Title

Human-gained heart enhancers are associated with species-specific cardiac attributes

Journal

Nature Cardiovascular Research, 1(9)

ISSN

2731-0590

Authors

Destici, Eugin
Zhu, Fugui
Tran, Shaina
et al.

Publication Date

2022-09-01

DOI

10.1038/s44161-022-00124-7

Peer reviewed

HHS Public Access

Author manuscript

Nat Cardiovasc Res. Author manuscript; available in PMC 2023 February 17.

Published in final edited form as:

Nat Cardiovasc Res. 2022 September ; 1(9): 830–843. doi:10.1038/s44161-022-00124-7.

Human-gained heart enhancers are associated with species-specific cardiac attributes

Eugin Destici[1,#], **Fugui Zhu**[1,#], **Shaina Tran**[1], **Sebastian Preissl**[2,3], **Elie N. Farah**[1], **Yanxiao Zhang**[2], **Xiameng Hou**[3], **Olivier B. Poirion**[3], **Ah Young Lee**[2], **Jonathan D. Grinstein**[1], **Joshua Bloomekatz**[4], **Hong Sook Kim**[1,^], **Robert Hu**[1], **Sylvia M. Evans**[1,5,6], **Bing Ren**[2,3,7,8,*], **Chris Benner**[1,*], **Neil C. Chi**[1,7,*]

[1]Department of Medicine, University of California, San Diego, La Jolla, CA, 92093, USA

[2]Ludwig Institute for Cancer Research, La Jolla, CA, 92093, USA

[3]Center for Epigenomics, University of California, San Diego, La Jolla, CA, 92093, USA

[4]Department of Biology, University of Mississippi

[5]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, CA, 92093, USA

[6]Department of Pharmacology, University of California, San Diego, La Jolla, CA, 92093, USA

[7]Institute of Genomic Medicine, School of Medicine, University of California, San Diego, La Jolla, CA, 92093, USA

[8]Moores Cancer Center, School of Medicine, University of California, San Diego, La Jolla, CA, 92093, USA

Abstract

The heart, a vital organ which is first to develop, has adapted its size, structure and function in order to accommodate the circulatory demands for a broad range of animals. Although heart development is controlled by a relatively conserved network of transcriptional/chromatin regulators, how the human heart has evolved species-specific features to maintain adequate cardiac output and function remains to be defined. Here, we show through comparative epigenomic analysis the identification of enhancers and promoters that have gained activity in humans during cardiogenesis. These cis-regulatory elements (CREs) are associated with genes involved in heart development and function, and may account for species-specific differences between human and mouse hearts. Supporting these findings, genetic variants that are associated with human cardiac phenotypic/disease traits, particularly those differing between human and mouse, are

[*]Corresponding authors: Correspondence to N.C.C. (nchi@ucsd.edu), C.B (cbenner@ucsd.edu) and B.R. (biren@ucsd.edu).
[^]Current address: Department of Biological Sciences, Sungkyunkwan University, Suwon, South Korea
[#]Shared first authors

Author contributions:

E.D., B.R. and N.C.C. conceived the project and the overall design of the experimental strategy. C.B. designed and performed bioinformatics analyses. E.D, F.Z., S.P., E.F., X.H, A.Y.L., and J.G. conducted experiments. Y.Z., O.B.P. and R.H, helped with bioinformatics analysis. B.R., S.M.E., J.B. and H.S.K. provided critical intellectual input and data interpretation. E.D. and N.C.C prepared the manuscript with input from all authors.

Ethics declarations

The authors declare no competing interests.

enriched in human-gained CREs. During early stages of human cardiogenesis, these CREs are also gained within genomic loci of transcriptional regulators, potentially expanding their role in human heart development. In particular, we discovered that gained enhancers in the locus of the early human developmental regulator *ZIC3* are selectively accessible within a subpopulation of mesoderm cells which exhibits cardiogenic potential, thus possibly extending the function of *ZIC3* beyond its conserved left-right asymmetry role. Genetic deletion of these enhancers identified a human gained enhancer that was required for not only *ZIC3* and early cardiac gene expression at the mesoderm stage but also cardiomyocyte differentiation. Overall, our results illuminate how human gained CREs may contribute to human-specific cardiac attributes, and provide insight into how transcriptional regulators may gain cardiac developmental roles through the evolutionary acquisition of enhancers.

Cis-regulatory elements (CREs) are required for establishing the precise developmental and cell type-specific gene expression programs, which regulate the specification, differentiation and function of distinct cell types[1,2]. In particular, these CREs are crucial for creating the gene regulatory networks (GRNs) that control the expression of genes not only directing the differentiation of mesodermal progenitors into the diverse cardiovascular cell types comprising the heart but also regulating their function[3]. Highlighting the importance of these cardiovascular GRNs, previous studies have shown that their disruptions can result in congenital heart disease (CHD) or fetal demise[4–7] as well as adult heart diseases[8–10]. Thus, there has been increasing interest in defining these GRNs and their cognate genes during mammalian heart development[11–16].

Despite relative conservation in organ development, structure and function across vertebrate species, notable evolutionary differences do exist. Compared to well-studied smaller model organisms such as the mouse, human organs have distinct morphological and functional differences, including their size, structure and physiology[17,18]. While such species-specific differences are in some cases driven by altering gene function, either through changes in function of conserved genes or through generation of new genes (e.g., gene duplication), many of these differences can be driven by altering spatial and temporal expression of conserved genes[19,20]. CREs, including active enhancers, are crucial regulatory DNA elements that can control these gene expression dynamics and are thus key sources of evolutionary novelty[19,20]. As an early example of this mechanism, cis-regulatory variation was found to underly the evolution of butterfly wing patterns[21] and has been well studied by interrogating other morphological differences in *Drosophila* and stickleback fish[20]. More recent evolutionary comparative epigenomic studies in mammalian limb[22] and brain development[23,24] have identified human specific enhancers referred as “human-gained enhancers” (HGEs) or “human gain-of-activity enhancers” (hGoA) that are associated with human developmental specificity in gene expression programs for these organs. Some of these HGEs have been further studied using reporter assays and found to generate appropriate cell type-specific reporter activity[22,23]. More recently, one such HGE was functionally validated through loss-of-function experiments and shown to regulate expression of *FGFR2* in human neural progenitor cells during neurogenesis[24]. However, whether this species-specific divergence in enhancer usage may arise from modification of the underlying DNA sequences (in cis) and/or differences in transcription factor utilization/

activity (in trans) remains to be examined for these and many other recently identified HGEs[22–24].

