Olson**[1,2,3]

[1]Department of Molecular Biology, University of Texas Southwestern Medical Center, Dallas, TX, USA.

[2]The Hamon Center for Regenerative Science and Medicine, University of Texas Southwestern Medical Center, Dallas, TX, USA.

[3]Senator Paul D. Wellstone Muscular Dystrophy Specialized Research Center, University of Texas Southwestern Medical Center, Dallas, TX, USA.

[4]Department of Clinical Sciences, University of Texas Southwestern Medical Center, Dallas, TX, USA.

Abstract

Direct cardiac reprogramming of fibroblasts to cardiomyocytes presents an attractive therapeutic strategy to restore cardiac function following injury. Cardiac reprogramming was initially achieved through overexpression of the transcription factors Gata4, Mef2c and Tbx5; later, Hand2 and Akt1 were found to further enhance this process[1–5]. Yet, staunch epigenetic barriers severely limit the ability of these cocktails to reprogramme adult fibroblasts[6,7]. We undertook a screen of mammalian gene regulatory factors to discover novel regulators of cardiac reprogramming in adult fibroblasts and identified the histone reader PHF7 as the most potent activating factor[8]. Mechanistically, PHF7 localizes to cardiac super enhancers in fibroblasts, and through cooperation

eric.olson@utsouthwestern.edu.

Author contributions

G.A.G., K.C., H.Z. and H.H. designed and performed experiments, data analyses, discussion and writing. S.B. and M.G.M. contributed to experimental work and discussion. N.L. and R.B.-D. contributed to discussion and writing. E.N.O. supervised the study and contributed to discussion and writing.

Competing interests

The other authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41556-021-00668-z.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41556-021-00668-z.

Peer review information *Nature Cell Biology* thanks Kenneth Chien, Gerald Crabtree and the other, anonymous, reviewer for their contribution to the peer review of this work. Peer reviewer reports are available.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41556-021-00668-z.

with the SWI/SNF complex, it increases chromatin accessibility and transcription factor binding at these sites. Furthermore, PHF7 recruits cardiac transcription factors to activate a positive transcriptional autoregulatory circuit in reprogramming. Importantly, PHF7 achieves efficient reprogramming in the absence of Gata4. Here, we highlight the underexplored necessity of cardiac epigenetic readers, such as PHF7, in harnessing chromatin remodelling and transcriptional complexes to overcome critical barriers to direct cardiac reprogramming.

In recent years, direct reprogramming of resident cardiac fibroblasts (CFs) to induced cardiac-like myocytes (iCLMs) has emerged as a promising therapeutic strategy to induce remuscularization of the injured heart. However, epigenetic barriers severely limit the conversion efficiency of adult fibroblasts, thus constraining the utility of this approach. In development, chromatin remodelling complexes function in a highly cell-specific manner to dictate cellular fate by either restructuring the nucleosome or directly modifying histones[9]. In cardiovascular tissues, the SWI/SNF complex in particular interacts synergistically with cardiac transcription factors (TFs) to dictate chromatin accessibility and cardiac gene expression[10–14]. Cardiac chromatin remodelling complexes conceivably hold the potential to redefine the epigenetic landscape of somatic cells, yet overexpression of select remodellers has shown little or no benefit[1,15]. Despite the central influence of epigenomic structure on reprogramming, the enhancer landscape of iCLMs has only recently been defined, and limited efforts have been dedicated to enhancing our understanding of the role of chromatin regulatory factors and complexes in mediating this cellular process[16–20].

Adult fibroblast reprogramming with Gata4, Mef2c and Tbx5 (GMT), GMT plus Hand2 (GHMT), or GMT plus Hand2 and Akt1 (AGHMT) converts only 0.5–3% of adult tail-tip fibroblasts (TTFs) to iCLMs[1,3,8,20]. To identify potential modifiers of the barriers to reprogramming, we conducted a screen of more than 1,000 retroviruses encoding TFs and epigenetic regulators in AGHMT-induced reprogramming of adult TTFs[8]. From this screen, we identified the histone reader PHF7 as the strongest activator of direct cardiac reprogramming. A critical regulator of germline cell fate specification, PHF7 binds directly to histone H3 and to H3K4me3 and H3K4me2 marks[21–23]. However, its effect on chromatin structure and its role in cardiac biology remain undefined. Through interrogation of single-cell Mesp1+ cardiac progenitor analyses from embryonic day 6.5 (E6.5) to E7.5 murine tissues, we observed robust and sustained expression of *Phf7* in early cardiac progenitors, which suggests that Phf7 plays an endogenous role in cardiac stem cell biology[24] (Extended Data Fig. 1a).

We confirmed that PHF7 robustly induced reprogramming, as measured by activation of the α-myosin heavy chain promoter fused to a green fluorescent protein reporter (α-MHC–GFP) and cardiac troponin T (cTnT) expression in TTFs derived from adult mice bearing a transgene encoding α-MHC–GFP (Fig. 1a–c and Extended Data Fig. 1b). PHF7 enhanced direct reprogramming of fibroblasts derived from diverse embryonic and adult tissues, thereby confirming the cardiogenic potential of this factor (Fig. 1b,d,e). Flow cytometry demonstrated that the addition of PHF7 to AGHMT-infected TTFs generated ~20% double-positive (cTnT+ and α-MHC–GFP+) iCLMs, which represents a ten-fold increase in reprogramming efficiency above AGHMT alone (~2% double-positive cells) (Fig. 1f,g and

Extended Data Fig. 1c,d). Robust activation of reprogramming in PHF7-treated cells was observed as early as 1 day following induction (Fig. 1h). The addition of PHF7 enriched cardiomyocyte markers, including *Actc1, Myh6* and *Tnni3* (Extended Data Fig. 1e). Consistent with these findings, PHF7 increased the functional maturity of reprogrammed cells by accelerating and increasing spontaneous beating of mouse embryonic fibroblast (MEF) iCLMs by approximately two-fold and increasing the number of calcium fluxing iCLMs by ~1.5-fold (Fig. 1i,j and Supplementary Videos 1–3). Of paramount importance for therapeutic translation, PHF7 augmented the reprogramming of adult cardiac human fibroblasts, the most resistant cell type to reprogramming, by approximately three-fold to four-fold when added to a human reprogramming cocktail containing Myocardin and GHMT (My-GHMT) (Fig. 1d,e and Extended Data Fig. 1f,g). Beyond this, transient PHF7 expression achieved through a doxycycline-inducible system was sufficient to activate and maintain the conversion of adult TTF iCLMs, as evidenced by cardiac gene expression, sarcomere formation, spontaneous beating and calcium fluxing (Extended Data Fig. 2a–j and Supplementary Video 4). Together, these results suggest that this factor has therapeutic potential.

To decipher the effect of PHF7 on reprogramming globally, we performed RNA sequencing (RNA-seq) using adult TTFs reprogrammed for 7 days with either AGHMT plus empty vector or AGHMT plus PHF7 (Fig. 2a). PHF7 induced extensive transcriptomic changes, dysregulating the expression of over 700 genes, with ~500 upregulated and ~200 downregulated transcripts (fold-change (FC) > 2, false discovery rate (FDR) < 0.01) (Fig. 2b). Among the genes most upregulated by PHF7 were those encoding mature cardiac structural proteins such as *Ttn*, *Actc1*, *Actn2*, *Tcap*, *Tnni3* and *Tnni1*, as well as known critical cardiac developmental control genes such as *Tbx20*, *Smyd1* and *Myocd* (Fig. 2c and Extended Data Fig. 3a,b). All four reprogramming TFs (*Gata4, Hand2, Mef2c* and *Tbx5*) were significantly upregulated by PHF7 (Fig. 2c and Extended Data Fig. 3c,d). Examination of the 3′ untranslated regions (UTRs) of these factors confirmed that the increase in transcription originated from the endogenous loci (Extended Data Fig. 3e). Gene ontology (GO)-enrichment analysis revealed that PHF7 specifically upregulated genes associated with cardiac development and contractility (Fig. 2d). Downregulated GO terms were related to inflammatory, chemotactic, ERK and tumour necrosis factor signalling pathways, which are known to present barriers to direct cell fate conversion (Fig. 2e). More specifically, gene set enrichment analyses (GSEA) revealed an upregulation of myogenic targets and downregulation of E2F, Myc, G2M and interleukin-6–Jak–Stat targets, which represent proliferative pathways that are known to inhibit the reprogramming process (Extended Data Fig. 3f,g). However, EdU pulse labelling of TTF iCLMs at day 1 post-induction revealed no significant difference in the number of proliferating cells with the addition of PHF7, which suggests that there is an alternative mechanism by which PHF7 potently activates reprogramming (Extended Data Fig. 3h,i).

Given the marked activation of the cardiac transcriptome by PHF7, we hypothesized that PHF7 could activate reprogramming using fewer factors. Applying PHF7 or empty vector to both TTFs and MEFs in the presence of GMT, GHMT or AGHMT, we observed that PHF7 induced reprogramming to a similar extent in the context of all three cocktails (Extended Data Fig. 4a–c). We then tested PHF7 with various combinations of factors and found that

PHF7 induced reprogramming in the absence of Gata4; that is, with Mef2c and Tbx5 (MT) alone (Fig. 3a). This PHF7–Mef2c–Tbx5 cocktail, herein referred to as PMT, generated a substantial α-MHC–GFP+ population (~10–20% of total cells) at 7 days post-induction, which was confirmed by both immunocytochemistry and flow cytometry (Fig. 3b,c and Extended Data Fig. 4b). The addition of PHF7 to MT induced cTnT expression, as demonstrated by immunocytochemistry, flow cytometry and quantitative PCR (qPCR) (Fig. 3a–d). Furthermore, the cardiac transcripts *Actc1* and *Nppb* were enriched in the presence of PHF7 (Fig. 3d). Addition of PHF7 to MT induced a similar increase in the transcription of endogenous Gata4 as was observed following the addition of PHF7 to AGHMT (Fig. 3e). Finally, the PMT cocktail induced occasional spontaneous calcium sparks at 14 days post-induction, which suggests that these iCLMs have functional potential (Supplementary Video 5).

To define the mechanism by which PHF7 augments reprogramming, we identified genome-wide binding sites of PHF7 using chromatin immunoprecipitation (ChIP) followed by massively paralleled deep sequencing (ChIP-seq). MEFs were infected with epitope-tagged PHF7^{3xTy1} alone or with AGHMT reprogramming factors (AGHMT + PHF7^{3xTy1}) and collected 2 days following induction for ChIP analyses (Fig. 4a and Extended Data Fig. 5a,b). Differential peak analysis showed that PHF7 occupied dramatically different genomic loci in the setting of reprogramming factors, with 12,710 peaks present and 22,369 peaks absent in the setting of AGHMT (Fig. 4b). For both conditions, PHF7 bound to predominantly intergenic and intronic genomic regions, which is consistent with TF binding patterns and accession of regulatory elements (Extended Data Fig. 5c,d). Peaks enriched in the presence of PHF7 alone had GO pathways related to cell–cell adhesion and promoter-binding processes, and motif analyses revealed enrichment of CTCFL, AP-1 and Runx motifs (Extended Data Fig. 5e,g). In the context of reprogramming factors, GO biological process terms demonstrated strong recruitment of PHF7 to genomic regions regulating cardiac development and function (Extended Data Fig. 5f). Searching for motifs in these PHF7 reprogramming peaks revealed enrichment of Gata4, AP-1, TEAD and Mef2 motifs (Fig. 4c). Evaluating the motif frequency among peak clusters confirmed the enrichment of Gata4, Mef2c and Hand2 motifs specifically in PHF7 reprogramming peaks, which implies that there is substantial co-occupancy of these TFs with PHF7 during reprogramming (Extended Data Fig. 5h).

Given the results of the motif-enrichment analyses, we interrogated the degree of genome-wide co-occupancy of PHF7 with reprogramming TFs in AGHMT-induced iCLMs and observed remarkable colocalization of all four cardiac TFs at these PHF7-reprogramming peak regions[16] (Fig. 4d and Extended Data Fig. 5i). Observing the high-fidelity binding of PHF7 to multivalent cardiac-lineage-defining TF loci, we hypothesized that PHF7 localized to cardiac super enhancers (SEs) and cooperated with cardiac TFs as a core transcriptional complex to dictate cardiac cell identity[25]. To address this hypothesis, we called SEs using H3K27ac ChIP obtained from postnatal day 4 (P4) mouse ventricle and annotated 1,251 cardiac SEs by ranking H3K27ac signal intensity (Fig. 4e and Supplementary Table 1). By mapping the PHF7^{3xTy1} ChIP signal globally across cardiac SEs, we observed PHF7 binding events at all 1,251 SE regions, regardless of the presence of reprogramming factors (Fig. 4f and Supplementary Table 1). However, binding of PHF7 to cardiac SEs was significantly

enriched in the presence of AGHMT (Fig. 4f). We hypothesized that PHF7 localized to these cardiac regulatory regions in fibroblasts through its known ability to read and directly bind H3K4me2 and H3K4me3 modifications[21,23]. To address this hypothesis, we aligned histone ChIP in MEFs and found that PHF7 colocalized with H3K4me2, a modification that demarcates TF-binding sites[26,27]. PHF7 also aligned with the activating modifications H3K4me3, H3K27ac and H3K79me2, but was absent from sites with repressive H3K27me3 marks (Extended Data Fig. 6a). Mapping MEF histone ChIP signals globally across P4 heart enhancers showed that H3K4me2 marked cardiac enhancers in MEFs. This result suggests that this is the mechanism by which PHF7 identifies cardiac regulatory regions in fibroblasts (Fig. 4g and Extended Data Fig. 6b).

Strong PHF7 binding was observed at the MHC SE (*Myh6/Myh7*), a region known to govern cardiac myosin isoform switching and cardiac gene programmes through recruitment of the SWI/SNF complex[13] (Fig. 4e,h). We identified multivalent cardiac TF binding and modest H3K27ac deposition at this region in the absence of PHF7 (Fig. 4h). The addition of PHF7 markedly increased *Myh6/Myh7* enhancer activation, as measured by the level of H3K27ac deposition. Thus, we identified a critical barrier through which PHF7 promotes reprogramming (Fig. 4l). We then validated activation of the *Myh6* enhancer (*Myh6*^{enh}) as well as other enhancers (*Tns1*^{enh}, *Thra*^{enh} and *Tbx20*^{enh}) identified from an in silico SE analysis through the generation of enhancer–hsp68–mCherry retroviral constructs. Delivery of these reporter constructs with AGHMT factors induced mCherry expression, and the addition of PHF7 further increased the activation of these cardiac enhancer elements (Fig. 4i and Extended Data Fig. 7a–c).

Previous studies demonstrated that Gata4, Hand2, Mef2c and Tbx5 are core TFs that can bind to both self and one another at their respective SEs, thereby creating a core transcriptional regulatory circuit[28,29]. PHF7 and cardiac TF binding events were identified at SEs annotated to the core reprogramming TFs themselves (Extended Data Fig. 8a–e and Supplementary Table 1). Furthermore, a mouse P4 cardiac SE was annotated to *Phf7* (rank 331 out of 1,251), and we identified PHF7 binding to its own SE at sites co-bound by cardiac TFs (Extended Data Fig. 8f and Supplementary Table 1). TF ChIP from in vivo P4 mouse ventricle further identified endogenous Gata4, Tbx5 and Nkx2–5 binding throughout the *Phf7* transcription start site (TSS) and the SE, which suggests that this regulatory network persists in vivo (Extended Data Fig. 8g). Together, our ChIP and transcriptomic data suggest that PHF7 participates with Gata4, Hand2, Mef2c and Tbx5 in a cardiac TF autoregulatory circuit.

To validate PHF7 co-occupancy with cardiac reprogramming factors, we performed immunoprecipitation assays to examine interactions between PHF7 and cardiac TFs. We observed strong interactions between PHF7$^{Flag\text{-}HA}$ and Gata4^{Myc}, Hand2^{Myc} and Mef2c^{Myc} by Flag-tagged immunoprecipitation (Fig. 4j and Extended Data Fig. 8h). Given these interactions, we hypothesized that PHF7 influences the binding of these cardiac TFs to cardiac enhancers. Gata4 ChIP showed that PHF7 strengthened the binding of Gata4 to both itself and its targets, which reinforces the observed positive autoregulatory circuit (Fig. 4k). Furthermore, PHF7 activated these cardiac enhancers as demonstrated by H3K27ac ChIP (Fig. 4l).