In addition to the aforementioned species-specific differences in the brain and limb, human hearts also display distinct structural and physiologic differences from other mammalian hearts[25], but the underlying mechanisms driving the (evolutionary) acquisition of these human-specific cardiac attributes remains to be identified. However, defining HGEs during heart development through comparative epigenomics and functionally examining them and their underlying cis- or trans-regulatory changes may address these critical gaps of knowledge. Thus, utilizing a human pluripotent stem cell (hPSC)-derived cardiac/cardiomyocyte differentiation system, we not only identified CREs and GRNs directing hPSCs into cardiomyocytes (CMs) but also examined how they may diverge from those regulating CM differentiation from mouse pluripotent stem cells (mPSCs). Comparative epigenomic analyses between these CREs revealed HGEs that are active across distinct cardiac developmental stages and associated with specific gene expression programs, which may explain species-specific differences between mouse and human hearts. Genetic variants associated with human cardiac phenotypic/disease traits, particularly those differing between human and mouse, are enriched in these HGEs. Furthermore, we discovered HGEs within genomic loci of transcriptional regulators during early stages of human cardiogenesis, potentially expanding the role of these regulators in human heart development. Supporting this notion, we identified HGEs in the genomic locus of the known early left-right asymmetry developmental transcription factor (TF), *ZIC3*, including one mesoderm-specific HGE that we show may facilitate a gained role for *ZIC3* in CM differentiation. Closer examination of this *ZIC3* HGE reveals cis-regulatory changes that may have enabled humans to evolutionarily gain this enhancer. Together, our results provide insight into how the human heart may have adapted its structure and function through acquiring enhancers that rewire the cardiac GRNs controlling the expression of cardiac genes critical for heart development and function.

Results

Dynamic cardiac gene regulatory programs control human cardiomyocyte differentiation.

To illuminate GRNs that control human heart development, we interrogated the chromatin landscape during hPSC cardiac differentiation, which includes the earliest stages of human cardiogenesis. Utilizing a *MYL2*:H2B-GFP hPSC ventricular cardiomyocyte (vCM) reporter line[26,27], we performed comprehensive epigenomic and gene expression profiling across a broad range of key cardiac developmental stages during CM (and more specifically vCM) differentiation (Fig. 1a and Extended Data Fig. 1a). These stages include the following: hPSCs, mesoderm (Mes), cardiac mesoderm (CMes), cardiac progenitor (CP), CM and vCM. In general, we discovered that identified CREs have H3K27ac signal that is largely stage- and cell type-specific as revealed by comparing H3K27ac signal between stages as well as to other, non-cardiac cell types, respectively (Extended Data Fig. 1b). Furthermore, these CREs primarily reside in intergenic and intronic regions (Extended Data Fig. 1c). Consistent with previous studies investigating enhancer dynamics in other differentiating hPSC-derived cell types[28,29], we observed a decrease in accessible chromatin regions from

hPSC to mesodermal stages (Extended Data Fig. 1d), but, in contrast to these findings[28,29], an overall increase in enhancer utilization from the cardiac progenitor to CM/vCM stages as indicated by an increased number of H3K27ac peaks and decreased number of H3K27me3 marks (Extended Data Fig. 1e, f). Integrating ATAC-seq data with corresponding ChIP-seq data obtained from the same samples revealed distinct epigenetic patterns that correlate with gene expression during heart development (Fig. 1b). These identified CREs were primarily divided into two major classes. The first class corresponds mainly to constitutive CREs comprising a) promoters as indicated by the presence of the promoter mark H3K4me3[30] and b) chromatin sites lacking H3K4me3 that are likely to be bound by CTCF[30] as supported by the strong CTCF motif enrichment in these sites (Figure 1b). The second class corresponds mainly to dynamic CREs consisting of primarily active enhancers (H3K4me1$^+$/H3K27ac$^+$). Consistent with previous results from mouse PSC-derived CM differentiation[15], we observed that many of these enhancers are primed (H3K4me1$^+$/H3K27ac$^-$) prior to activation as observed by the appearance of the H3K4me1 signal before acquisition of H3K27ac marks (Fig. 1b, clusters IV-VI). These enhancers could be further subdivided into four sub-classes that are active at different stages of CM differentiation (Fig. 1b, clusters III-VI). Examining corresponding RNA-seq from these stages revealed that the enhancers correlate closely with gene expression (Fig. 1b, RNA), thus supporting that these enhancers control dynamic gene expression during human CM differentiation. Consistent with these findings and previous bulk and single-cell studies[14,15,31–33], these genes associate with developmental Gene Ontology (GO) categories at the earlier mesodermal/cardiac progenitor stages (Fig. 1b, clusters III, IV) and cardiac GO categories at later CM differentiation stages (Fig. 1b, clusters V, VI). To determine the *in vivo* relevance of our *in vitro* cardiac enhancer maps, we interrogated H3K27ac ChIP-seq data from purified human fetal vCMs[12], and confirmed that many of the *in vitro* enhancers, in particular those from the vCM stage, are also active *in vivo* (Extended Data Fig. 1g). Additional analyses of accessible chromatin regions overlapping active enhancers uncovered not only TF binding site motifs associated with cardiac development that have been identified previously[12,15,34–36], such as T-box, GATA, bHLH and MADS, but also several binding site motifs for TFs that have not been as well studied during human cardiac development, such as ZIC and NFI (CTF motif) at the mesoderm and vCM stages, respectively (Fig 1b, Extended Data Fig. 2a). Integrating corresponding RNA-seq data including TF expression with this motif analysis revealed specific TFs, such as GATA4 and TBX5, that may bind to these sites in a stage-specific manner to regulate the stepwise fate commitment of pluripotent stem cells (PSCs) into CMs (Extended Data Fig. 2a). Consistent with these findings, we identified dynamic super-enhancers (SEs) using H3K27ac ChIP-seq signal as described previously[37,38] (Extended Data Fig. 2b), and their cognate genes (Extended Data Fig. 2c), which include many TFs with highly conserved roles in heart development[34], whose expression correlate with the activation of stage-specific enhancers harboring their TF binding motif (Extended Data Fig. 2a). Together these results provide insight into the epigenetic developmental patterns and dynamic TF utilization during human CM/vCM differentiation.

Comparative epigenomic analysis of cardiac gene regulatory programs reveals human gained cis-regulatory elements, which are associated with cardiac attributes.