We hypothesized that these mechanisms were maintained in the absence of Gata4 and could account for the enhanced reprogramming efficiency observed using the PMT cocktail. To explore this hypothesis, we performed PHF7^{3xTy1} ChIP in the context of reprogramming with PHF7^{3xTy1}, Mef2c and Tbx5 (PMT) and identified strong binding of PHF7^{3xTy1} to the *Myh6*^{enh} and the *Gata4* TSS (Extended Data Fig. 8i,j). As we had observed an interaction between Mef2c and PHF7, we hypothesized that PHF7 also affects the strength of Mef2c binding to its targets. Mef2c^{3xTy1} ChIP showed that PHF7 strengthened the binding of Mef2c^{3xTy1} to the *Myh6*^{enh} and the *Gata4* TSS in the context of PMT compared with MT alone (Extended Data Fig. 8i,j). The addition of PHF7 to MT increased the deposition of active enhancer marks at the *Myh6* and *Gata4* loci, as determined by H3K27ac ChIP (Extended Data Fig. 8i,j). These data suggest that PHF7 promotes the activation of cardiac enhancers, increases cardiac TF binding and activates a core cardiac autoregulatory circuit in the absence of the pioneer factor Gata4.

To assess global changes in chromatin accessibility, we performed assay for transposase-accessible chromatin using sequencing (ATAC-seq) in day 2 AGHMT-induced MEF iCLMs in the presence and absence of PHF7. PHF7 induced broad changes in chromatin architecture, yielding 5,808 peaks with increased chromatin accessibility and 3,774 peaks with decreased accessibility (FC > 2) (Fig. 5a). GO pathway analysis showed that open chromatin peaks were annotated to regions regulating cyclic-AMP-mediated signalling and cardiac morphogenesis (Fig. 5b). These accessible regions were enriched for AP-1, CTCF and TEAD motifs by de novo motif analysis (Fig. 5d). Closed chromatin peaks were related to Wnt signalling, a pathway prohibitive to direct reprogramming (Fig. 5c).

Given the ability of PHF7 to modify chromatin structure, we hypothesized that PHF7 recruits remodelling complexes to enact changes in accessibility. To identify interacting partners, we utilized a miniTurbo biotinylation-based proximity ligation assay, infecting MEFs with either PHF7^{mTurbo} or AGHMT + PHF7^{mTurbo}. Among the proteomics hits shared between PHF7^{mTurbo} and AGHMT + PHF7^{mTurbo} conditions, we identified several histones (histone H2a and histone H3.2) and heterochromatin binding proteins (Hp1bp3, Mybbp1a and Las1l) (Extended Data Fig. 9a and Supplementary Table 2). However, the top hit in both conditions was SMARCD3 (also known as BAF60c), a SWI/SNF complex subunit known to orchestrate cardiac development and disrupt nucleosomal stability to regulate cardiac TF binding, proliferation and Wnt signalling[10,30,31]. Through Flag co-immunoprecipitation, we observed a strong interaction between PHF7^{3xTy1} and SMARCD3$^{Flag-HA}$ (Fig. 5e). Shared hits were enriched in ribonucleoprotein complex factors, while interacting proteins identified as unique to reprogramming conditions were almost exclusively cardiac and muscle-specific factors and included the reprogramming factor Gata4 (Extended Data Fig. 9b–d and Supplementary Table 2). These results are consistent with the known ability of SMARCD3 to recruit and interact with Gata4 and other cardiac TFs[10,14].

We were interested in whether the observed interaction between PHF7 and SMARCD3 was a causal mediator of the reprogramming phenotype. Forced overexpression of retroviral SMARCD3 in AGHMT reprogramming neither phenocopied PHF7 nor augmented its effect (Extended Data Fig. 9e). Furthermore, knockdown of *Smarcd3* in AGHMT TTF

reprogramming did not decrease reprogramming efficiency (Fig. 5g,h and Extended Data Fig. 9f). However, when a short hairpin RNA (shRNA) targeting *Smarcd3* (shSmarcd3) was applied to AGHMT + PHF7 reprogramming, knockdown of *Smarcd3* significantly attenuated the effect of PHF7, yielding a reprogramming efficiency closer to that of AGHMT (Fig. 5g,h). These findings suggest that SMARCD3 utilizes a histone reader such as PHF7 to identify cardiogenic sites and exert its chromatin-modifying effects in fibroblasts (Fig. 5i).

While substantial attention has been focused on TF biology in cardiac reprogramming, our understanding of and ability to harness epigenetic regulatory mechanisms remain in their infancy. Thus far, few epigenetic regulators have been shown to affect direct reprogramming. Knockdown of the polycomb complex protein Bmi1 can enhance reprogramming and permit cell fate conversion in the presence of MT[20]. Overexpression of SMARCD3 alongside Gata4 and Tbx5 is critical in mesodermal differentiation to cardiomyocyte-like cells[14]. However, overexpression of SWI/SNF and other cardiac-specific epigenetic regulatory complexes have shown limited or no benefit in fibroblast reprogramming[8,15].

In this study, we defined the histone reader PHF7 as a critical factor in overcoming barriers of direct reprogramming through its ability to stabilize TF binding at cardiac enhancers. PHF7 recognized cardiac SEs in fibroblasts in the absence of reprogramming factors, which suggests that this reader is involved in the earliest steps of cardiac enhancer recognition and activation. Notably, the addition of reprogramming factors enhanced the recruitment of PHF7 to cardiac SEs. Through cooperation with the SWI/SNF complex, PHF7 increased chromatin accessibility at these SEs, thereby informing the expression of cardiac gene programmes. Furthermore, PHF7 participated in and activated a cardiac TF autoregulatory circuit, regulating endogenous transcription of the core TFs themselves. Consistent with its function as a histone reader recruiting the SWI/SNF complex, PHF7 was dependent on endogenous SMARCD3 expression for its phenotype in reprogramming. Conceivably, PHF7 may facilitate SWI/SNF eviction of the polycomb repressive complex from cardiac enhancers, potentially linking the mechanisms observed with Bmi1 knockdown and PHF7 overexpression[20,32]. This is a hypothesis deserving of further investigation.

Minimizing factors for therapeutic administration has been of great interest to the reprogramming field. We showed that PHF7 enhanced the efficiency of all TF cocktails and accomplished reprogramming in the absence of the pioneer factor Gata4. In this setting, PHF7 not only strongly strengthened Mef2c binding to its targets but also recruited Mef2c to induce activation of the TF autoregulatory circuit in the absence of exogenous Gata4 (ref. [16]). Through these mechanisms, PHF7 rapidly overcomes epigenetic barriers to reprogramming at cardiac regulatory elements that additional cardiac TFs fail to efficiently achieve on their own.

Direct cardiac reprogramming holds rich potential for therapeutic translation, but numerous roadblocks persist, the most dominant of which involves efficient rewiring of the human fibroblast epigenome to a cardiac fate. TFs are powerful in exacting these cell fate changes, yet we know from developmental studies that their effects are often intricately interdependent on chromatin remodelling machinery[9,11,33]. In spite of this, relatively little is

known regarding the potential for chromatin remodellers in direct cardiac reprogramming. Our studies highlight the therapeutic potential of PHF7 through its ability to rapidly and persistently achieve cardiac cell fate conversion in adult and human fibroblasts. We believe that the identification and definition of epigenetic factors, such as PHF7, merits further investigation to not only shed light on the mechanistic basis of cardiac reprogramming but to also more rapidly propel these technologies past the epigenomic barriers of human cells to the clinic.

Methods

Mice.

All experiments involving animals were approved by the Institutional Animal Care and Use Committee at the University of Texas Southwestern Medical Center. All mice used in this study were housed at the Animal Resource Center at the University of Texas Southwestern Medical Center. All animals were bred inside a specific pathogen-free facility with 12-h light–dark cycles and monitored daily; no health problems were reported. All animals were housed in groups of maximum five per cage with ad libitum access to food and water. The temperature and humidity of all animal rooms is electronically monitored and regulated. α-MHC–GFP mice were maintained on a C57BL/6 background[3].

Isolation and culture of mouse fibroblasts.

Adult mouse TTFs and CFs from 4–6-week-old male and female C57BL/6 or α-MHC–GFP mice were prepared as previously described[8] and cultured in fibroblast growth medium (DMEM with 10% fetal bovine serum (FBS), 1% penicillin–streptomycin). MEFs derived from embryos of C57BL/6 or α-MHC–GFP timed pregnant females were collected at E13.5–14.5 and were prepared as previously described[34]. All primary cells were maintained in DMEM with 10% FBS and 1% penicillin–streptomycin.

Retrovirus production and cardiac reprogramming.

Construction of pMXs retroviral expression vectors encoding Gata4, Hand2, Mef2c, Tbx5, Akt1 and PHF7 has been previously described[3,4,8]. Retroviral constructs of shRNA targeting *Smarcd3* and scramble sequence 5′-CTACACAAATCAGCGATTT cgaaAAATCGCTGATTTGTGTAG-3′ and shSmarcd3 sequence 5′-GGTGGAATT CAGTGGTCAAGAcgaaTCTTGACCACTGAATTCCACC-3′ were cloned into an entry vector using a BLOCK-iT U6 Entry Vector kit (Thermo Scientific) and cloned into a pMXs-GW vector by Gateway cloning. pMXs-GW was a gift from S. Yamanaka (Addgene, plasmid 18656)[35]. Retroviruses were generated using Platinum E cells as previously described[8]. Briefly, retroviral expression vectors were transfected into Platinum E cells using FuGENE 6. Twenty-four hours following transfection, fibroblasts were seeded onto precoated plates with SureCoat (Cellutron) or Matrigel (Corning). Forty-eight hours after transfection, fibroblasts were infected by replacing growth medium with the indicated virus supplemented with polybrene at 8 μg ml^{-1}. Viral infection was repeated after 24 h. Twenty-four hours following the second infection, the virus medium was replaced with induction medium composed of DMEM/199 (4:1), 10% FBS, 5% horse serum, 1% penicillin–streptomycin, 1% B-27, 1% insulin–selenium–transferrin, 1% nonessential amino acids, 1%

essential amino acids, 1% vitamin mixture and 1% sodium pyruvate (Invitrogen), which was replaced every 2 days. For human CF reprogramming, 1 day after viral transduction (day 1), the virus medium was replaced by induction medium composed of DMEM/199 (4:1), 10% FBS, 1% nonessential amino acids, 1% penicillin–streptomycin every 2 days until day 4. On day 4, the medium was changed to 75% induction medium and 25% RPMI + B27. On day 7, the medium was changed to 50% iCM and 50% RPMI + B27. On day 11, the medium was changed to 25% iCM and 75% RPMI + B27. On day 14, the medium was changed to RPMI + B27 + FFV (10 ng ml^{-1} rhFGF, 15 ng ml^{-1} rhFGF-10 and 5 ng ml^{-1} rhVEGF) every other day until day 21. Immunostaining and gene expression analysis were performed using day 21 samples.

Quantitative mRNA measurement.

Total RNA was extracted using TRIzol (Invitrogen) according to the vendor's protocol and was reverse-transcribed to complementary DNA using iScript Supermix (Bio-Rad). qPCR was performed using KAPA SYBR Fast (Kapa Biosystems). Gene expression analysis was performed using the Ct method. qPCR was performed using Taq-man probes (Applied Biosystems, Mm03053570_s1) for *Gata4* targeted at the 3′ UTR. The following SYBR primer pairs were used: *Myh6* Fw 5′-GCC CAG TAC CTC CGA AAG TC-3′ and Rv 5′-GCC TTA ACA TAC TCC TTG TC-3′; *Tnnt2* Fw 5′-GTA GAG GAC ACC AAA CCC AAG-3′ and Rv 5′- GAG TCT GTA GCT CAT TCA GGT C-3′; *Actc1* Fw 5′-CGG ACA ATT TCA CGT TCA GCA-3′ and Rv 5′-CTG GAT TCT GGC GAT GGT GTA-3′; *Tnni3* Fw 5′-TCT GCC AAC TAC CGA GCC TAT-3′ and Rv 5′-CTC TTC TGC CTC TCG TTC CAT-3′; *Nppb* Fw 5′-GAGGTCACTCCTATCCTCTGG-3′ and Rv 5′-GCCATTTCCTCCGACTTTTCTC-3′; *PHF7* Fw 5′-GACTAGGAG GGTAACCCAGAG-3′ and Rv 5′-TGCACGCTGATATTGTCTTTCT-3′; and *Smarcd3* Fw 5′-CCC GAG TCC CAG GCT TAC A-3′ and Rv 5′-GCT TTC GCT TTT GCT TCA TGG-3′. For input normalization, we used 18S (Applied Biosystems, 4319413E) or *Gapdh* Fw 5′-AGG TCG GTG TGA ACG GAT TTG-3′ and Rv 5′-TGT AGA CCA TGT AGT TGA GGT CA-3′.

Immunocytochemistry.

Immunocytochemistry was performed as previously described[4]. Briefly, cells were fixed in 4% paraformaldehyde (PFA) for 15 min at room temperature, permeabilized with 0.1% Triton X-100 and blocked with 10% goat serum. Treated cells were then incubated with mouse monoclonal anti-Tnnt2 antibody (1:500; Thermo Scientific, MA5–12960), rabbit anti-GFP antibody (1:500; Thermo Scientific, A-11122) or mouse anti-α-actinin (1:500; Sigma, A7811) in 1% goat serum at 4 °C overnight. After three washes with PBS, cells were incubated with Alexa Fluor secondary antibodies (1:500; Invitrogen) for 1 h at room temperature. Image acquisition was performed using a BZ-X710 or BZ-X800 (Keyence). Cells were manually quantified in ten randomly selected low-power fields of view from each well in three independent experiments. Results were then averaged to yield an individual replicate.

Western blot analysis.

Western blot analyses were performed as previously described[3]. Briefly, cell lysates were prepared using RIPA buffer with complete phosphatase and protease inhibitor cocktail

tablets (Roche) and boiled in Laemmli buffer for 5 min at 95 °C. Antibodies utilized were anti-Mef2c antibody (1:1,000; Cell Signaling, 5030), anti-Gata4 antibody (1:500; Santa Cruz, sc25310), anti-Tnnt2 antibody (1:500; Thermo Scientific, MA5–12960), anti-GFP antibody (1:500; Thermo Scientific, A-11122), anti-PHF7 antibody (1:500; LSBio, B11090), anti-Ty1 antibody (1:1,000; Diagenode, C15200054), anti-Myc (1:1,000; Invitrogen, 46–0603), anti-Flag (1:1,000; Sigma, F7425), anti-mCherry antibody (1:1,000; Abcam, ab167453) and anti-GAPDH antibody (1:1,000; Merck Millipore, MAB374).

Flow cytometry.

Flow cytometry was performed as previously described[8]. Briefly, reprogrammed iCLMs were trypsinized using TrypLE Express (Gibco) and resuspended into a single-cell suspension. Then, cells were fixed and permeabilized using BD Cytofix/Cytoperm (BD Biosciences). Cells were stained with rabbit anti-GFP antibody (1:200; Thermo Scientific, A-11122) and mouse monoclonal anti-Tnnt2 antibody (1:200; Thermo Scientific, MA5–12960) and secondarily stained using donkey anti-mouse Alexa Fluor 647 (1:200; Invitrogen, A-31571) and goat anti-rabbit Alexa Fluor 488 (1:200; Invitrogen, A-11008). Flow cytometry was performed using a FACSCalibur instrument (BD Biosciences) and analysis was performed using FlowJo software.

Beating cell analysis and calcium assay.

Spontaneous-beating cell assays were performed as previously described[4]. Cells were reprogrammed on a Matrigel-coated dish (Corning, 354248). Beating cells were manually counted in ten randomly selected high-power fields per well in three independent experiments. Spontaneous calcium flux assays were performed using a Fluo-4 NW Calcium Assay kit (Thermo Scientific, F36206) as previously described[4]. Calcium flux was assessed according to the manufacturer's protocol on fibroblasts at indicated times following reprogramming induction. After replacing culture medium with the loading solution, plates were incubated at 37 °C for 30 min. They were then incubated at room temperature for 30 min immediately before visualization and measurement. Calcium-flux-positive cells were manually counted in ten randomly selected high-power fields per well and averaged to yield a single replicate, across three independent experiments.

Inducible reprogramming assays.