To discover the evolutionary genomic changes that may underlie differences in human cardiac GRNs and contribute to human-specific aspects of heart function and development, we identified human-gained enhancer and promoter activity through a comparison of our human PSC CM differentiation epigenomic data with epigenomic data reported from mouse PSC CM differentiation[15,27]. Toward this end, we examined orthologous DNA regions[22] between mouse and human genomes and compared our PSC, Mes, CP and CM samples with respective samples from mouse PSC-derived cardiac differentiations. Regions for which the H3K27ac signal was 4-fold stronger in human cells compared to those from corresponding mouse cells at each respective stage were considered human-gained promoters/enhancers, while promoters/enhancers below this cutoff were considered stable (Fig. 2a). As a result, we discovered several thousand HGEs (Fig. 2a), which reside mainly in intronic and intergenic regions (Extended Data Fig. 3a), including those within SEs at each cardiac developmental stage (Fig. 2a, Extended Data Fig. 3b, Supplementary Tables 1 and 2). On the other hand, only hundreds of human-gained promoters were observed (Fig. 2a). Comparing the HGEs of each stage individually to all examined mouse stages combined confirmed that the HGEs had gained activity across all mouse stages (Extended Data Fig. 3c). Examining the expression of genes nearest to each HGE revealed that HGEs appear to drive more stage-specific gene expression compared to stable enhancers as indicated by the lower Shannon entropy score[39] (Extended Data Fig. 3d), thus suggesting that these HGEs may be involved in more specialized cardiac cellular processes/functions. To investigate whether these HGEs may be gained through cis-regulatory mechanisms, we interrogated the underlying DNA sequences for TF motifs. Enhancers active in both human and mouse CMs (stable enhancers) were similarly enriched for TF motifs in both human and mouse genomes (Fig. 2b). However, HGEs were enriched for these TF motifs in the human genome as compared to the orthologous regions in the mouse genome, whereas, conversely, mouse-gained enhancers (MGEs, Supplementary Table 3) were enriched for these motifs in the mouse genome compared to the orthologous regions in the human genome (Fig. 2b). These results further support the notion of cis-regulatory changes in TF motifs underlying evolutionary differences in enhancer activity[40]. To identify potentially important human-gained developmental regulatory elements, we further examined super-enhancer/SE domains[41], which are topological domains with usually one critical developmental regulator controlled by an extensive SE[41]. As a result, we observed several hundred of these gained SE domains at each cardiac developmental stage, which were frequently associated with TFs (Extended Data Fig. 3e, f). To further investigate whether identified HGEs may have *in vivo* activity, we examined the VISTA enhancer browser database[42]. This analysis confirmed that cardiac HGEs could drive cardiac reporter activity *in vivo* (Fig. 2c, Extended Data Fig. 4a and Supplementary Table 4). As examples of the *in vivo* cardiac HGE activity, HGEs within *SCNA5A*[43,44] and near *RHOBTB2*[45] were found to drive the expression of a LacZ-reporter in mouse embryonic hearts (Fig. 2c). Supporting that HGEs may provide additional gene regulation[22], both genes displayed increased expression in human CMs compared to mouse CMs (Fig. 2c).

To examine biological processes that may be modulated by HGEs, we analyzed GO term enrichment of neighboring genes using GREAT analysis[46]. GO-terms involved in developmental processes were enriched at the mesoderm stage, whereas GO-terms related to CM-relevant processes such as sarcomere organization and myofibril assembly were enriched at the cardiac progenitor (CP) and CM stages (Extended Data Fig. 4b). Furthermore, we discovered HGEs that harbor genetic sequence variants associated with cardiac traits which particularly differ between human and mouse including conduction-related phenotypes such as heart rate (Fig. 2d). Supporting these findings, we identified an HGE that contained a variant associated with an increased resting heart rate phenotype[47] (Fig. 2e). This HGE could drive human CM-induced reporter activity, which was reduced when introducing the risk allele variant into this HGE reporter (Fig. 2f). Notably, the human risk allele has the same sequence as in the mouse genome (Fig. 2e), further supporting that this HGE may contribute to the modulation of resting heart rate as mice have significantly faster resting heart rates than humans[17]. Furthermore, the activity of this HGE correlates with the expression of the neighboring gene *CPNE5* during cardiomyocyte development suggesting that this gene, which belongs to a family of proteins involved in calcium handling/signaling[48], may be the target gene for this HGE. In line with these findings, examination of additional HGE loci overlapping with GWAS variants associated with atrial fibrillation and QRS duration further corroborated the notion that HGEs may modulate human cardiac traits and disease (Extended Data Fig. 4c–f). These enhancers are near and within *HCN4* and *PRDM16*, respectively, which have roles in cardiac function and disease[49–52] and display gene expression levels that differ between mouse and human CMs (Extended Data Fig. 4c, e). To further interrogate whether HGEs may be also associated with human CHDs, we interrogated recently published whole genome sequencing data from CHD patients and their unaffected parents[7]. This analysis uncovered a subset of CHD *de novo* variants[7] that overlapped with HGEs across the four examined stages (Extended Data Fig. 4g). Overall, these results support that HGEs may enable the human heart to acquire its species-specific functional cardiac features through modulating the expression of genes that specifically regulate these cardiac processes/functions and that mutations in these HGEs may contribute to human CHD.

HGEs are associated with developmental regulators during early cardiogenesis.

Although we discovered HGEs that contain genetic variants associated with functional cardiac traits (Fig. 2d–f), we further investigated whether they may also participate in directing developmental/structural differences between human and mouse hearts. Because of their association to key developmental regulators including TFs[41], we particularly examined super-enhancer domains which may have acquired HGEs during early heart development (Extended Data Fig. 3e), including pluripotent stem cell and mesoderm stages. To investigate whether these HGEs may alter or expand the role of these regulators, we focused on HGEs that are associated with well-known developmental regulators whose roles in mammalian heart development remain less clear (Extended Data Fig. 3f). Because mutations in *ZIC3* can lead to left-right asymmetry disorders including congenital heart disease defects[53–55] but the role of *ZIC3* in directly regulating heart development is less clear[53,54,56], we specifically investigated whether HGEs may exist in the *ZIC3* locus to extend the function of this TF from its conserved role in embryonic primitive node