PHF7 expression was temporally induced and studied in reprogramming using a Retro-X Tet-One Inducible Expression System (Takara, 634304). An inducible PHF7 expression cassette was generated by cloning pRetro-X-PHF7 using the manufacturer's protocol. pRetroX-PHF7 was then retrovirally delivered to TTFs and MEFs with reprogramming factors. Doxycycline was added to doxycycline treatment groups at 1 μg ml^{-1} following infection and to reprogramming induction medium. Medium with doxycycline was changed every 2 days. Control groups included no doxycycline (no dox) and sustained doxycycline (D1 on) exposure, while treatment groups included removal of doxycycline at days 3 and 10 (D3 off and D10 off, respectively) following the addition of reprogramming induction medium. Gene and protein expression, calcium flux and beating cells were assessed either throughout or at day 28 post-induction medium, as indicated.

EdU labelling.

At day 1 post-induction medium treatment, adult TTF iCLMs were treated with 10 μM EdU (Lumiprobe, 10540) for 4 h. After that, cells were fixed with 4% PFA at room temperature for 10 min, permeabilized with 0.3% Triton X-100 in PBS, followed by EdU staining by click chemistry (a label mix containing 8 μM sulfo-Cy5-azide, 2 mM $CuSO_4 \cdot 5H_2O$ and 20 mg ml^{-1} ascorbic acid in PBS was applied to cells for 30 min). Cells were washed and nuclei were counterstained with Hoechst (1:2,000). For quantification, the percentage of EdU+ cells were manually quantified and averaged to yield an individual replicate in ten randomly selected low-power fields of view from each well in three independent experiments.

RNA-seq sample preparation.

Total RNA was extracted from TTFs 7 days after retroviral transduction using TRIzol (Invitrogen) according to the vendor's protocol and prepared for RNA-seq. Illumina RNA-seq was performed by the University of Texas Southwestern Microarray Core Facility.

ChIP-seq sample preparation.

For ChIP-seq sample preparation, MEFs or TTFs were prepared as previously described[16]. Briefly, 2 days after retroviral transduction, MEFs or TTFs were crosslinked with 1% formaldehyde in PBS for 15 min and neutralized by the addition of glycine to a final concentration of 0.125 M for 5 min. TTFs or MEFs were then collected and washed with cold PBS for ChIP. ChIP was then performed using ChIP-IT Express kits (Active Motif) following the vendor's protocol. In brief, cell lysates were sonicated (ten cycles of 30 s on/off) to shear DNA by using a Bioruptor Pico sonicator (Diagenode, B01060010). Chromatin was incubated with indicated antibodies overnight at 4 °C. The following antibodies were used for ChIP experiments: anti-Gata4 antibody (Santa Cruz Biotechnology, sc-1237), anti-Ty1 antibody (Diagenode, C15200054) and anti-H3K27ac antibody (Diagenode, C15410196). ChIP-seq libraries were generated using a KAPA Hyper Prep kit following the manufacturer's protocol (Kapa Biosystems), and single-end sequenced on an Illumina NextSeq500 system using a 75-bp high-output sequencing kit. Massive parallel sequencing was performed at the University of Texas Southwestern Next Generation Sequencing Core Facility. For ChIP–qPCR, chromatin fragments were analysed by qPCR using SYBR green fluorescence using the following primer sequences that had been previously validated: *Gata4* TSS: Fw 5′-CTG GGT AGG GGC TGG AGT AG-3′, Rev 5′-CTG GCC GAG AGC AGT ACG-3′; *Myh6* promoter Fw 5′-GCA GAT AGC CAG GGT TGA AA-3′ Rev 5′-TGG GTA AGG GTC ACC TCC TC-3′; and *Tbx5* Fw 5′-GCG AAG GGA TGT TTC AGC AC, Rev 5′-CAC GCC GTG AGT GTA GAG AA-3′.

ATAC-seq sample preparation.

ATAC-seq was performed as per the Omni-Seq protocol described by Corces et al.[36]. All ATAC-seq experiments were performed using 50,000 cells. Multiplexed paired-end 75-bp sequencing was performed using Illumina HiSeq 2500 using Nextera-compatible amplification primers.

Immunoprecipitation assays.

PHF7$^{Flag\text{-}HA}$ was co-expressed with GFP, Gata4^{Myc}, Hand2^{Myc} or Mef2c^{Myc}, or SMARCD3$^{Flag\text{-}HA}$ was co-expressed with PHF^{3xTy1} or GFP in HEK293 cells for 72 h. Experiments involving crosslinking were adapted from Zlatic et al.[37], whereby cells were treated with 200 μM dithiobis(succinimid ylpropionate) (DSP) or dimethylsulfoxide (DMSO) for 4 h at 4 °C and quenched with 25 mM Tris for 15 min. Pre-cleared lysates were incubated with Flag magnetic beads (Sigma, M8823) overnight. The Flag epitope tag was eluted using 0.5 mg ml^{-1} free 3× Flag peptide (Sigma). The final elution and input obtained before immunoprecipitation were analysed by western blotting using an anti-Myc (Invitrogen), anti-Ty1 (Diagenode) or rabbit anti-Flag antibody (Sigma).

In vitro transgenic reporter assays.

Putative enhancers were cloned into a hsp68–mCherry expression vector (Supplementary Table 3). Expression cassettes were cloned into pMXs retroviral vectors via Gateway cloning. MEFs were infected with pMXs-enhancer–hsp68–mCherry constructs along with reprogramming factors. Following reprogramming induction, immunostaining was performed for mCherry and Hoechst at indicated time points in iCLMs. For quantification, the percentage of mCherry+ cells was manually quantified and averaged to yield an individual replicate in ten randomly selected low-power fields of view from each well in three independent experiments. The genomic coordinates of all enhancers are listed in Supplementary Table 3.

miniTurbo BioID assay.

Proximity biotinylation (BioID) was adapted from Branon et al.[38]. Briefly, MEFs were infected with empty vector, pMXs-puro-PHF7-miniTurbo or AGHMT + pMXs-puro-PHF7-miniTurbo and exposed to reprogramming induction medium as described above for 7 days. Cells were then exposed to 200 μM biotin for 4 h or no biotin control. Cell lysates were extracted in 1 ml of lysis buffer (6 M urea and 10% SDS, supplemented with protease inhibitor) and mechanically lysed. Lysates were added to streptavidin magnetic beads and rotated for 24 h at 4 °C (Thermo Fisher, 88816), with final elution by boiling at 95 °C for 5 min. Pulldown was assessed by silver staining (Thermo Fisher, LC6070) and peptide identification was performed by the Proteomics Core Facility at the University of Texas Southwestern Medical Center. Data were normalized to the total protein submitted per sample as well as to empty vector/biotin negative control. Proteomics hits were validated by western blotting of the submitted sample.

Bioinformatics and computational analysis.

RNA-seq analysis. RNA-seq analyses were performed as described in Zhou et al.[8]. Briefly, reads were aligned to the mouse reference genome GRCm38 (mm10) using the Hisat (v.2.0.0) aligner using default settings. Aligned reads were counted using featureCounts (v.1.4.6) per gene ID. Differential gene expression analysis was done using the R package edgeR (v.3.8.6). Cut-off values of FC > 2 and FDR < 0.01 were used to select for differentially expressed genes. The DAVID gene functional annotation and classification tool was used to annotate a list of differentially expressed genes. GO analysis was performed to

determine biological functional categories, and enrichment plots were performed using GSEA software.

ChIP-seq analysis.

Raw sequencing reads with >30% nucleotide with phred quality scores <20 were filtered. Single-end sequencing reads were then aligned to the mouse reference genome GRCm38 (mm10) using bowtie2 aligner (v.2.3.4.3) with default parameters. For TF ChIP-seq data, peaks were called using the HOMER software package (v.4.9) findpeaks command, with the parameter '-style factor', peaks were called with more than twofold enrichment over the input controls and more than fourfold enrichment over the local tag count, and the FDR threshold was set to 10^{-3}. For histone marker ChIP-seq data, peaks were called using the findpeaks command with the parameter '–style histone', peaks were called with more than twofold enrichment over the input controls and more than fourfold enrichment over the local tag counts, and the FDR threshold set to 10^{-3}. ChIP-seq peaks within a 1,000-bp range were stitched together to form broad regions. To identify differential peaks between two samples, called peaks were merged from each sample and the raw read count within each peak region was calculated. Then, differential peaks were identified using the R package DEseq (v.3.8). PHF7 peaks with more than twofold change were designated as DE peaks. To identify potential TF motifs enriched in peaks-of-interest regions, we used the findMotifsGenome. pl command from the HOMER software package, using peak region.bed file as the input, with parameter '–size 200' to search for motif enrichment in a 200-bp window surrounding the peak centre. To analyse the functional significance of bound peaks, genomic regions enrichment of annotations tool (GREAT) was used with mm10 as the background genome and other parameters set as default. To annotate SEs, the algorithm ranking of super-enhancer (ROSE) was used (http://younglab.wi.mit.edu/super_enhancer_code.html) with H3K27ac ChIP-seq data from P4 mouse ventricle cells as input. A custom R script was used to plot ranking of SEs and selected SE-associated genes.

ATAC-seq analysis.

Paired-end raw reads were trimmed 30 bp from the 3′ end to remove possible adaptor sequences and mapped to the mouse reference genome (GRCh38/mm10) using bowtie2 (v.2.3.4.3) with the parameter '–very-sensitive' enabled. Read duplication and reads that mapped to chrM were removed from downstream analysis. ATAC peaks were then called using the ENCODE ATAC-seq pipeline (https://github.com/ENCODE-DCC/atac-seq-pipeline) with an IDR threshold of 0.05. Called peaks were merged from all samples, and read counts were calculated and a raw count matrix produced. Differential peaks were identified using the R package DEseq (v.3.8). Peaks with more than twofold change were designated as DE peaks. To analyse the functional significance of peaks, GREAT was used with mm10 as the background genome and other parameters set as default.

Statistics and reproducibility.

All data are presented as the mean ± s.d. (error bars), unless otherwise specified in the legends. Representative reprogramming experiments are shown, unless otherwise specified. Reprogramming experiments and the associated assays (quantification, flow cytometry, beating cell, calcium flux, qPCR and co-immunoprecipitation) were performed three times

unless otherwise specified in the legends. For reprogramming experiments and the associated assays (quantification, flow cytometry, beating cell, calcium flux and qPCR), quantification and statistics were derived from $n = 3$ independent experiments unless specified in the legends. Sample sizes for next-generation sequencing were determined following ENCODE consortium guidelines; RNA-seq was performed using $n = 3$ biological replicates per group; ChIP-seq was performed using $n = 2$ biological replicates per group; ATAC-seq was performed using $n = 3$ biological replicates per group, with the exclusion of one replicate outlier from both control and treatment samples by principal component analysis. The criteria were pre-established. Statistical analyses were performed using GraphPad Prism 9. P values were calculated using unpaired two-tailed Student's t-test or one-way analysis of variance (ANOVA) with modification for multiple comparison, unless otherwise indicated in the legends. Significance is depicted by $*P < 0.05$, $**P < 0.01$, $***P < 0.001$ and $****P < 0.0001$.

Reporting Summary.

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

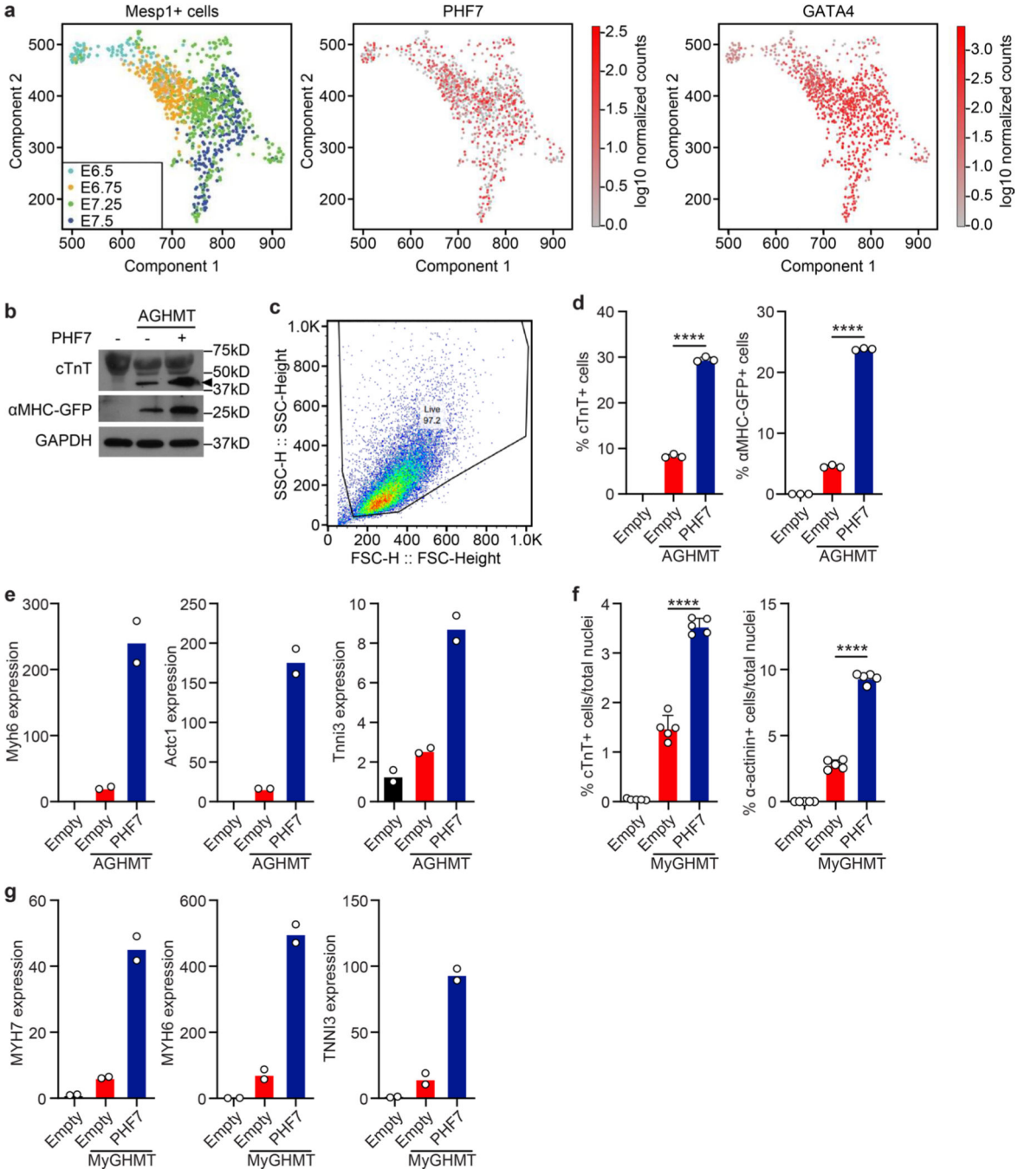
Data availability

All RNA-seq, ChIP-seq and ATAC-seq data that support the findings of this study have been deposited in the Gene Expression Omnibus (GEO) under accession code GSE151328. Previously published ChIP-seq and single-cell RNA-seq data that were re-analysed here are available under accession codes GSE90893, GSE112315, GSE100471 and http://singlecell.stemcells.cam.ac.uk/mesp1. Source data are provided with this paper. All other data supporting the findings of this study are available from the corresponding author upon reasonable request.

Code availability

The custom R script for plotting the ROSE results have been deposited to GitHub with the access link as https://github.com/chenkn009/PlotROSE.