development[53–55] to an acquired role in early human heart development. Examination of the *ZIC3* locus identified several active enhancers surrounding *ZIC3*, which is the only coding gene in a ~1 Mb window (Fig. 3a). Hi-C chromatin organization analysis[27] revealed that *ZIC3* and its corresponding enhancers interact with the *ZIC3* promoter and form a dense gene/super-enhancer chromatin interaction network that is contained within an insulated neighborhood (i.e. topologically associating domain/TAD)[37,38,41], and these interactions exist mainly at the mesoderm stage (Fig. 3a, Extended Data Fig. 3f). One of these enhancers is active in both mouse and human PSC and mesoderm (Fig. 3a, St1) and corresponds with *ZIC3* expression at these stages (Fig. 3b), whereas the other four enhancers are active mainly at mesoderm stage (Fig. 3a, St2 and HGE1–3), with HGE1–3 being specific to human mesoderm (each HGE has H3K27ac enrichment that is ~10-fold higher compared to the orthologous mouse region). While the St1 and St2 enhancers may contribute to expression of *ZIC3* at the PSC and Mes stages, the human-gained activity of HGE1–3 may contribute to not only the increased *ZIC3* expression in human mesoderm (Fig. 3b) but also the gene regulatory and expression differences between human and mouse mesoderm including enrichment of ZIC motifs in human mesoderm active enhancers (Extended Data Fig. 5a, b, Supplementary Table 5). Supporting these findings, the chromatin organization of the human *ZIC3* locus is dynamically regulated during early cardiogenesis to tightly control the temporal expression of *ZIC3*. Corresponding to *ZIC3* stage-specific enhancer activity and gene expression induction, distinct chromatin enhancer-enhancer and enhancer-gene interactions involving the St2 enhancer (St1 is within the *ZIC3* gene and its interactions cannot be resolved by Hi-C) and HGE1–3 enhancers within the *ZIC3* TAD increase in strength and become more distinct from the PSC to mesoderm stages; however, the majority of these interactions are lost at the cardiac progenitor stage when *ZIC3* expression has diminished (Fig. 3a, b). Together, these findings reveal potential mechanisms for how transcriptional regulators may modify their developmental roles through the acquisition of enhancers which alter spatio-temporal control of gene expression.

Because *ZIC3* has a well-established and highly conserved role in left-right asymmetry specification but a less clear function during cardiogenesis[53,54], we investigated whether *ZIC3* enhancers including HGEs are differentially activated in specific subpopulations of early mesodermal progenitor cells to potentially direct specific developmental events including human cardiac specification. To this end, we interrogated chromatin accessibility of hPSC-derived mesodermal cells at single cell resolution using single nucleus (sn)ATAC-seq[57,58] and obtained accessible chromatin profiles for 27,972 nuclei, with a median of 6318 fragments mapped per nucleus (Supplementary Table 6). Applying ArchR[59] to this snATAC-seq data, we identified nine clusters (Fig. 3c), which were annotated based on marker genes identified by gene scores that predict how highly expressed a gene will be based on the accessibility of the promoter and distal CREs in the vicinity of the gene (Fig. 3d, Supplementary Table 7). For example, *POU5F1* and *SOX2* were used to classify cluster 1 as cells resembling PSCs, whereas *PDGFRA*, *GATA4*, and *RBM20* were used to annotate cluster 7 as pre-cardiac mesoderm (Fig. 3d). Thus, single nuclear chromatin accessibility analyses revealed distinct subpopulations of progenitor cells which may participate in early cell fate decisions of the mesoderm including the specification of early cardiac progenitors.

To discover putative promoters and enhancers regulating these subpopulations at the mesoderm stage, we aggregated snATAC-seq data from each cell cluster and identified their accessible chromatin regions using MACS2[60]. To this end, snATAC-seq peaks from all nine cell clusters were initially merged into a collection of 250,715 CREs (Supplementary Table 8), of which 70.1% of these CREs were at least 2 kb away from annotated promoter regions, and 77,605 CREs displayed cluster-specific accessibility (Extended Data Fig. 6a, Supplementary Table 9). Examination of TF motif enrichment in cluster-specific snATAC-seq CREs revealed a wide range of TF binding signatures that vary between the clusters (Extended Data Fig. 6b, Supplementary Table 10). *ZIC3* motif enrichment is present across all clusters and may be more enriched in the LPM and PCM clusters (Extended Data Fig. 6b, c). Together, these results reveal specific candidate CREs and developmental regulators that establish gene regulatory networks directing the mesoderm to distinct progenitor subtypes including pre-cardiac progenitors.

To examine whether specific enhancers within the *ZIC3* locus may regulate *ZIC3* expression in distinct subpopulations at the mesoderm stage, we interrogated the chromatin accessibility profiles of the cell types identified by our snATAC-seq studies. Combining this analysis with our bulk ChIP- and ATAC-seq data, we observed that the *ZIC3* stable enhancers and HGE1 were accessible across all identified cell subpopulations, whereas HGE2 and HGE3 were more selectively accessible in the lateral plate (cluster 6) and pre-cardiac mesoderm (cluster 7) subpopulations (Fig. 3e). Consistent with these findings, we further discovered that the *ZIC3* motif, along with motifs of known cardiac-associated TFs, were enriched in the lateral plate and pre-cardiac mesoderm clusters (Extended Data Fig. 6b, c), supporting a potential key role for *ZIC3* in directing early cardiogenesis in humans. Thus, these results suggest that specific enhancers may have been gained to expand the role of *ZIC3* in human cardiac development through regulating the increased *ZIC3* expression in specific subpopulations of human mesoderm involved in early cardiogenesis. In the line with this notion, we discovered from *in vivo* mouse mesoderm scRNA-seq data[61] that *Zic3* was expressed in only a subset of the mesoderm populations and that those populations were not LPM or cardiac progenitors (Extended Data Fig. 6d).

Human *ZIC3* directs human cardiac progenitor differentiation into cardiomyocytes.

To investigate whether human *ZIC3* and its HGEs may direct human cardiogenesis, we created and examined hPSCs harboring CRISPR-Cas9 genome-edited *ZIC3* gene and HGE deletions. We initially investigated whether *ZIC3* is involved in directing human CM development and observed that *ZIC3* gene-deleted hPSC lines (Fig. 4a, b) displayed impaired CM differentiation (Fig. 4c), supporting that *ZIC3* may have a role in human cardiac lineage specification. Examining expression profiles between wild-type and *ZIC3* gene-deleted hPSCs at the mesoderm stage revealed 1053 differentially regulated genes (Fig. 4d, Supplementary Table 11) out of which 656 were down-regulated while 397 genes were upregulated (Fig. 4d, Supplementary Table 11). Consistent with the conserved role of ZIC3 in left-right asymmetry and its potential cardiac role in human cardiogenesis, many of these downregulated genes included not only *NODAL* and several NODAL target genes, such as *LEFTY2,* but, notably, also genes involved in early cardiac development such as *MESP1/2*, *HOPX, APLNR* and *WNT3A* (Fig. 4d, e, Supplementary Table 11). Further

supporting these findings, Gene Ontology (GO) enrichment analysis revealed that the down-regulated genes are associated with “Embryonic morphogenesis”, “Tissue morphogenesis” and “Heart development” GO terms (Fig. 4d), whereas up-regulated genes are associated with “Cell morphogenesis involved in differentiation” and “Epithelial cell differentiation” GO terms (Fig. 4d). Notably, *ZIC3*-dependent genes (i.e., genes whose expression are downregulated after *ZIC3* ablation) are enriched among the genes up-regulated in human mesoderm compared to mouse mesoderm and include early cardiac developmental genes such as *APLNR*, *PDGFRA*, *MESP1* and *WNT5A* (Fig. 4f). Thus, these results suggest that *ZIC3* may have adopted an expanded cardiac developmental role in humans when compared to mice through participating in the gene regulation and expression of known early cardiac developmental regulators.