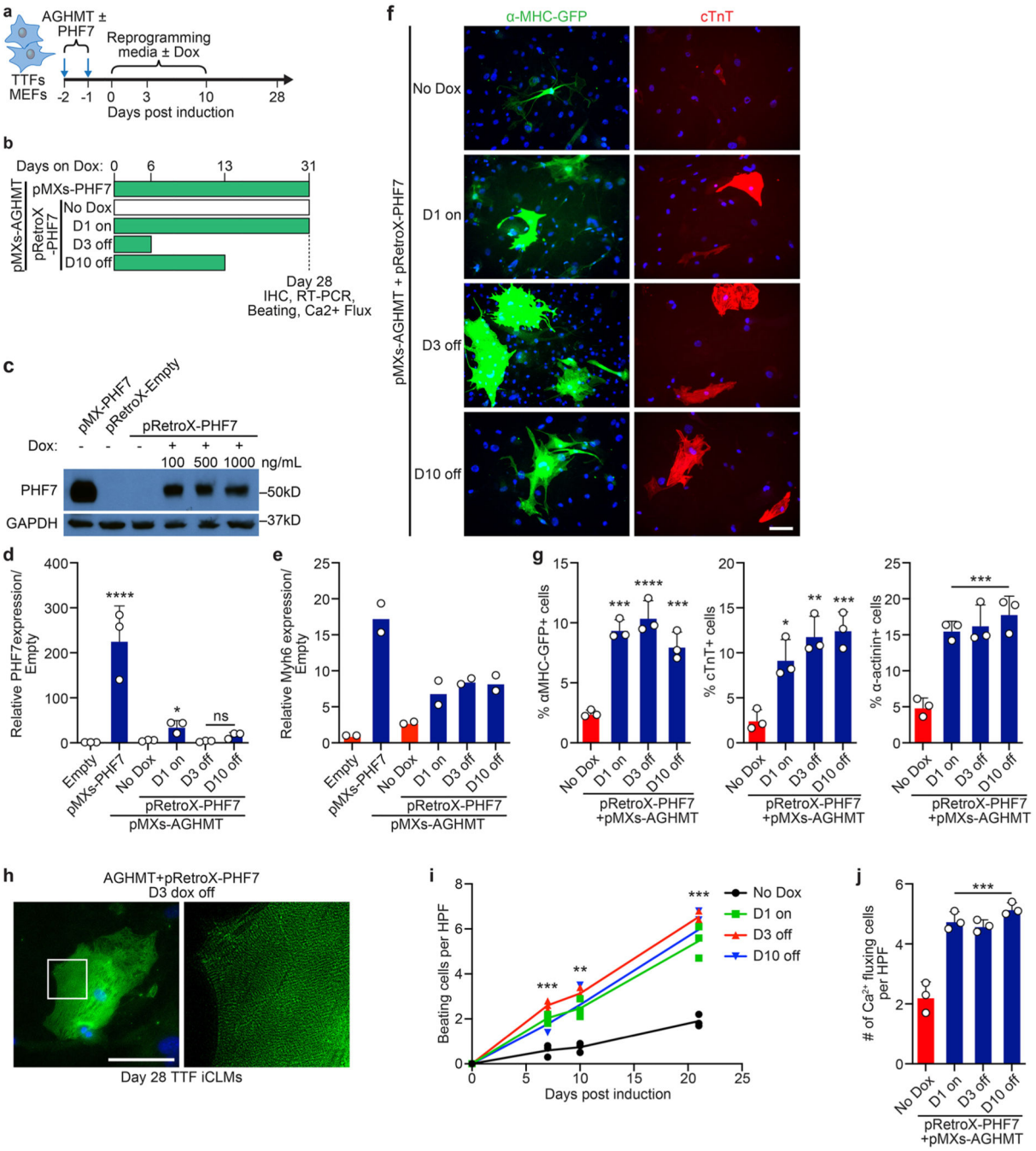
Extended Data



Extended Data Fig. 1 |. PHF7 augments cardiac reprogramming.

(**a**) tSNE plots from single-cell RNA Seq analysis of Mesp1+ cardiac progenitors from E6.5-E7.5 demonstrate robust expression of *Phf7* and *Gata4* throughout the existence of these cells. **b**, Western blot demonstrating increased cTnT and αMHC-GFP protein expression in day7 AGHMT ±PHF7 TTF iCLMs. Arrow indicates relevant cTnT band. GAPDH is a loading control. Biologically independent experiments were performed with similar results

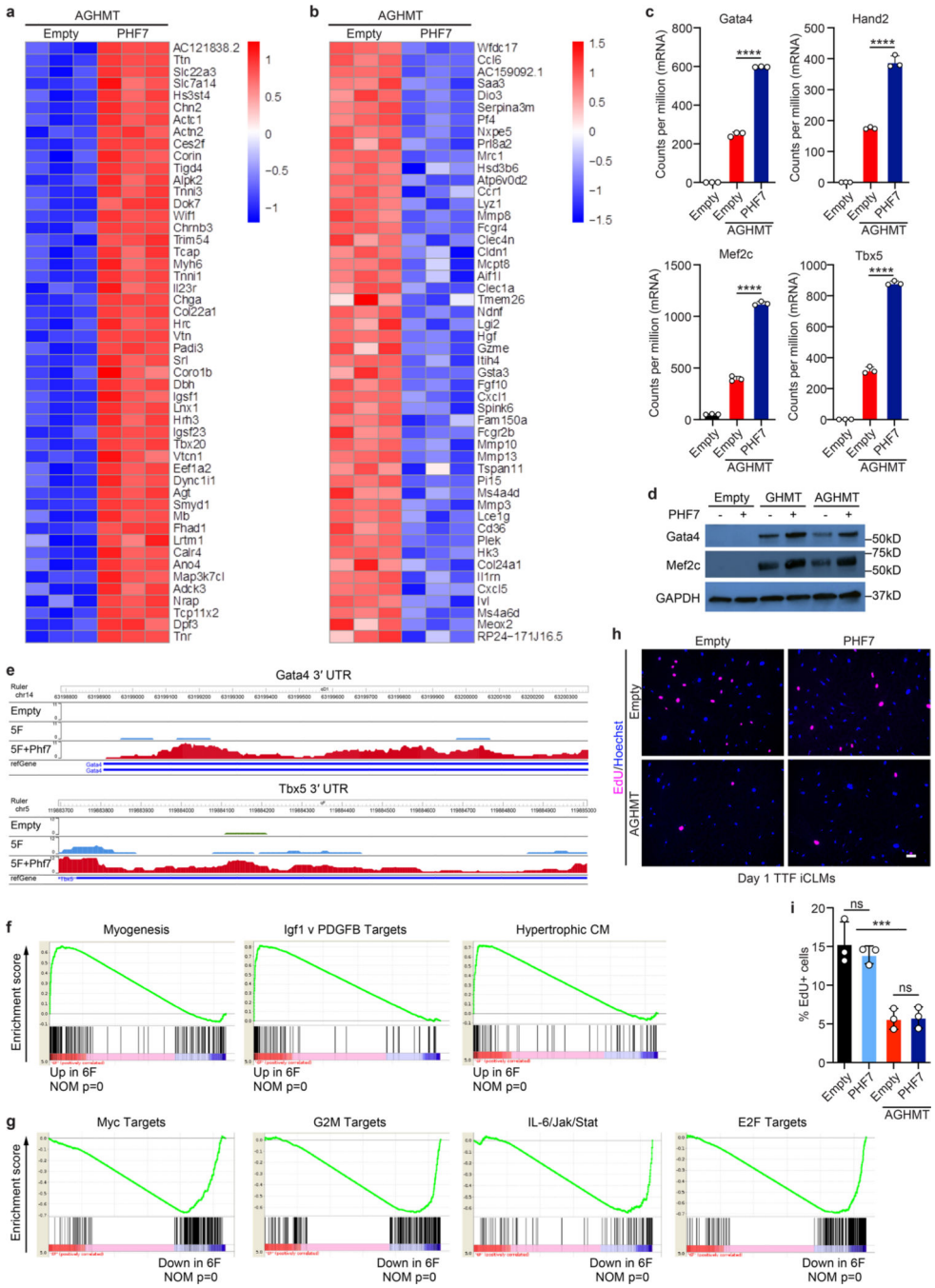
two times. **c**, Representative FACs plot demonstrating side and forward scatter gating of live cells. Biologically independent experiments were performed three times with similar results across all FACs experiments in this manuscript. **d**, Quantitative analysis by flow cytometry demonstrates % αMHC-GFP+ and % cTNT+ cells in day 7 AGHMT ±PHF7 TTF iCLMs. n=3 biologically independent samples. p****<0.0001. Data are presented as mean ±SD values. (**e**) qPCR of TTF iCLMs 7 days after infection with AGHMT ±PHF7 for cardiac markers. Relative to Empty virus control. Normalized to GAPDH. n=2 biologically independent samples. **f**, Quantitative analysis of immunocytochemistry in Day 21 My-GHMT±PHF7 human CFs. n=4 biologically independent samples. p****<0.0001. Data are presented as mean ±SD values. (**g**) qPCR of Day 21 My-GHMT±PHF7 human CF iCLMs for cardiac markers. Relative to Empty virus control. Normalized to GAPDH. n=2 biologically independent samples. All statistical comparisons between groups were evaluated by one-way ANOVA analysis, with modification for multiple comparisons. Source data are provided as a Source data file.



Extended Data Fig. 2 |. Transient PHF7 expression augments cardiac reprogramming.

a, Schematic for doxycycline-inducible expression strategy of PHF7 in reprogramming. **b**, Schematic representation of doxycycline (dox) dosing strategy. Dox was administered for different intervals; 4 weeks on (D1 on), 6 days (D3 off), and 13 days (D10 off), immediately following infection of TTFs or MEFs with pRetroX-PHF7/pMXs-AGHMT/pMXs-PHF7. Cardiac reprogramming media was added at D0 and changed along with dox every 2 days. **c**, Western blot demonstrating activation of PHF7 expression in the presence of dox (100–1000ng/mL). GAPDH is a loading control. Biologically independent experiments were

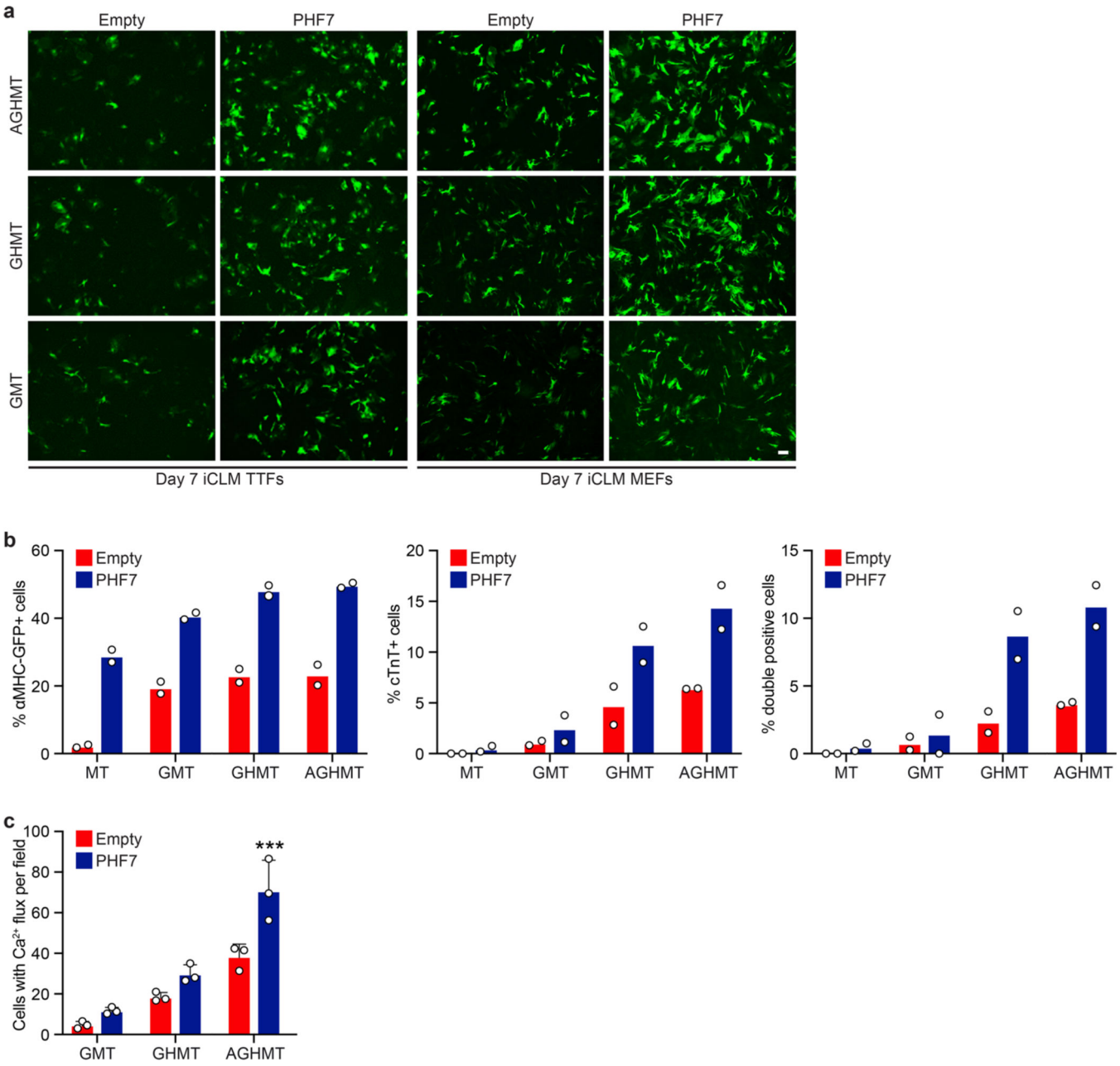
performed with similar results two times. (**d**) qPCR for PHF7 expression iCLM MEFs in the indicated samples after 28 days. n=3 biologically independent samples, p*=0.0468, p****<0.0001. Data are presented as mean ±SD values. (**e**) qPCR for Myh6 expression in iCLM TTFs in the indicated samples after 28 days. n=2 biologically independent samples. **f**, Representative immunocytochemistry for αMHC-GFP (green) and cTNT (red) in indicated treatment groups for Day 28 TTF iCLMs and (**g**) quantification of %αMHC-GFP+, % cTnT +, and %α-actinin+ cells in indicated treatment groups Hoechst (blue). Scale=100μm. n=3 biologically independent samples. %αMHC-GFP: D1 on p***=0.0001, D3 off p****<0.0001, D10 off p***=0.0007. % cTnT+: D1 on p*=0.0117, D3 off p**=0.0015, D10 off p***=0.001. %α-actinin+: D1 on p***=0.001, D3 off p***=0.0006, D10 off p***= 0.0003. Data are presented as mean ±SD values. **h**, Immunocytochemistry for α-actinin in D3 off TTF iCLMs at day 28. Zoomed view demonstrates sarcomere organization. Scale=100μm. Biologically independent experiments were performed with similar results three times. **i**, Quantification of number of spontaneously beating cells per high power field in MEF iCLMs. n=3 biologically independent experiments. Day 7: D1 on p*** 0.0007, D3 off p****<0.0001, Day 10 off p**=0.0028. Day 10: D1 on p**=0.0071, D3 off p***=0.0009, D10 off p**=0.0041. Day 21: D1 on p***=0.0006, D3 off p****<0.0001, D10 off p***=0.002. **j**, Quantification of intracellular calcium flux by Fluo-4 assay in day 28 iCLM MEFs. n=3 biologically independent samples. p****<0.0001. Data are presented as mean ±SD values. All statistical comparisons between groups were evaluated by one-way ANOVA analysis, with modification for multiple comparisons. Source data are provided as a Source data file.



Extended Data Fig. 3 |. PHF7 activates the cardiac transcriptome.

a, Unbiased heatmap displaying the 50 most upregulated transcripts by RNA-Seq in the presence of AGHMT+PHF7. **b**, Unbiased heatmap displaying the 50 most downregulated transcripts by RNA-Seq in the presence of AGHMT+PHF7. Color scale by Z score. **c**, Plots of cardiac reprogramming transcription factors upregulated in the presence of PHF7 by RNA-Seq analyses (normalized counts per million). n=3 biologically independent replicates. p****<0.0001. Data are presented as mean ±SD values. **d**, Western blot of lysates from day 7 TTF iCLMs following GHMT and AGHMT reprogramming ±PHF7 blotting for Gata4

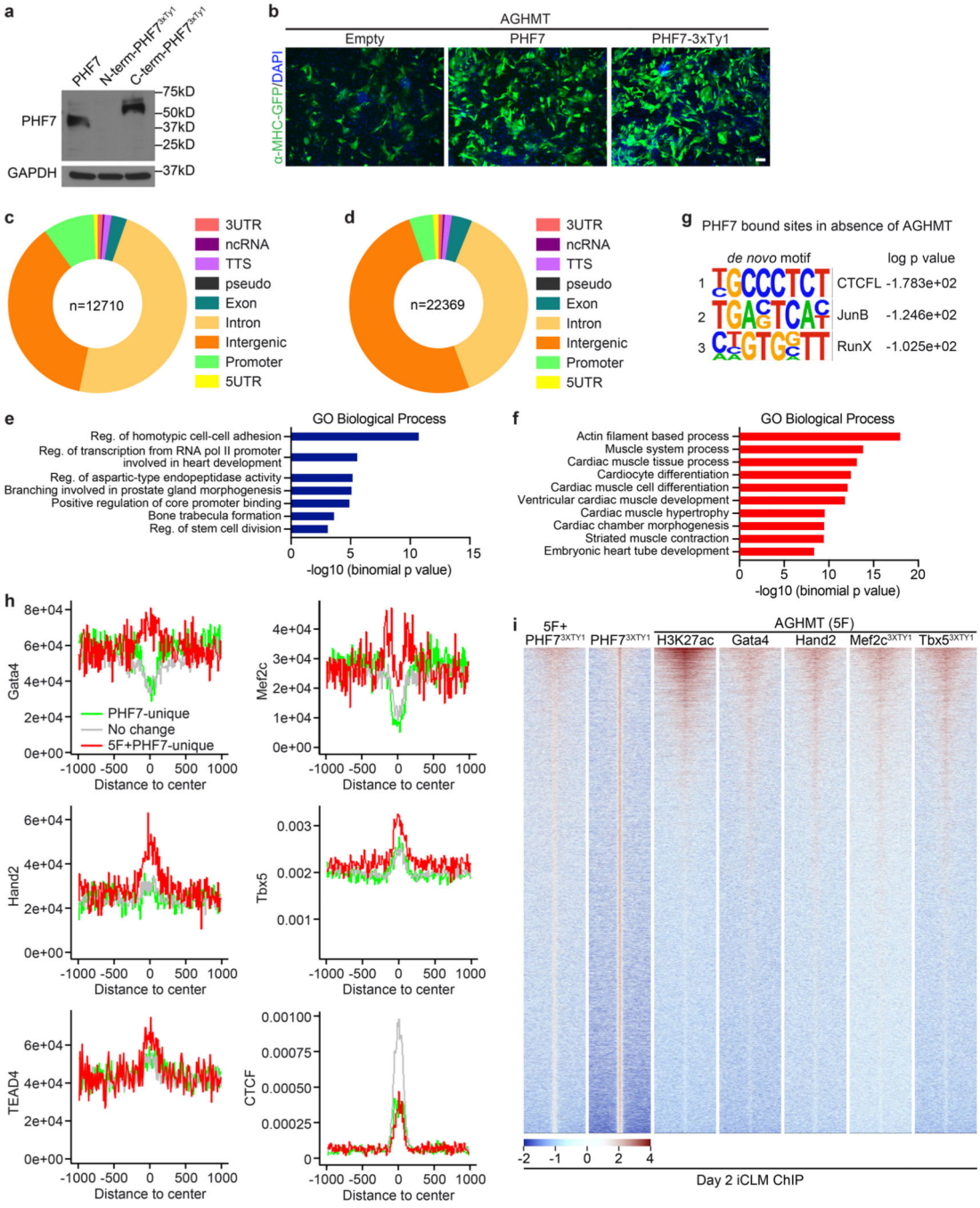
and Mef2c expression. GAPDH is a loading control. Biologically independent experiments were performed with similar results two times. **e**, RNA-seq tracks at Gata4 3' UTR and Tbx5 3'UTR from Empty, AGHMT (5F), and AGHMT+PHF7 (5F+PHF7) samples. **f, g**, Enrichment plots of indicated gene sets and their nominal p-value of genes upregulated by PHF7 (**f**) and downregulated by PHF7 **(g**). **h**, EdU pulse labeling of day 1 TTF iCLMs in the presence of Empty, PHF7, or AGHMT±PHF7. EdU (magenta), Hoechst (blue). Scale=100μm. Biologically independent experiments were performed with similar results three times. (**i**) Quantification of EdU pulse labeling of day 1 TTF iCLMs in the presence of Empty, PHF7, or AGHMT±PHF7. n=3 biologically independent samples. Empty v AGHMT: p***0.0008. AGHMT±PHF7: ns=0.9993. Data are presented as mean ±SD values. All statistical comparisons between groups were evaluated by one-way ANOVA analysis, with modification for multiple comparisons. Source data are provided as a Source data file.



Extended Data Fig. 4 |. PHF7 augments reprogramming with GMT, GHMT, and AGHMT.

a, Representative immunocytochemistry images demonstrate that PHF7 augments GMT, GHMT, or AGHMT reprogramming in both αMHC-GFP transgenic TTF and MEF iCLMs at day 7. αMHC-GFP (green), scale=200μm. Biologically independent experiments were performed with similar results three times. **b**, Quantitative analysis of % of αMHC-GFP+, cTnT+, and double positive TTF iCLMs 7d after infection with MT, GMT, GHMT, or AGHMT ±PHF7. n=2 biologically independent samples. **c**, Quantification number of cells with active calcium flux per high-powered field by Fluo-4 assay in Day 14 iCLM MEFs infected with GMT, GHMT, or AGHMT ±PHF7. n=3 biologically independent samples, p***=0.0003. Data are presented as mean ±SD values. All statistical comparisons between

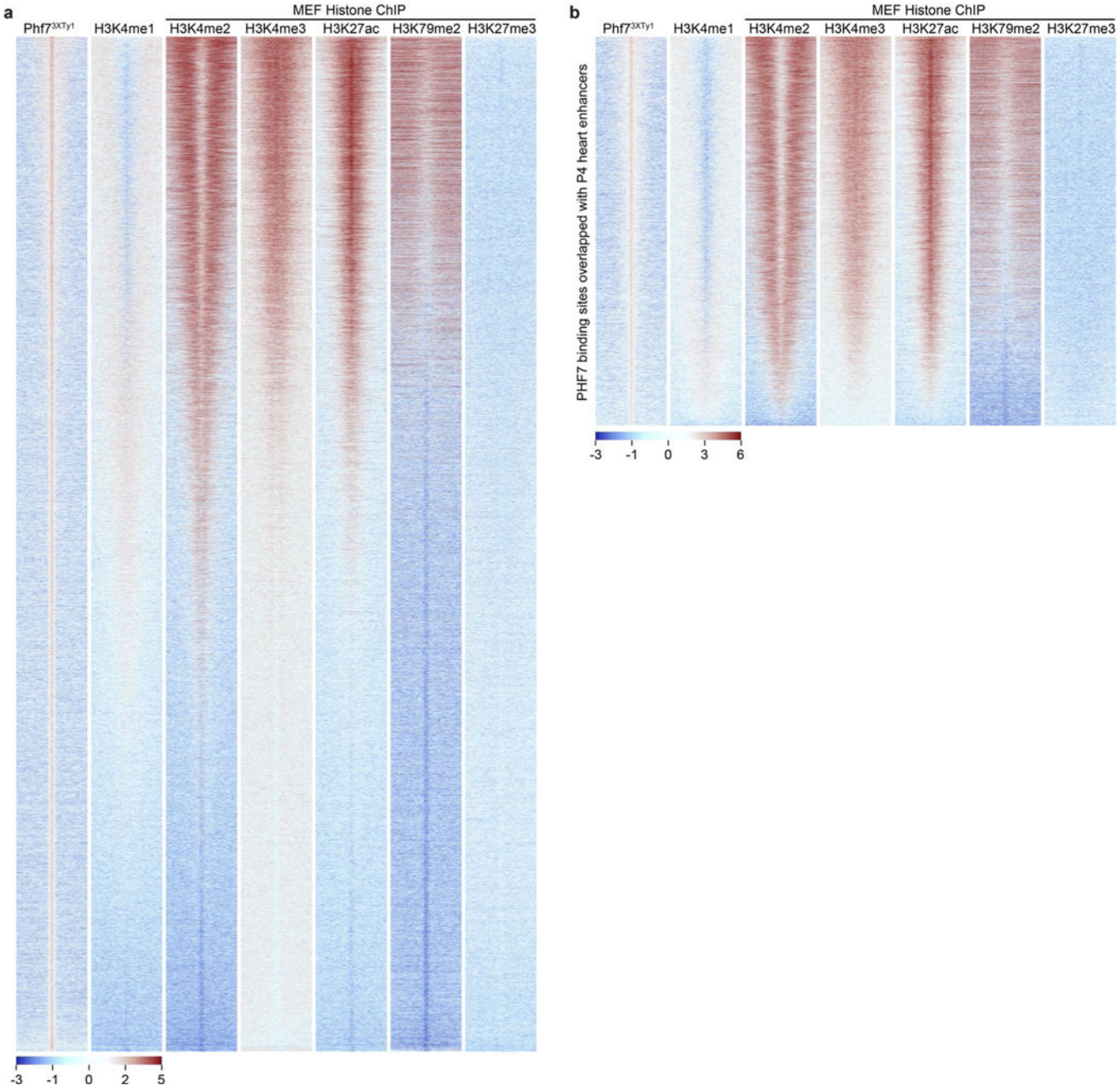
groups were evaluated by one-way ANOVA analysis, with modification for multiple comparisons. Source data are provided as a Source data file.



Extended Data Fig. 5 |. PHF7 binds to cardiac enhancers.

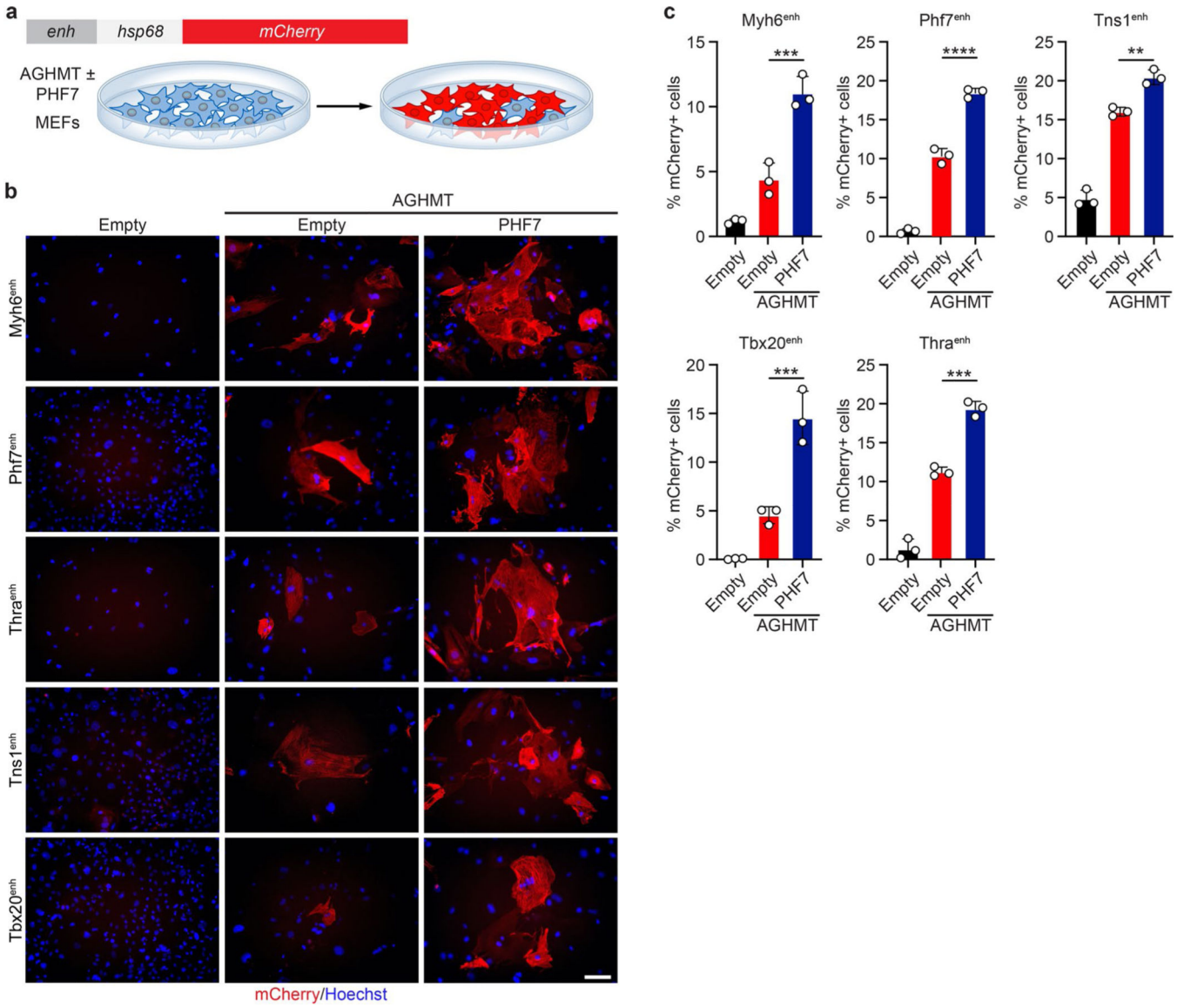
a, Western blot for PHF7 of lysates from HEK293 cells transfected with PHF7, N-terminal 3x-Ty1-PHF7, or C-terminal PHF7–3xTy1. GAPDH is a loading control Biologically independent experiments were performed with similar results two times. **b**, Representative immunocytochemistry images demonstrate that PHF^{3xTy1} augments reprogramming to a

similar extent as untagged PHF7 in day7 MEF iCLMs. Scale=200μm. Biologically independent experiments were performed with similar results three times. **c**, Genomic location of annotated PHF7^{3xTy1} peaks in the presence of reprogramming factors (AGHMT+ PHF7^{3xTy1} -unique). **d**, Genomic location of annotated PHF7^{3xTy1} peaks in the absence of reprogramming factors (PHF7^{3xTy1} -unique). **e**, GO enrichment analysis of PHF7^{3xTy1} peaks in the absence of AGHMT (PHF7^{3xTy1} -unique) as determined by GREAT analysis. **f**, GO enrichment analysis of PHF7^{3xTy1} peaks in the presence of AGHMT (AGHMT+ PHF7^{3xTy1} -unique) as determined by GREAT analysis. (**g**) *de novo* motif analysis of PHF7^{3xTy1} peaks in the absence of AGHMT by HOMER (PHF7^{3xTy1} -unique). (FDR threshold 10^{-3}). **h**, Motif frequency analysis of Gata4, Mef2c, Hand2, Tbx5, TEAD, and CTCF centered on PHF7 peak signal in reprogramming (AGHMT+PHF7^{3xTy1})-unique peaks (red), unchanged PHF7^{3xTy1} peaks (gray), or PHF7^{3xTy1}-unique peaks in the absence of reprogramming factors (green). **i**, Heatmap ordered by PHF7^{3xTy1} signal in absence of AGHMT (PHF7^{3xTy1} -unique peaks) aligned with H3K27ac, Gata4, Hand2, Mef2c, Tbx5 ChIP signal from day2 AGHMT iCLMs. Normalized ChIP-seq signal with ±2kb window centered around peak summit and sorted in descending order by signal intensity. Source data are provided as a Source data file.