A *ZIC3* HGE regulates *ZIC3* cardiac mesoderm expression and cardiogenesis.

To further investigate whether *ZIC3* HGEs may regulate *ZIC3* mesoderm-specific expression and early human cardiac development, we examined hPSCs containing individual and combinatorial CRISPR-Cas9 genome-edited deletions of the three HGEs in the *ZIC3* locus (Fig. 5a and Extended Data Fig. 7). Among the individual HGE hPSC deletion lines, only the HGE3-deleted hPSC line displayed a significant decrease in *ZIC3* expression in hPSC-mesoderm cells (Fig. 5b and Extended Data Fig. 8a). Deletion of all three HGEs in combination (HGE TKO) did not lead to a further decrease in *ZIC3* expression compared to HGE3-deleted hPSC lines (Fig. 5b), supporting that HGE3 may be crucial for driving increased *ZIC3* mesoderm expression. In line with the mesoderm-specific activity of HGE3 (Fig. 5a), *ZIC3* gene expression after HGE3 or TKO deletion was not altered at the PSC stage (Extended Data Fig. 8b). Consistent with *ZIC3* and its HGEs residing in their own topological domain, the expression of genes neighboring *ZIC3* (*RBMX*, *FGF13*) in adjacent TADs was not affected, confirming that HGE3 specifically controls *ZIC3* mesoderm induction (Extended Data Fig. 8c). Supporting that *ZIC3* HGE3 may specifically regulate human cardiac development, expression of *HOPX* and *APLNR*, both of which are *ZIC3*-dependent genes (Fig. 4d, e), was downregulated in the HGE3 and TKO lines (Fig. 5b and Extended Data Fig. 8a), but expression of *NODAL* and the *NODAL* target gene *LEFTY2*, which are involved in left-right asymmetry, was not significantly altered (Fig. 5b and Extended Data Fig. 8a). In line with the downregulation of *HOPX* and *APLNR*, we discovered that CM differentiation was reduced significantly in the HGE3 and TKO lines (Fig. 5c and Extended Data Fig. 8d). Thus, these results support that HGE3 increases *ZIC3* expression in human mesoderm cells to facilitate a human-gained role for *ZIC3* in CM differentiation, which may be distinct from its highly conserved role in Nodal signaling/left-right asymmetry. Moreover, together with our snATAC-seq data showing that HGE3 is accessible only in a subset of cells at the mesoderm stage, these data further suggest that HGE3 drives the gained role of ZIC3 through mesodermal cell subpopulation-specific activity.

To investigate whether cis-regulatory changes may have contributed to the evolutionary emergence of HGE3, we examined the DNA sequence within HGE3, focusing particularly on regions exhibiting high chromatin accessibility (ATAC-seq signal, Fig. 5d), which is indicative of TF binding. We found that all sequence motifs are conserved between

mouse and human except for a FORKHEAD motif in HGE3 (Fig. 5d). While the pioneer TFs FOXA1–3 are more well known for their role in endoderm and ectoderm formation[62], a recent report identified a role for *Foxa2* in mouse cardiac development by regulating a progenitor pool that gives rise to ventricular CMs[63]. To explore whether FOXA1/2 contribute to HGE3 enhancer activity and regulation of *ZIC3* expression, we examined FOXA2 ChIP-seq data from the human mesendoderm stage[64], which precedes the mesoderm stage, and found that FOXA2 binds to HGE3 (Fig. 5d). Using ChIP-qPCR, we confirmed specific binding of both FOXA1 and FOXA2 to HGE3 in hPSC-derived mesodermal cells (Fig. 5e). To investigate whether this evolutionary divergent DNA sequence may contribute to HGE3 activity, we employed an enhancer-dependent luciferase reporter assay in hPSC-derived mesoderm cells and discovered that HGE3 was able to drive mesoderm-induced reporter activity (Fig. 5f). Notably, this reporter activity was impaired after altering the human FORKHEAD motif to the corresponding mouse sequence (Fig. 5f). Thus, these results support that the human (and primate) genome may have acquired changes in the HGE3 region that generate a TF binding site for the pioneer TFs FOXA1/2 (Fig. 5d), which may underly not only the evolutionary acquisition of HGE3 that directs mesoderm-specific *ZIC3* expression but also *ZIC3*'s subsequent function in human cardiac development.

Discussion

Species-specific attributes are often driven by the evolutionary creation of new CREs, including enhancers[1,19]. Recent studies have shown the acquisition of such enhancers associated with human-specific attributes for limb and brain during late development[22,23]. Here, we reveal underlying cis-regulatory mechanisms for how the human heart may have gained specific early developmental and functional cardiac attributes to meet the circulatory and metabolic requirements of the human body through similar evolutionary gain of human enhancers.

From our comparative epigenomic analyses between developing human and mouse cardiomyocytes, we uncovered human-gained activity at promoters and enhancers during CM differentiation that contribute to human-specific cardiac function and development, and furthermore observed that genetic variants associated with human cardiac phenotypic/disease traits, particularly those differing between human and mouse, are enriched in these HGEs. Additionally, we discovered HGEs within genomic loci of transcriptional regulators at early stages of human cardiogenesis (i.e., mesoderm stage) that may potentially expand the role of these regulators in human heart development. Specifically, we identified HGEs in the *ZIC3* locus that induce *ZIC3* expression at the mesoderm stage, which may extend the function of *ZIC3* beyond its highly conserved role in early left-right asymmetry determination[53–55,65–68]. Further supporting the notion that *ZIC3* may have an expanded cardiac role in humans, another *ZIC* family member, *ZIC2*, was recently reported to control the cell fate of early mesodermal precursors to human heart progenitors[69]. Through exploring the mesoderm chromatin landscape at single-cell resolution, we identified HGEs within the *ZIC3* locus that display differential chromatin accessibility profiles between mesoderm subpopulations. In particular, the *ZIC3* HGE3 enhancer exhibited increased accessibility in lateral plate and pre-cardiac mesodermal subpopulations suggesting that

HGE3 may have facilitated an acquired developmental role for *ZIC3* by inducing *ZIC3* expression in specific human mesodermal cell subpopulations. Supporting this notion, genetic ablation of these HGEs revealed that HGE3 is required for mesoderm induction of *ZIC3* as well as human CM differentiation. Thus, the subpopulation-specific accessibility of *ZIC3* HGE enhancers in the mesoderm may function to extend the role of ZIC3 in the lateral plate mesoderm and more specifically pre-cardiac mesodermal progenitors and their subsequent cardiac developmental derivatives in humans. As additional cardiac developmental single-cell RNA-seq and ATAC-seq data from other species, including those from non-human primates[70], becomes available, it will be interesting to investigate whether this induction of *ZIC3* is human-specific or a more broad phenomenon across other large mammals.

The evolutionary acquisition of enhancers is likely to be driven by the gain of TF motifs[1,19]. Recent work using massively parallel reporter assays identified human-specific sequence substitutions that affected cortical neuron HGE reporter activity and in some cases altered TF motifs[40]. Our findings provide further insight into human cis-regulatory evolution by revealing that the functionally validated *ZIC3* HGE3 enhancer acquired a FORKEAD motif (when compared to the orthologous region in rodents) that can be bound by the pioneer TFs FOXA1 and FOXA2 in human mesoderm cells and is required for enhancer activity in these mesoderm cells. Interestingly, *Foxa2* was recently found to mark and control a subset of cardiac progenitors that contributes to the differentiation of ventricular cardiomyocytes and ventricular chamber size in mice[63]. Given our demonstration that FOXA2 binds HGE3 and that its FORKHEAD motif is required for full enhancer activity in reporter assays, our results suggest the possibility that the HGE-driven *ZIC3* gene expression induction in human mesoderm cell (sub)populations may expand the cardiac progenitor pool that contributes to the ventricles of the heart.

Finally, while many patients with *ZIC3* mutations exhibit a spectrum of developmental left-right asymmetry defects[71,72], some patients with *ZIC3* mutations display isolated CHD[65,73,74], including ventricle-related cardiac defects such as ventricular septal defects[55,65,74]. Through the expanded role of *ZIC3* in cardiac-like mesoderm, our findings may provide insight as to how *ZIC3* mutations lead to these isolated CHDs through directly affecting cardiogenesis rather than left-right asymmetry patterning. Supporting this notion, we observed that both *ZIC3* gene-deleted and *ZIC3* HGE3-deleted hPSC lines display impaired CM differentiation.

While informative, our study has limitations. Results from our studies are based on *in vitro* differentiations using hPSCs to study the earliest stages of human cardiac development that cannot be easily studied *in vivo* for ethical and technical reasons. However, future work may use *in vitro* cardiac organoid models, which allow cells to be studied in a more complex, *in vivo*-like context[75–79]. Although our human stages are similar to the corresponding mouse stages based on gene expression, differences due to timing variation between the human and mouse cardiac differentiations cannot be excluded. Finally, we cannot discount that some of the HGE-orthologous mouse regions do have high enhancer activity in other cell types not studied here or perhaps even in similar cell types but in a *in vivo* context.

Overall, our results illuminate how the gene regulatory networks controlling human heart development and function may have evolved through the acquisition of human/primate-specific enhancers whose regulatory sequences have been altered to enable additional TF binding that mediates gained enhancer activity. Thus, these findings may provide new avenues in the future for exploring how HGEs and their genetic modifications may participate in directing human-specific aspects of both heart development and disease.

Materials and Methods

Experimental procedures

Cell culture, differentiation, and collections for large-scale genomic studies.— For the large-scale genomic and transcriptomic human pluripotent stem cell (hPSC) cardiac studies (i.e., ChIP-Seq, RNA-Seq, and ATAC-seq), an engineered hPSC H9-*MYL2:H2B-GFP* reporter transgenic line that specifically expresses H2B-GFP (the human histone H2B fused to green fluorescent protein) in differentiated ventricular cardiomyocytes was used as previously described[26,27]. This H9 *MYL2:H2B-GFP* line was differentiated into cardiomyocytes and their corresponding developmental stages utilizing a well-established WNT-based cardiomyocyte differentiation protocol[80] that we have previously used to efficiently create hPSC-derived cardiomyocytes[27]. Specifically, the following developmental cardiac cell populations were generated and collected: day 0 (hPSC); day 2 (mesoderm, Mes); day 5 (cardiac mesoderm, CMes); day 7 (cardiac progenitors, CP); day 15 (cardiomyocytes, CMs); and day 80 (ventricular cardiomyocytes, vCMs). The differentiation efficiency was assessed at specific developmental cardiac stages by flow cytometry using antibodies against BRACHYURY (AF2085; R&D Systems) for day 2/mesoderm; KDR and PDGFRA (AF357P and AF1264A, respectively; R&D Systems) for day 5/cardiac mesoderm; and TNNT2 (MA5–12960; Thermo Fisher Scientific) for day 15/cardiomyocytes. The differentiation success for collected samples at days 2–7 was further confirmed by reserving a portion of the cells gathered for collection and then maintaining them until day 15, when cardiomyocyte differentiation could be assessed using TNNT2 flow analysis. The collected samples that resulted in high differentiation efficiency (>90%) were then further processed and analyzed for molecular studies (see below). For the day 80 ventricular cardiomyocytes, we additionally sorted and purified these cells based on *MYL2:H2B-GFP* before their processing and analyses. For the RNA-seq studies, ~2 million dissociated cells were washed in phosphate buffered saline (PBS), lysed in TRIzol and then stored at −80°C until further processing. For the ChIP-seq studies, dissociated cells were washed in PBS, fixed in 1% formaldehyde for 10 minutes (min), quenched with 0.125 M glycine for 5 min at room temperature (RT) followed by 10 min on ice, and finally washed twice with PBS. These fixed cell pellets were then snap-frozen and stored at −80°C until further processing. For ATAC-seq, cells were slow-frozen overnight (O/N) in freezing media (E8 supplemented with 10% DMSO) at −80°C and stored in liquid nitrogen until further processing. RNA-seq, H3K27ac ChIP-seq and Hi-C samples and data sets were generated as previously described[27] and available online from GEO under accession number GSE116862.

Flow cytometry.—PSC, Mes and CMes cells were dissociated using Accutase (STEMCELL) while CP, CM and vCM cells were dissociated using collagenase IV

(Invitrogen) followed by TrypLE (Invitrogen). Dissociation was quenched using RPMI + 20% FBS and cells were centrifuged for 5 min at 200*g*. For TNNT2 flow cytometry, cells were resuspended and fixed in PBS/4% para-formaldehyde, while for PDGFRA/KDR flow cytometry and GFP fluorescence, cells were examined without fixation. Fixed cells were washed and stained with antibodies in the dark for 1 hour (h) at RT or at 4°C O/N. Cells were washed twice and then incubated with secondary antibodies for 1 h at RT. Cells were filtered to obtain a single-cell suspension for flow cytometry analysis on a Guava easyCyte 8HT (Millipore). Data was processed and analyzed using FlowJo software (BD Biosciences). For each experiment, WT cells and a set of clones for each KO were differentiated, processed, and analyzed together for flow cytometry analysis.