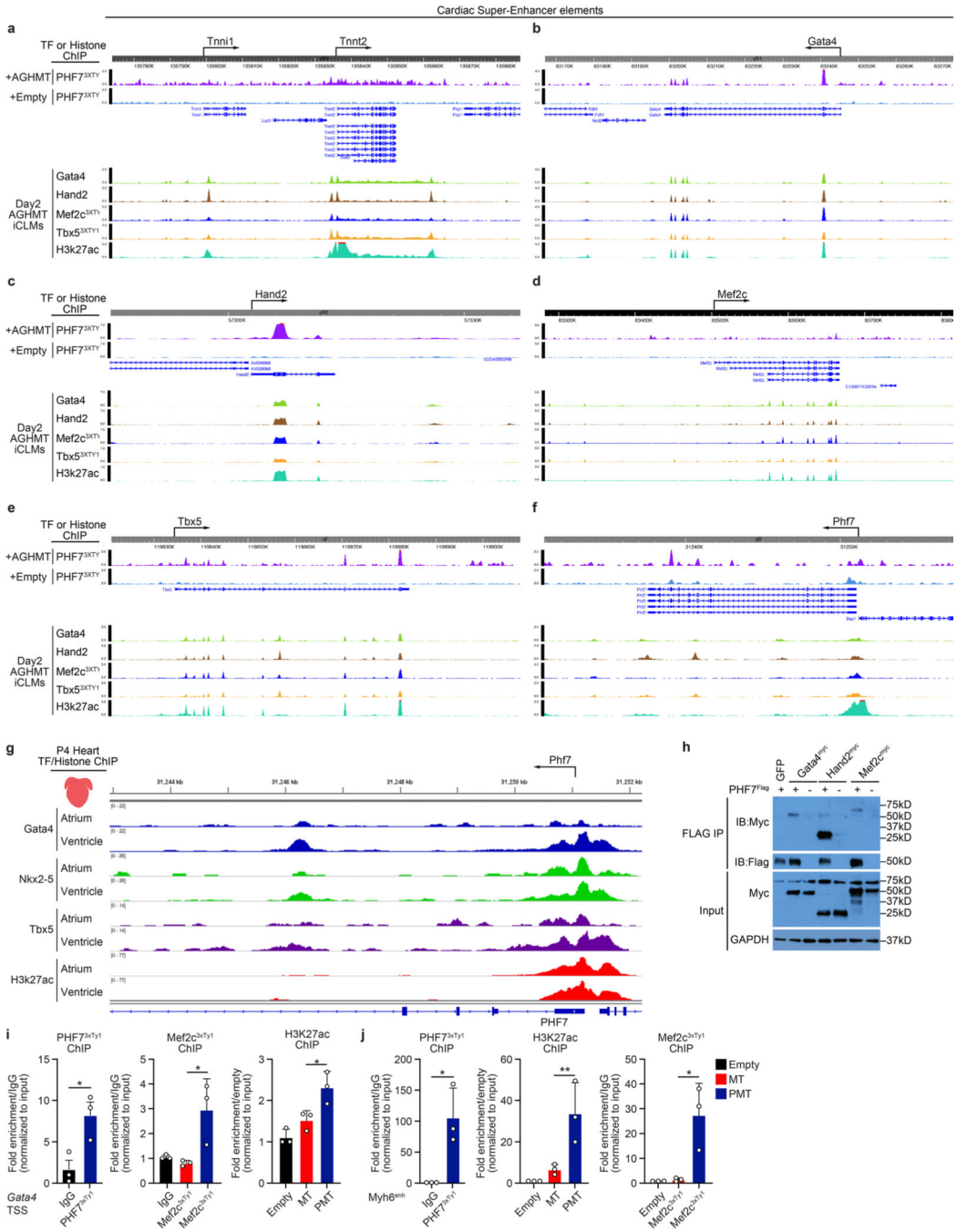
Extended Data Fig. 6 |. PHF7 aligns with H3K4me2 and other active histone marks.

a, Heatmap ordered by PHF7^{3xTy1} peaks aligned with H3K4me1, H3K4me2, H3K4me3, H3K27ac, H3K79me2, and H3K27me3 histone ChIP from uninduced MEFs. Normalized ChIP-seq signal with ±2kb window centered around peak summit and sorted in descending order by signal intensity. **b**, Heatmap ordered by PHF7^{3xTy1} peaks bound to P4 heart enhancers aligned with H3K4me1, H3K4me2, H3K4me3, H3K27ac, H3K79me2, and H3K27me3 histone ChIP from uninduced MEFs. Normalized ChIP-seq signal with ±2kb window centered around peak summit and sorted in descending order by signal intensity.



Extended Data Fig. 7 |. PHF7 activates cardiac enhancers.

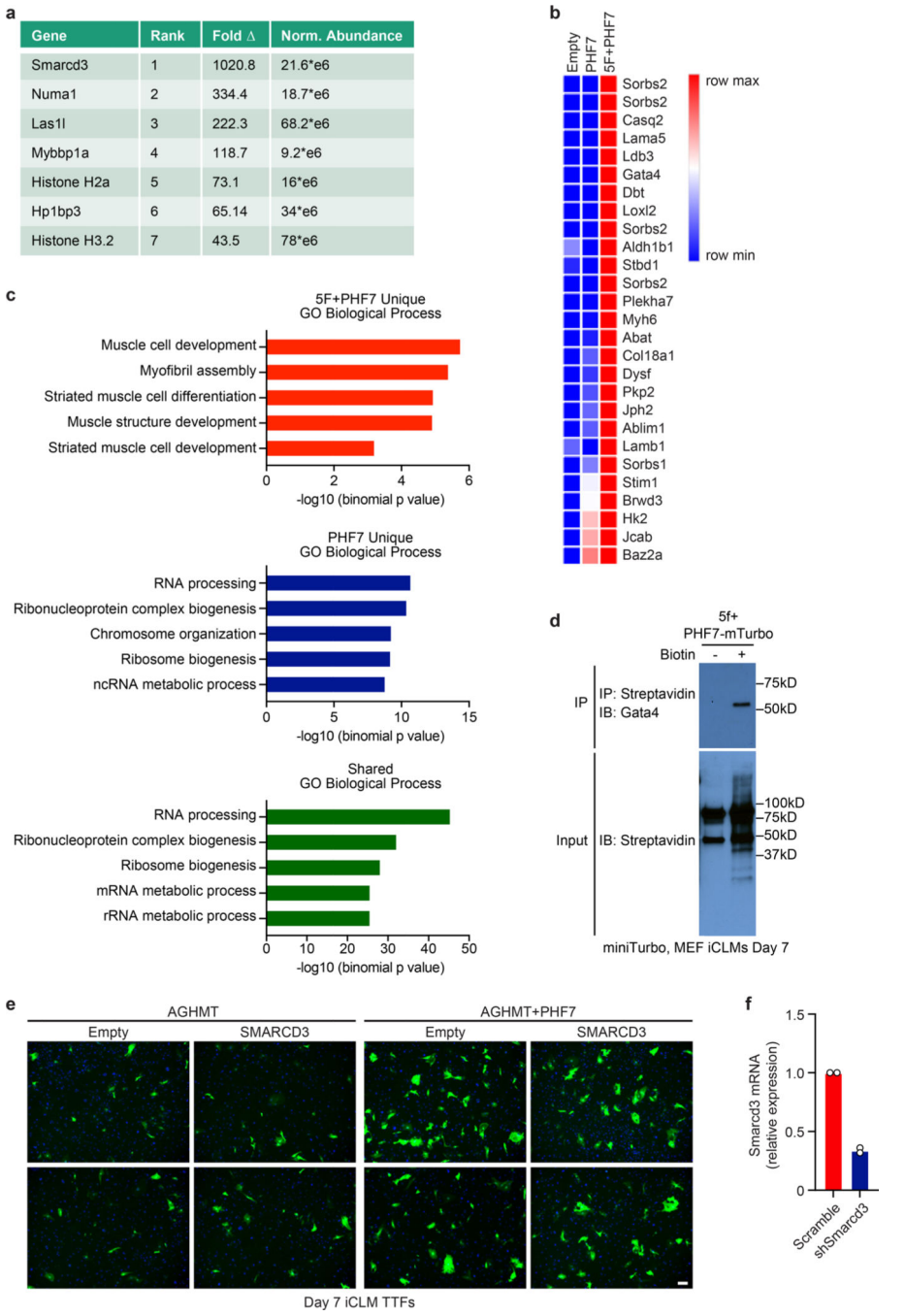
a, Schematic of enhancer-hsp68-mCherry retroviral generation and application to AGHMT ±PHF7 MEF iCLM reprogramming. **b**, Representative images of Myh6^{enh}, Phf7^{enh}, Thra^{enh}, Tns1^{enh}, and Tbx20^{enh}-hsp68-mCherry activation in Day 4 AGHMT±PHF7 MEF iCLMs. mCherry (red), Hoechst (blue). Scale=100μm. Biologically independent experiments were performed with similar results three times. **c**, Quantification of enhancer activation by %mCherry+ cells in Day 4 AGHMT±PHF7 MEF iCLMs treated with respective enhancer-hsp68-mCherry constructs. n=3 biologically independent samples. Myh6^{enh}: p***=0.0006, Phf7^{enh}: p****<0.0001, Thra^{enh}: p***=0.0001, Tns1^{enh}: p**=0.0023, Tbx20^{enh}: p***=0.0008. Data are presented as mean ±SD values. All statistical comparisons between groups were evaluated by one-way ANOVA analysis, with modification for multiple comparisons. Source data are provided as a Source data file.



Extended Data Fig. 8 |. PHF7 participates in a core TF autoregulatory circuit.

a-f, Genome browser shots of PHF7^{3xTy1} binding ±AGHMT at the a) *Tnni1/Tnnt2*, b) *Gata4*, (**c**) *Hand2*, (d) *Mef2c*, (**e**) *Tbx5*, and (**f**) *Phf7* super-enhancer loci, aligned with binding profiles of cardiac TFs by Gata4, Hand2, Mef2c, Tbx5, and H3K27ac ChIP in day 2 AGHMT iCLMs. **g**, Genome browser shot of P4 mouse atrium and ventricle TF (Gata4, Nkx2–5, and Tbx5) and H3K27ac ChIP alignment at the *Phf7* locus. **h**, FLAG co-immunoprecipitation in HEK293 cells transfected with PHF7^{FLAG} and GFP, Gata4^{myc}, Hand2^{myc}, or Mef2c^{myc}. (IP) Immunoprecipitation. (IB) Immunoblot. GAPDH is a loading

control. Biologically independent experiments were performed with similar results three times. **i**, PHF7^{3xTy1} ChIP qPCR in day2 iCLMs at the *Gata4* locus treated with PMT as compared to IgG control. n=3 biologically independent samples, p*=0.0261, by unpaired two-tailed Student's t-test. Data are presented as mean ±SEM values. Mef2c^{3xTy1} ChIP qPCR at the *Gata4* locus in day 2 iCLMs treated with Mef2c^{3xTy1}+Tbx5 (MT) or PHF7+Mef2c^{3xTy1}+Tbx5 (PMT) as compared to IgG control. n=3 biologically independent replicates. p*=0.0250, by one-way ANOVA with adjustment for multiple comparisons. Data are presented as mean ±SD values. H3K27ac ChIP qPCR at the *Gata4* locus in day2 iCLMs treated with Empty vector, MT, or PMT cocktail. n=3 biologically independent samples. p*=0.0320. Data are presented as mean ±SD values. **j**, PHF7^{3xTy1} ChIP qPCR in day2 iCLMs at the *Myh6* locus treated with PMT as compared to IgG control. n=3 biologically independent samples. p*=0.0184, as determined by unpaired two-tailed Student's t test. Mef2c^{3xTy1} ChIP qPCR at the *Myh6* locus in day 2 iCLMs treated with Mef2c^{3xTy1}+Tbx5 (MT) or PHF7+Mef2c^{3xTy1}+Tbx5 (PMT) as compared to IgG control. n=3 biologically independent samples. p*=0.0118, as determined by one-way ANOVA with adjustment for multiple comparisons. H3K27ac ChIP qPCR at the *Myh6* locus in day2 iCLMs treated with Empty vector, MT, or PMT cocktails. n=3 biologically independent samples. p*=0.0202, as determined by one-way ANOVA with adjustment for multiple comparisons. Data are presented as mean ±SD values. Source data are provided as a Source data file.



Extended Data Fig. 9 |. PHF7 interacts with cardiac proteins in reprogramming.

a, Table with top shared proteomics hits ranked based on fold change relative to control and normalized abundance from miniTurbo proximity biotinylation assay performed in PHF7^{miniTurbo} and AGHMT+PHF7^{miniTurbo} infected MEFs. **b**, Heatmap demonstrating AGHMT+PHF7^{miniTurbo} -unique proteins (related to Supplementary Data Table 2). Biologically independent experiments for AGHMT+PHF7^{miniTurbo} samples were performed with similar results two times. **c**, GO Pathway analysis of unique and shared proteins between PHF7^{miniTurbo} and AGHMT+PHF7^{miniTurbo} biotin-treated samples. **d**, Streptavidin

IP of AGHMT(5F)+PHF7^{miniTurbo} infected MEFs in biotin-treated and untreated control followed by western blot. (IP) Immunoprecipitation. (IB) Immunoblot. Biologically independent experiments were performed with similar results two times. **e**, Immunocytochemistry images demonstrate overexpression of pMXs-SMARCD3 with AGHMT or AGHMT+PHF7 does not augment reprogramming in day 7 TTF iCLMs. αMHC-GFP (green), Hoechst (blue). Biologically independent experiments were performed with similar results three times. **f**, Realtime PCR validation of *Smarcd3* transcript knockdown by shSmarcd3 in murine neuro2a (N2a). n=2 biologically independent samples. Source data are provided as a Source data file.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank J. Cabrera for graphical assistance; J. Xu, X. Liu and the Sequencing Core Facility at the Children's Research Institute and the Genomics and Next Generation Sequencing Core Facility at UT Southwestern for performing the Illumina sequencing. We are grateful to A. Mobley and the Flow Cytometry Core Facility for their assistance. This work was supported by grants from the NIH (HL-130253, HL-138426 and HD-087351), the Foundation Leducq Transatlantic Networks of Excellence in Cardiovascular Research and the Robert A. Welch Foundation (grant 1-0025 to E.N.O.). G.A.G. was supported by a NIH T32 Training grant (5T32HL125247-04). H.Z. was supported by a predoctoral fellowship (14PRE20030030) from the American Heart Association. H.H. was supported by a Uehara Memorial Foundation Postdoctoral Fellowship and a Kanae Foreign Study Grant. M.G.M. was a Pew Latin American Fellow in the Biomedical Sciences, supported by the Pew Charitable Trusts.

E.N.O. is a co-founder and member of the Scientific Advisory Board of Tenaya Therapeutics and holds equity in the company.

References

1. Ieda M. et al. Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. Cell 142, 375–386 (2010). [PubMed: 20691899]
2. Qian L. et al. In vivo reprogramming of murine cardiac fibroblasts into induced cardiomyocytes. Nature 485, 593–598 (2012). [PubMed: 22522929]
3. Song K. et al. Heart repair by reprogramming non-myocytes with cardiac transcription factors. Nature 485, 599–604 (2012). [PubMed: 22660318]
4. Zhou H, Dickson ME, Kim MS, Bassel-Duby R. & Olson EN Akt1/protein kinase B enhances transcriptional reprogramming of fibroblasts to functional cardiomyocytes. Proc. Natl Acad. Sci. USA 112, 11864–11869 (2015). [PubMed: 26354121]
5. Zhang Z, Zhang W. & Nam YJ Stoichiometric optimization of Gata4, Hand2, Mef2c, and Tbx5 expression for contractile cardiomyocyte reprogramming. Sci. Rep 9, 14970 (2019). [PubMed: 31628386]
6. Kurotsu S, Suzuki T. & Ieda M. Direct reprogramming, epigenetics, and cardiac regeneration. J. Card. Fail 23, 552–557 (2017). [PubMed: 28529134]
7. Vaseghi H, Liu J. & Qian L. Molecular barriers to direct cardiac reprogramming. Protein Cell 8, 724–734 (2017). [PubMed: 28389873]
8. Zhou H. et al. ZNF281 enhances cardiac reprogramming by modulating cardiac and inflammatory gene expression. Genes Dev. 31, 1770–1783 (2017). [PubMed: 28982760]
9. Wamstad JA et al. Dynamic and coordinated epigenetic regulation of developmental transitions in the cardiac lineage. Cell 151, 206–220 (2012). [PubMed: 22981692]
10. Lickert H. et al. Baf60c is essential for function of BAF chromatin remodelling complexes in heart development. Nature 432, 107–112 (2004). [PubMed: 15525990]