CRISPR/Cas9-mediated genome editing.—To interrogate the functional significance of *ZIC3* and its human-gained enhancer (HGE) elements, selected sequences were genetically deleted in H9 *MLC2v:H2B-GFP* hPSCs using CRISPR/Cas9 genome editing strategies. For *ZIC3* gene knockout (KO), exon1 was targeted. For *ZIC3* HGE1–3 KOs, the following regions were targeted (using hg19 assembly): HGE1 - chrX:136,506,096–136,512,989; HGE2 - chrX:136,997,377–136,999,652; HGE3 - chrX:137,142,246–137,143,948. Specific guide RNA (gRNA) pairs (Supplementary Table 12) with binding sites surrounding the target region were designed using the web-based software tool CRISPOR[81], which generates gRNA sequences for optimal targeting of regions of interest while minimizing potential off-target effects. These identified gRNAs were then synthesized in vitro using the GeneArt Precision gRNA Synthesis kit (Invitrogen) following the manufacturer's protocol. 1×10^5 H9 *MLC2v:H2B-GFP* hPSCs were seeded in 12-well plates, and one day later, a pair of ribonucleoprotein complexes containing 2 μg Cas9 protein (New England Biolabs) and 400 ng of in vitro-transcribed gRNA were transfected using Lipofectamine Stem Transfection Reagent (Invitrogen). Three days after transfection, the cells were diluted and clonally expanded for clonal line generation and PCR-based genotyping. To generate the HGE triple KO (TKO) lines, HGE2 and HGE3 were sequentially deleted in an HGE1 KO clone. After confirmation of genome-edited clones by PCR using external and internal primers for HGE deletions, and external PCR primers and western blot/WB for the *ZIC3* gene deletions, at least two successfully targeted clones per deletion were selected for downstream analyses.

ChIP-seq.—ChIP-seq experiments for H3K4me1, H3K4me3 and H3K27me3 at all conditions were performed as previously described using chromatin samples from the same collections[27]. The experiments were conducted according to experimental guidelines and read depth standards of the Encyclopedia of DNA Elements (ENCODE) consortium. Harvested cells were washed in PBS and fixed in 1% formaldehyde for 10 min at RT. The reaction was quenched with 125 mM glycine for 5 min at RT, followed by 10 min incubation on ice. Next, the cells were washed twice with PBS. Crosslinked cells were lysed and chromatin sheared using a Branson Sonifier 450 (20 cycles; 15 seconds (s) on and 45 s off at power 3). Detailed protocols can be found on the ENCODE homepage for immunoprecipitation: (https://www.encodeproject.org/documents/89795b31-e65a-42ca-9d7bd75196f6f4b3/@@download/attachment/Ren%20Lab%20ENCODE%20Chromatin%20Immunoprecipitation%20Protocol_V2.pdf)

and library preparation (https://www.encodeproject.org/documents/4f73fbc3-956e-47ae-aa2d41a7df552c81/@@download/attachment/Ren_ChIP_Library_Preparation_v060614.pdf). All ChIP-seq libraries were sequenced on a HiSeq 2500/4000 [50-base pair (bp) single ends (SE); Illumina]. Antibody information can be found in the Reporting Summary.

ATAC-seq.—ATAC-seq was performed on 50,000 cells as described previously[82] with minor modifications. Cell suspensions were thawed at 37°C and pelleted for 5 min at 500*g* at 4°C. Cell pellets were resuspended in 250 μL permeabilization buffer [5% BSA, 0.2% IGEPAL-CA630, 1 mM DTT and cOmplete EDTA-free protease inhibitor cocktail (Roche) in PBS] and incubated on an overhead shaker for 5 min at 4°C, followed by centrifugation for 5 min at 500 g. The cell pellet was resupended in 25 μL tagmentation buffer (36.3 mM Tris-acetate (pH = 7.8), 72.6 mM K-acetate, 11 mM Mg-acetate, 17.6% DMF) and counted on a hemocytometer. 40,000 nuclei were used for tagmentation and the reaction volume was adjusted to 19 μL using tagmentation buffer. After addition of 1 μL Tn5 transposomes (Illumina), tagmentation was performed at 37°C for 60 min with shaking (500 rpm). Next, samples were purified using MinElute columns (Qiagen), PCR-amplified for 9 cycles with NEBNext High-Fidelity 2X PCR Master Mix (New England Biolabs, 72°C 5 min, 98°C 30 s, [98°C 10 s, 63°C 30 s, 72°C 60 s] × 9 cycles, 4°C ∞). Amplified libraries were purified using MinElute columns (Qiagen) and Ampure XP Bead (Beckmann Coulter). Sequencing was carried out on a HiSeq2500 or 4000 (50 bp paired end (PE), Illumina).

snATAC-seq.—Mesoderm cells from three independent differentiations were dissociated with Accutase (STEMCELL) at 37°C for 10 min and processed fresh for single cell combinatorial indexing (sci)ATAC-seq[58,83]. Dissociated cells were stained with DAPI (Thermo Fisher Scientific). 500,000 viable cells (DAPI negative) were sorted using a SH800 sorter (Sony). Cells were pelleted with a swinging-bucket centrifuge (500 g, 5 min, 4°C; 5920R, Eppendorf) and resuspended in 1 mL OMNI Buffer [10 mM Tris-HCL (pH 7.5), 10 mM NaCl, 3 mM $MgCl_2$, 0.1% Tween-20 (Sigma), 0.1% IGEPAL-CA630 (Sigma) and 0.01% Digitonin (Promega) in water[84]] and pelleted with a swinging-bucket centrifuge (500*g*, 5 min, 4°C; 5920R, Eppendorf). After centrifugation, permeabilized nuclei were resuspended in 500 μL of high-salt tagmentation buffer [36.3 mM tris acetate (pH 7.8), 72.6 mM K-acetate, 11 mM Mg-acetate, and 17.6% *N,N*′-dimethylformamide] and counted using a hemocytometer. The concentration was adjusted to 2000 nuclei/9 μL, and 2000 nuclei were dispensed into each well of a 96-well plate per sample (96 tagmentation wells per sample; samples were processed in batches of two to four samples). For tagmentation, 1 μL of barcoded Tn5 transposomes was added using a BenchSmart 96 (Mettler Toledo), mixed five times, and incubated for 60 min at 37°C with shaking (500 rpm). To inhibit the Tn5 reaction, 10 μL of 40 mM EDTA (20 mM final) were added to each well with a BenchSmart 96 (Mettler Toledo), and the plate was incubated at 37°C for 15 min with shaking (500 rpm). Next, 20 μL of 2× sort buffer (2% BSA and 2 mM EDTA in PBS) were added using a BenchSmart 96 (Mettler Toledo). All wells were combined into a separate fluorescence-activated cell sorter tube for each sample and stained with DRAQ7 at 1:150 dilution (Cell Signaling Technology). Using an SH800 (Sony), 20 nuclei per sample were sorted per well into eight 96-well plates (total of 768 wells) containing 10.5 μL of