11. Sun X. et al. Cardiac-enriched BAF chromatin-remodeling complex subunit Baf60c regulates gene expression programs essential for heart development and function. Biol. Open 10.1242/bio.029512 (2018).
12. Hota SK et al. Dynamic BAF chromatin remodeling complex subunit inclusion promotes temporally distinct gene expression programs in cardiogenesis. Development 10.1242/dev.174086 (2019).
13. Hang CT et al. Chromatin regulation by Brg1 underlies heart muscle development and disease. Nature 466, 62–67 (2010). [PubMed: 20596014]
14. Takeuchi JK & Bruneau BG Directed transdifferentiation of mouse mesoderm to heart tissue by defined factors. Nature 459, 708–711 (2009). [PubMed: 19396158]
15. Christoforou N. et al. Transcription factors MYOCD, SRF, Mesp1 and SMARCD3 enhance the cardio-inducing effect of GATA4, TBX5, and MEF2C during direct cellular reprogramming. PLoS ONE 8, e63577 (2013).
16. Hashimoto H. et al. Cardiac reprogramming factors synergistically activate genome-wide cardiogenic stage-specific enhancers. Cell Stem Cell 25, 69–86.e5 (2019). [PubMed: 31080136]
17. Liu Z. et al. Re-patterning of H3K27me3, H3K4me3 and DNA methylation during fibroblast conversion into induced cardiomyocytes. Stem Cell Res. 16, 507–518 (2016). [PubMed: 26957038]
18. Stone NR et al. Context-specific transcription factor functions regulate epigenomic and transcriptional dynamics during cardiac reprogramming. Cell Stem Cell 25, 87–102.e9 (2019). [PubMed: 31271750]
19. Zhou Y. et al. A loss of function screen of epigenetic modifiers and splicing factors during early stage of cardiac reprogramming. Stem Cells Int. 2018, 3814747 (2018).
20. Zhou Y. et al. Bmi1 is a key epigenetic barrier to direct cardiac reprogramming. Cell Stem Cell 18, 382–395 (2016). [PubMed: 26942853]
21. Yang SY, Baxter EM & Van Doren M. *Phf7* controls male sex determination in the *Drosophila* germline. Dev. Cell 22, 1041–1051 (2012). [PubMed: 22595675]
22. Yang SY et al. Control of a novel spermatocyte-promoting factor by the male germline sex determination factor PHF7 of *Drosophila melanogaster*. Genetics 206, 1939–1949 (2017). [PubMed: 28588035]
23. Wang X. et al. PHF7 is a novel histone H2A E3 ligase prior to histone-to-protamine exchange during spermiogenesis. Development 10.1242/dev.175547 (2019).
24. Lescroart F. et al. Defining the earliest step of cardiovascular lineage segregation by single-cell RNA-seq. Science 359, 1177–1181 (2018). [PubMed: 29371425]
25. Whyte WA et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell 153, 307–319 (2013). [PubMed: 23582322]
26. Wang Y, Li X. & Hu H. H3K4me2 reliably defines transcription factor binding regions in different cells. Genomics 103, 222–228 (2014). [PubMed: 24530516]
27. Chronis C. et al. Cooperative binding of transcription factors orchestrates reprogramming. Cell 168, 442–459 e420 (2017).
28. He A, Kong SW, Ma Q. & Pu WT Co-occupancy by multiple cardiac transcription factors identifies transcriptional enhancers active in heart. Proc. Natl Acad. Sci. USA 108, 5632–5637 (2011). [PubMed: 21415370]
29. Saint-Andre V. et al. Models of human core transcriptional regulatory circuitries. Genome Res. 26, 385–396 (2016). [PubMed: 26843070]
30. Devine WP, Wythe JD, George M, Koshiba-Takeuchi K. & Bruneau BG Early patterning and specification of cardiac progenitors in gastrulating mesoderm. eLife 10.7554/eLife.03848 (2014).
31. Takeuchi JK et al. Baf60c is a nuclear Notch signaling component required for the establishment of left-right asymmetry. Proc. Natl Acad. Sci. USA 104, 846–851 (2007). [PubMed: 17210915]
32. Stanton BZ et al. Smarca4 ATPase mutations disrupt direct eviction of PRC1 from chromatin. Nat. Genet 49, 282–288 (2017). [PubMed: 27941795]
33. Han P, Hang CT, Yang J. & Chang CP Chromatin remodeling in cardiovascular development and physiology. Circ. Res 108, 378–396 (2011). [PubMed: 21293009]

References

34. Nam YJ et al. Induction of diverse cardiac cell types by reprogramming fibroblasts with cardiac transcription factors. Development 141, 4267–4278 (2014). [PubMed: 25344074]
35. Takahashi K. & Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126, 663–676 (2006). [PubMed: 16904174]
36. Corces MR et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat. Methods 14, 959–962 (2017). [PubMed: 28846090]
37. Zlatic SA, Ryder PV, Salazar G. & Faundez V. Isolation of labile multi-protein complexes by in vivo controlled cellular cross-linking and immuno-magnetic affinity chromatography. J. Vis. Exp 10.3791/1855 (2010).
38. Branon TC et al. Efficient proximity labeling in living cells and organisms with TurboID. Nat. Biotechnol 36, 880–887 (2018). [PubMed: 30125270]

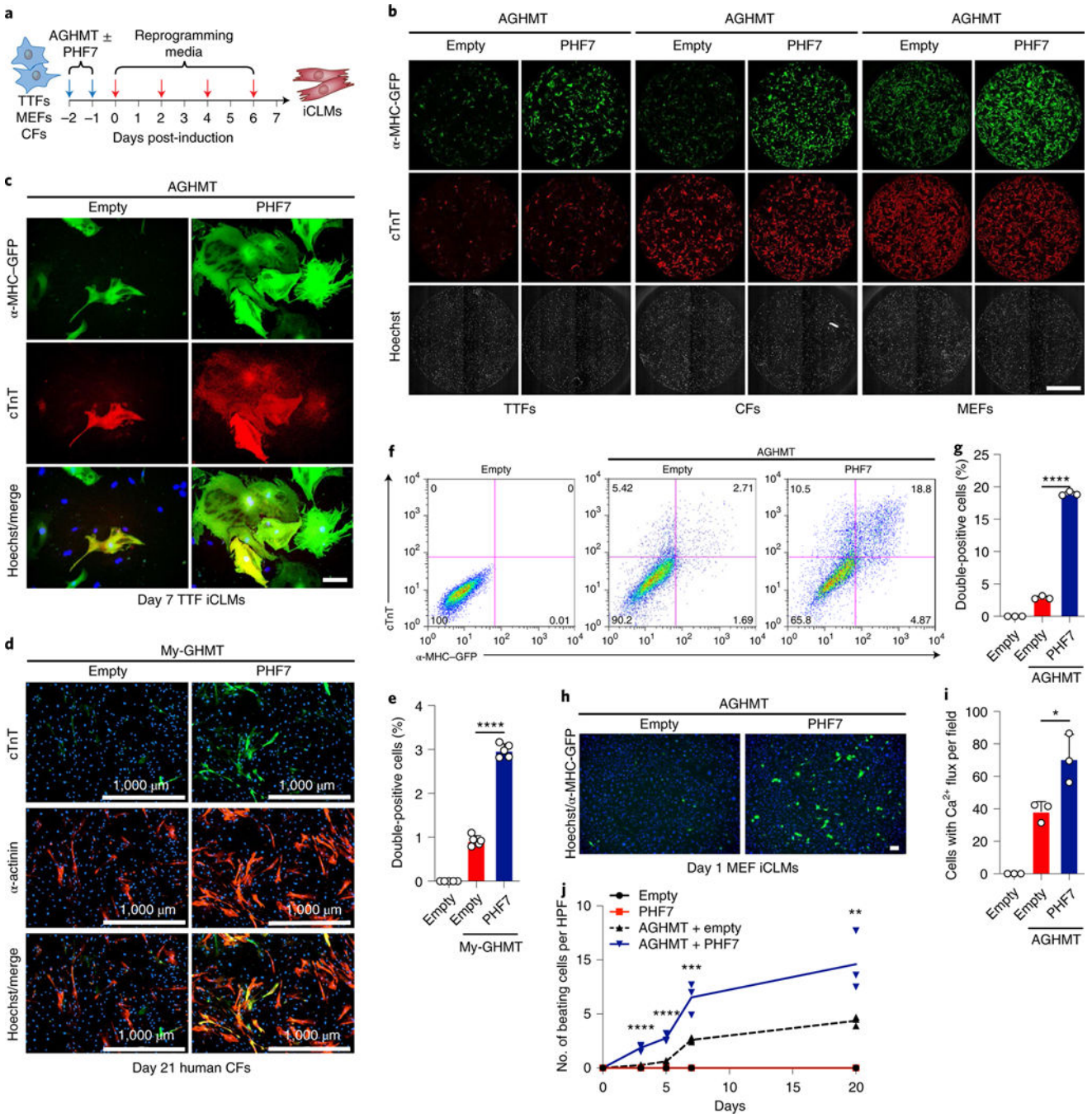


Fig. 1 |. PHF7 promotes direct cardiac reprogramming.

a, Schematic of direct cardiac reprogramming assay. **b**, Immunostaining of day 7 α-MHC–GFP transgenic TTF, CF or MEF AGHMT ± PHF7 iCLMs. α-MHC–GFP (green), cTnT (red), Hoechst (bright-field). Scale bar, 2 mm. **c**, Immunostaining of day 7 α-MHC–GFP transgenic TTF AGHMT ± PHF7 iCLMs. α-MHC–GFP (green), cTnT (red), Hoechst (blue). Scale bar, 100 μm. Three biologically independent experiments were performed, with similar results obtained, for **b** and **c**. **d**,**e**, Immunostaining (**d**) of day 21 human CF (NHCF-V) My-GHMT ± PHF7 iCLMs and quantification (**e**) of the percentage of double-positive cells.

cTnT (green), α-actinin (red), 4,6-diamidino-2-phenylindole (DAPI; blue). Scale bar, 1,000 μm. $n = 4$ biologically independent samples, ****$P < 0.0001$. Data are presented as the mean ± s.d. **f**,**g**, Representative flow cytometry plots (**f**) of day 7 TTF iCLMs and quantitative analysis (**g**) of the percentage of double-positive (α-MHC–GFP+cTnT+) cells. $n = 3$ biologically independent samples, ****$P < 0.0001$. Data are presented as the mean ± s.d. α-MHC–GFP (*x* axis, FITC), cTnT (*y* axis, Alexa Fluor 647). **h**, Immunocytochemistry of day 1 AGHMT ± PHF7 MEF iCLMs. α-MHC–GFP (green), Hoechst (blue). Scale bar, 200 μm. Three biologically independent experiments were performed, with similar results obtained. **i**, Number of Ca^2+-fluxing cells per high-power field (HPF) in day 14 AGHMT ± PHF7 MEF iCLMs. $n = 3$ biologically independent samples, *$P = 0.0135$. Data are presented as the mean ± s.d. **j**, Quantification of the number of spontaneously beating cells per HPF in MEF iCLMs infected with empty virus, PHF7 or AGHMT ± PHF7. $n = 3$ biologically independent samples, ****$P < 0.0001$, ***$P < 0.001$, **$P < 0.01$. All statistical comparisons between groups were evaluated by one-way ANOVA analysis, with modification for multiple comparisons. Source data are provided as a source data file.

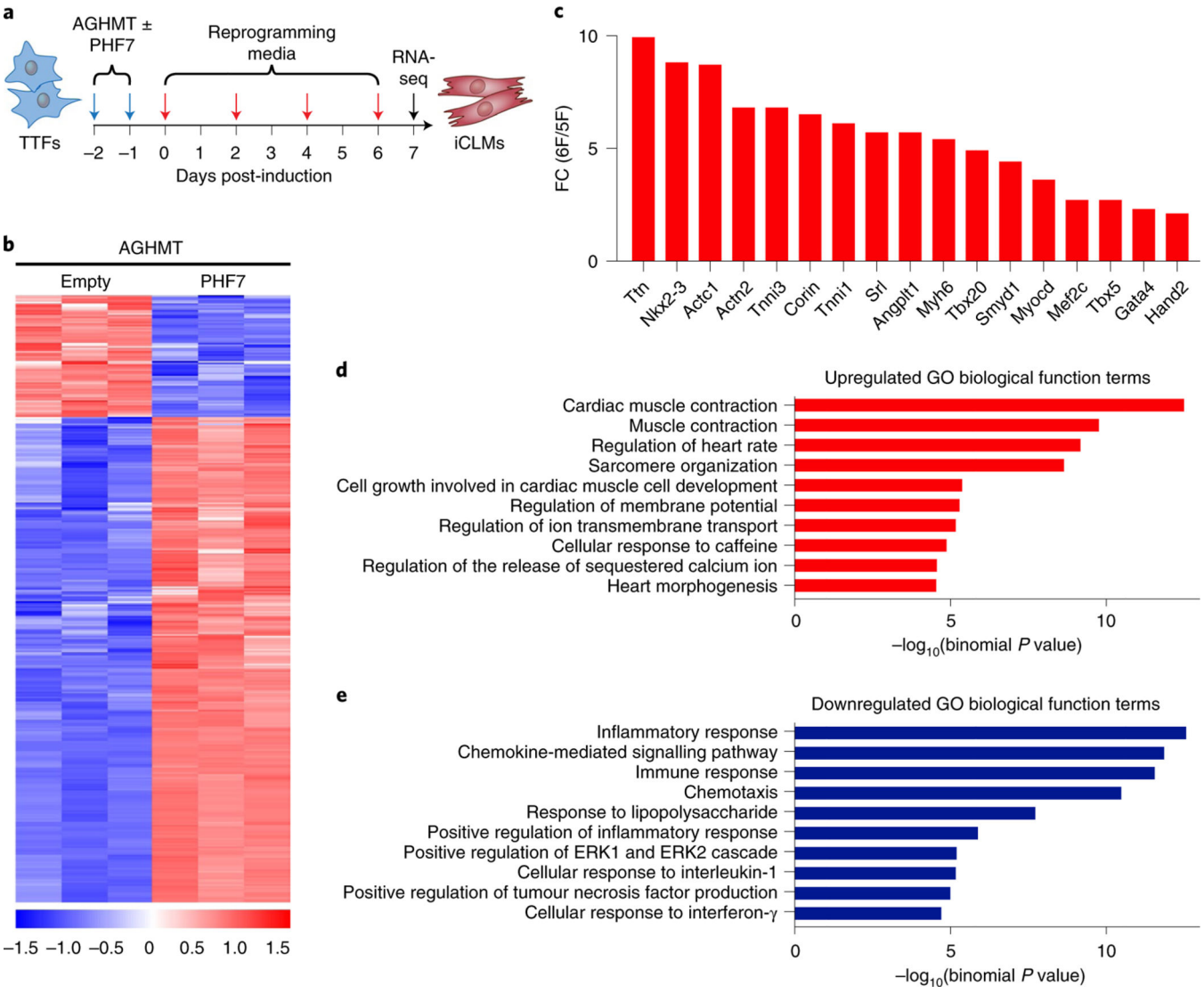


Fig. 2 |. PHF7 globally activates the cardiac transcriptome.

a, Schematic of direct cardiac reprogramming assay followed by RNA-seq. **b**, Heatmap comparing differentially expressed transcripts in day 7 AGHMT ± PHF7 TTF iCLMs identified by RNA-seq. Upregulation (red), downregulation (blue). Colour scale represents the *Z*-score. $n = 3$ biologically independent samples per group. (FDR < 0.1, FC > 2). **c**, FC between AGHMT + PHF7 (6F) and AGHMT (5F) for selected cardiac markers and TFs. **d**,**e**, GO analysis showing biological processes associated with genes upregulated (**d**) and downregulated (**e**) as assessed using RNA-seq. Source data are provided as a source data file.

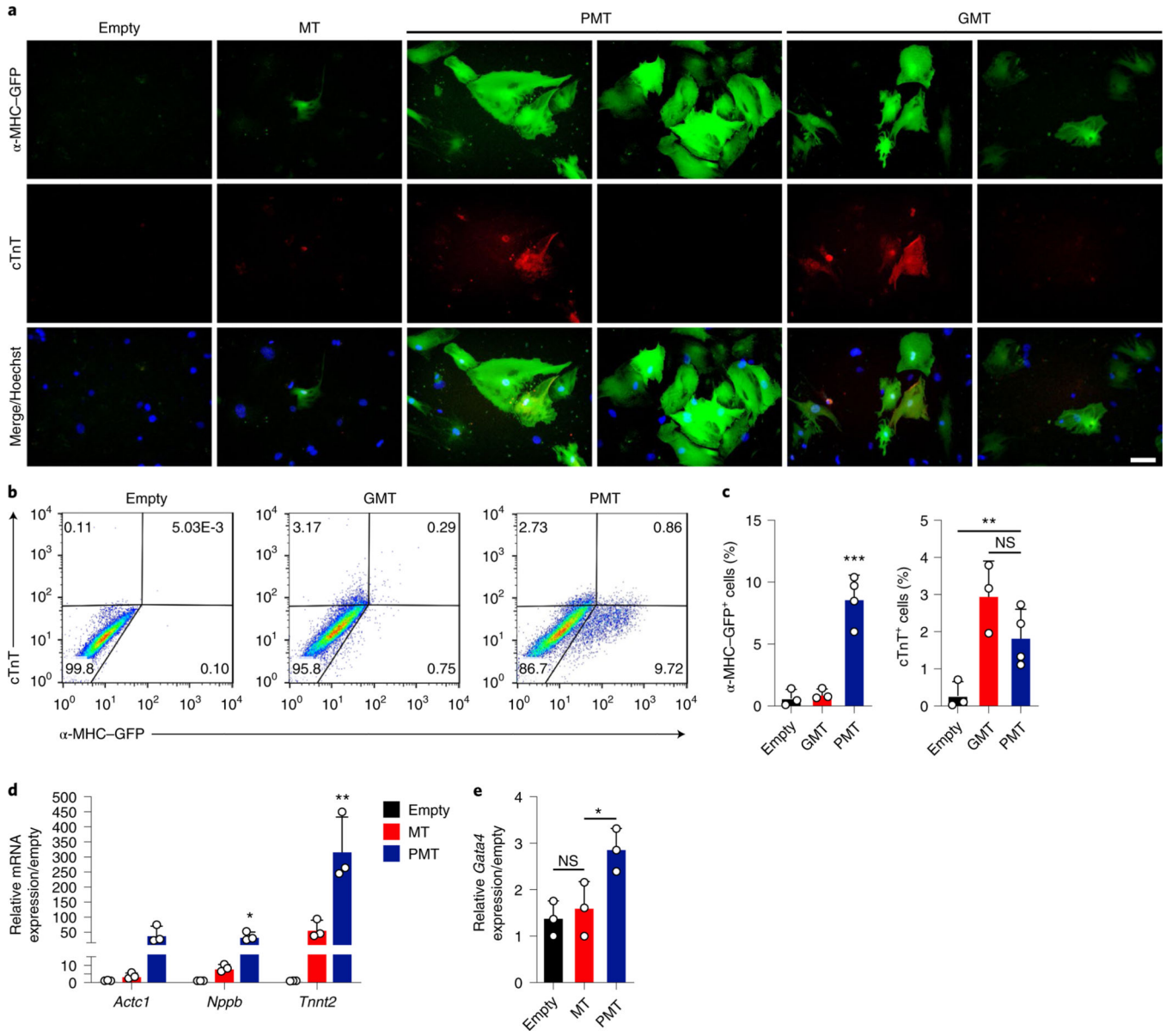


Fig. 3 |. PHF7 reprogrammes cells to a cardiac fate in the absence of exogenous Gata4.

a, Immunocytochemistry of day 7 TTF iCLMs treated with empty virus, MT, PMT or GMT. α-MHC–GFP (green), cTnT (red), Hoechst (blue). Scale bar, 100 μm. Three biologically independent experiments were performed, with similar results obtained. **b**, Representative flow cytometry plots of day 7 TTF iCLMs treated with empty virus, GMT or PMT. α-MHC–GFP (*x* axis, FITC) and cTnT (*y* axis, Alex Fluor 647). **c**, Quantitative analysis of the percentage of αMHC–GFP+ cells and cTNT+ cells in day 7 TTF iCLMs by flow cytometry. $n = 3$ biologically independent samples for empty and GMT, $n = 4$ biologically independent samples for PMT; ***$P = 0.0001$, **$P = 0.0083$, NS, not significant. Data are presented as the mean ± s.d. **d**, PCR with reverse transcription (RT–PCR) analysis of day 7 TTF iCLMs infected with empty vector, MT or PMT (relative to empty vector, normalized to *Gapdh*). $n =$ 3 biologically independent samples, *$P = 0.0189$, **$P = 0.0078$. Data are presented as the

mean ± s.d. **e**, RT–PCR analysis probing the 3′ UTR of *Gata4* in day 7 TTF iCLMs infected with empty vector, MT or PMT (relative to empty vector, normalized to *Gapdh*). $n = 3$ biologically independent replicates, $*P = 0.0187$. Data are presented as the mean ± s.d. All statistical comparisons between groups were evaluated by one-way ANOVA analysis, with modification for multiple comparisons. Source data are provided as a source data file.

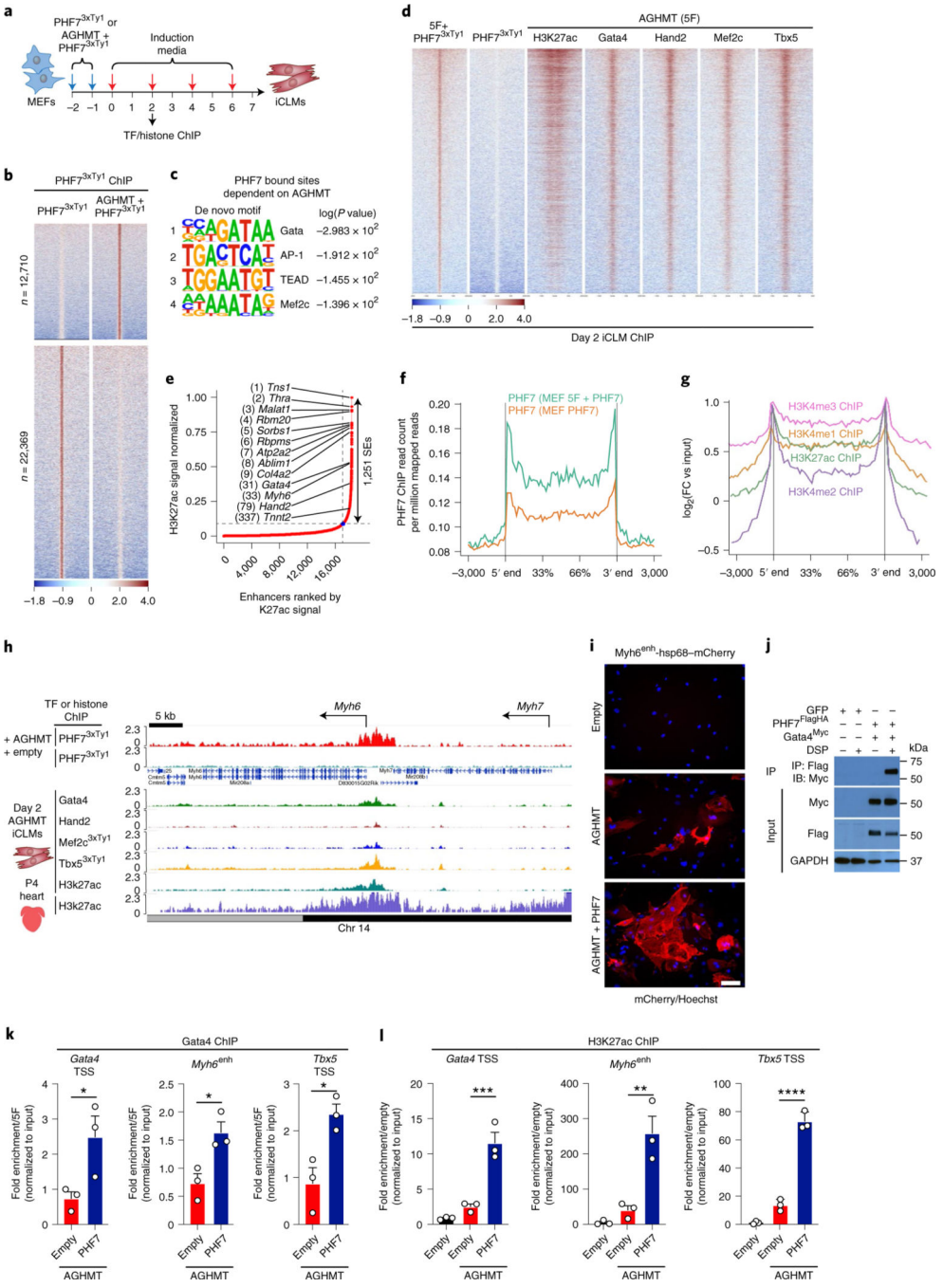


Fig. 4 |. PHF7 binds to and activates cardiac enhancers.

a, Schematic of direct cardiac reprogramming assay followed by ChIP-seq. **b**, Heatmap of differential PHF7^{3xTy1} ChIP-seq peaks in the presence or absence of AGHMT (FC > 2) (± 2-kb window centred on the peak summit). $n = 2$ biologically independent replicates per group. Scale indicates $\log_2$(fold change v. input). **c**, De novo PHF7^{3xTy1} motif analysis of reprogramming peaks by HOMER (AGHMT + PHF7^{3xTy1}-unique) (FDR threshold of 10^{-3}). **d**, Heatmap ordered by the 5F + PHF7^{3xTy1} peak signal aligned with H3K27ac, Gata4, Hand2, Mef2c and Tbx5 ChIP in day 2 AGHMT iCLMs (± 2-kb window). Scale indicates

log_2(fold change v. input). **e**, SE regions ranked and annotated by ROSE ($n = 1,251$) plotted in increasing order based on their input-normalized H3K27ac ChIP-seq signal from P4 mouse ventricle. **f**, Metagene plot of PHF7^{3xTy1} ChIP signals mapped to cardiac SEs with ± 3-kb flanking region in the presence (green) or absence (orange) of AGHMT. **g**, Metagene plot of H3K4me1, H3K4me2, H3K4me3 and H3K27ac ChIP signals derived from uninduced MEFs mapped to cardiac SEs with ± 3-kb flanking region. **h**, Genome browser shots of PHF7^{3xTy1} ± AGHMT ChIP at the *Myh6/Myh7* SE aligned with Gata4, Hand2, Mef2c^{3xTy1}, Tbx5^{3xTy1} and H3K27ac ChIP in day 2 AGHMT-induced iCLMs and P4 heart H3K27ac ChIP. **i**, Immunostaining for Myh6^{enh}–hsp68–mCherry in day 4 AGHMT ± PHF7 MEF iCLMs. mCherry (red), Hoechst (blue). Scale bar, 100 μm. Three biologically independent experiments were performed, with similar results obtained; see Extended Data Fig. 7c for quantification. **j**, Flag co-immunoprecipitation assay in HEK293 cells transfected with GFP, PHF7$^{Flag-HA}$ and/or Gata4^{Myc} ± DSP crosslinker. GAPDH was used as the loading control. IB, immunoblot; IP, immunoprecipitation. Three biologically independent experiments were performed, with similar results obtained. **k**, Gata4 ChIP qPCR in day 2 iCLMs treated with AGHMT ± PHF7. $n = 3$ biologically independent samples. *Gata4* TSS, $*P = 0.0482$. *Myh6*^{enh}, $*P = 0.023$. *Tbx5* TSS, $*P = 0.0197$. Statistical analyses performed using unpaired two-tailed Student's *t*-test. Data are presented as the mean ± s.d. **l**, H3K27ac ChIP qPCR in day 2 iCLMs treated with empty vector or AGHMT ± PHF7. $n = 3$ biologically independent samples. *Gata4* TSS, $***P = 0.001$. *Myh6*^{enh}, $**P = 0.0043$. *Tbx5* TSS, $****P < 0.0001$. Data are presented as the mean ± s.d. Statistical analyses performed using one-way ANOVA analysis, with modification for multiple comparisons. Source data are provided as a source data file.

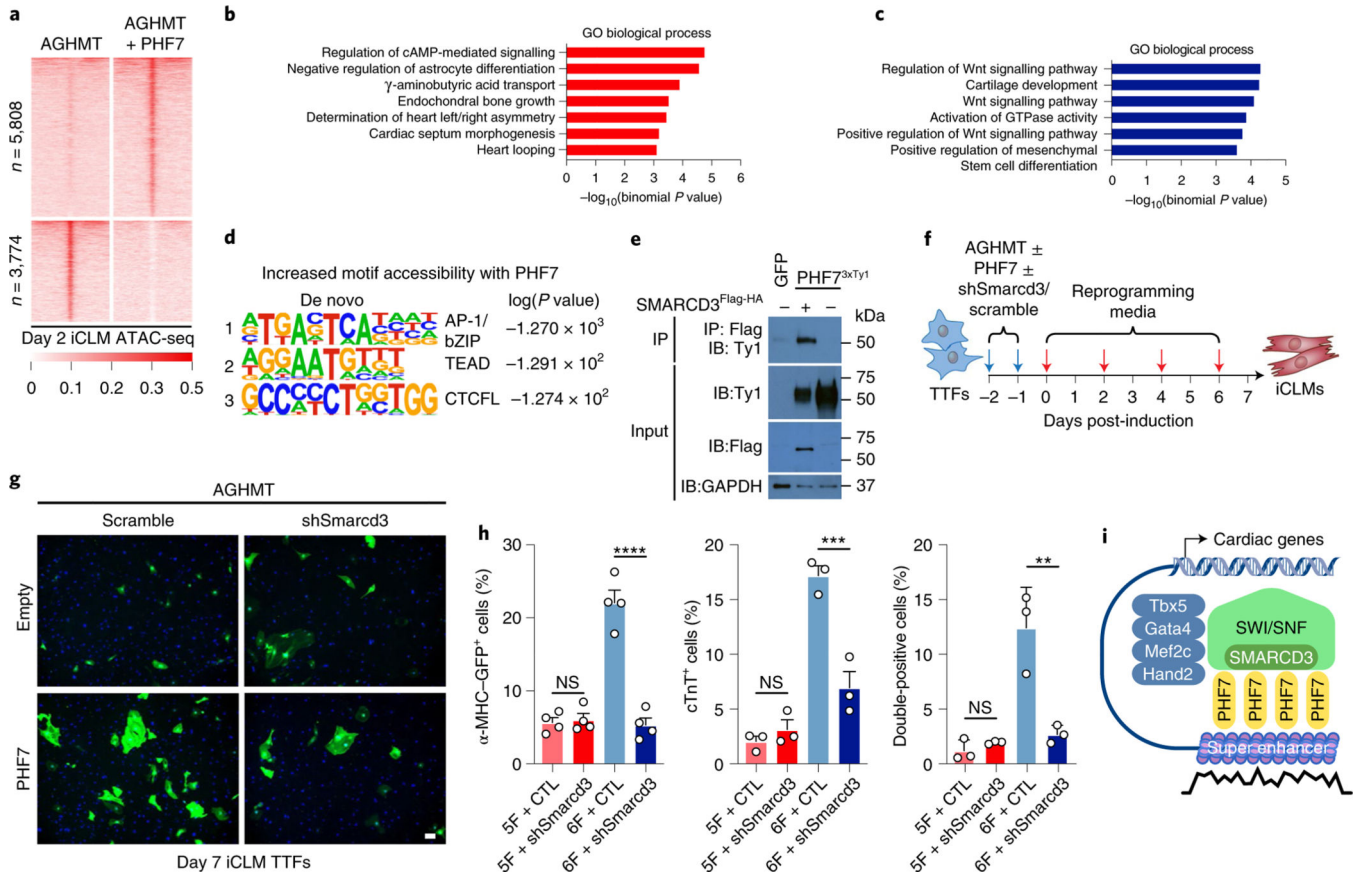


Fig. 5 |. PHF7 interacts with SMARCD3 to promote reprogramming.

a, Heatmap demonstrating differential peaks by ATAC-seq in day 2 AGHMT ± PHF7 iCLMs (FC > 2). **b**, GO pathway analysis of regions with increased accessibility in the presence of PHF7 as determined by GREAT analysis. **c**, GO pathway analysis of regions with decreased accessibility in the presence of PHF7 as determined by GREAT analysis. **d**, De novo motif analysis of regions with increased accessibility in the presence of PHF7 (FDR threshold of 10^{-3}). **e**, Flag co-immunoprecipitation in HEK293 cells transfected with GFP, $PHF7^{3xTy1}$ and/or $SMARCD3^{Flag\text{-}HA}$. GAPDH was used as the loading control. Biologically independent experiments were performed with similar results three times. **f**, Schematic of direct cardiac reprogramming assay and knockdown strategy. **g**, Immunocytochemistry of day 7 α-MHC–GFP TTF iCLMs after infection with AGHMT ± PHF7 and scramble or shSmarcd3. α-MHC–GFP (green), Hoechst (blue). Scale bar, 200 μm. **h**, Quantification of the percentage of α-MHC–GFP+, cTnT+ and double-positive cells. CTL is scramble control. ($n = 4$ biologically independent samples for the percentage of α-MHC–GFP+ samples and $n = 3$ biologically independent samples for the percentage of cTnT+ cells and double-positive samples). ***$P < 0.0001$, ***$P < 0.0004$, **$P < 0.0013$. Data are presented as the mean ± s.d. **i**, Graphical model of PHF7 recognition of histone modifications at cardiac SEs and recruitment of the SMARCD3–SWI/SNF complex and cardiac TFs to regulate cardiac gene transcription. All statistical comparisons between groups were

evaluated by one-way ANOVA analysis, with modification for multiple comparisons. Source data are provided as a source data file.