Elution Buffer (EB; Qiagen) [25 pmol of primer i7, 25 pmol of primer i5, and 200 ng of BSA (Sigma-Aldrich)]. Preparation of sort plates and all downstream pipetting steps were performed on a Biomek i7 automated workstation (Beckman Coulter). After addition of 1 μL of 0.2% SDS, samples were incubated at 55°C for 7 min with shaking (500 rpm). 1 μL of 12.5% Triton X was added to each well to quench the SDS. Next, 12.5 μL of NEBNext High-Fidelity 2× PCR Master Mix (New England Biolabs) were added, and samples were PCR-amplified [72°C 5 min, 98°C 30 s, (98°C 10 s, 63°C 30 s, 72°C 60 s) × 12 cycles, held at 12°C]. After PCR, all wells were combined. Libraries were purified according to the MinElute PCR Purification Kit manual (Qiagen) using a vacuum manifold (QIAvac 24 Plus, Qiagen), and size selection was performed with SPRISelect reagent (Beckman Coulter; 0.55× and 1.5×). Libraries were purified one more time with SPRISelect reagent (Beckman Coulter; 1.5×). Libraries were quantified using a Qubit fluorimeter (Life Technologies), and a nucleosomal pattern of fragment size distribution was verified using a TapeStation (High Sensitivity D1000, Agilent). Libraries were sequenced on NextSeq 500 or or HiSeq4000 sequencers (Illumina) using custom sequencing primers with the following read lengths: 50 + 10 + 12 + 50 (read 1 + index 1 + index 2 + read 2).

RNA-seq.—RNA-seq was performed as previously published[27]. For *ZIC3* KO and wild-type (WT) control RNA-seq, total RNA from mesoderm-stage cells was extracted and purified using TRIzol (Thermo Fisher Scientific) from three independent replicates with each replicate representing a pool of three culture wells. RNA-seq libraries were prepared from 1–2 μg total RNA using the Illumina TruSeq Stranded mRNA Library Prep Kit Set A (RS-122–2101; Illumina) or Set B (RS-122–2102; Illumina). Sequencing was carried out on a HiSeq 2500/4000 (50/100 bp PE; Illumina).

In situ Hi-C.—In situ Hi-C data sets were previously published[27] and visualized using HOMER[85] and Juicebox[86].

Real-time quantitative/qPCR.—For quantitative gene expression analysis, RNA was extracted using Trizol (Invitrogen) following manufacturers protocol. cDNA was transcribed from 1 μg RNA using iScript Supermix (Bio-Rad). Gene expression was analyzed using PowerUp SYBR Green Mastermix (ThermoFisher) and run on a CFX Connect qPCR machine (Bio-Rad). Primer sequences are listed in Supplementary Table 12.

ChIP-qPCR.—ChIP-qPCR studies were performed similar to ChIP-seq studies except that chromatin was sheared to a larger average size of 0.5–1 kb. Chromatin was immunoprecipitated using FOXA1 and FOXA2 antibodies (Active Motif) or IgG as a control. Enrichment was assessed using primers within *ZIC3* HGE3 or a control region lacking enhancer activity. qPCR was performed using PowerUp SYBR Green Mastermix (ThermoFisher) and run on a CFX Connect qPCR machine (Bio-Rad). Data was plotted as percentage of input to correct for potential differences in input amounts. Primer sequences are listed in Supplementary Table 12.

Western blot.—To confirm *ZIC3* KO clones, lysates were prepared using NuPAGE LDS sample buffer (Invitrogen) and run on a 4–12% NuPAGE gradient gel (Invitrogen) for Western analysis. Gels were transferred using standard procedures[87]. Nitrocellulose

membranes were stained with antibodies for ZIC3 (Abcam) and CTCF (Abcam) as loading control.

Luciferase reporter assays.—Synthetic DNA fragments containing 160 bp of the wild-type (WT) and mutant (Mut) versions for the regions harboring the GWAS variants rs235353 (chr21:46,188,785–46,188,945), rs7172038 (chr15:73,667,174–73,667,334), and rs79823300 (chr1:3,244,715–3,244,895) and *ZIC3* HGE3 (chrX:137,143,536–137,143,696) were obtained from IDT and cloned into pGL4.23 [luc2/minP] (Promega E8411) luciferase reporter vector using NEBuilder HiFi DNA assembly (New England Biolabs E5520S). 1×10^5 of hPSCs were plated in a Geltrex-coated 24-well plate, and hPSCs were transfected with 100 ng of pGL4.23 plasmid (with no enhancer as a control or with WT/Mut enhancers) and 2 ng of TK:Renilla-luc as an internal transfection control using Lipofectamine Stem Transfection Reagent (Invitrogen STEM00003). Media was replaced with fresh media after 24 hours of transfection. At 48 h after transfection, half of the wells were differentiated into mesoderm cells as described above while the other half of the wells were left untreated to serve as controls representing the PSC stage. All wells were washed with PBS 48 h later and processed for luminescence assays using a Dual-Luciferase Reporter Assay System (Promega) following manufacturer's protocol

Data Analysis

ChIP-seq data processing

Read alignment.: Sequencing reads were aligned to the human GRCh37/hg19 or mouse GRCm38/mm10 genomes using Bowtie2[88] using default parameters. PCR duplicates were removed using Picard (http://broadinstitute.github.io/picard/). Only reads aligning uniquely to a single location were used for downstream analysis (MAPQ > 10).

ChIP-seq peak calling.: The identification of ChIP-seq peaks (bound regions) was performed using HOMER's getDifferentialPeaksReplicates.pl tool using ChIP-seq input sequencing as a control. For each ChIP-seq experiment, ChIP-enriched regions (peaks) were found by first identifying significant clusters of ChIP-seq tags and then filtering these clusters for those significantly enriched relative to background sequencing and local ChIP-seq signal. First, we centered raw ChIP-seq tags representing the edge of ChIP fragments by 75 bp to mark the approximate center of fragments isolated in the ChIP experiment. We considered one tag from each unique position to eliminate peaks resulting from clonal amplification of fragments that can arise during library preparation and cluster generation steps in the ChIP-Seq protocol. Putative peaks were then identified by searching for clusters of tags within a sliding window of 200 bp. Putative peaks were then filtered based on the following four conditions. (1) The number of tags in each cluster must exceed a threshold corresponding to a false discovery rate of 0.1% based on the assumption that randomly distributed tags would naturally form clusters given the large number of tags sequenced. This threshold was empirically calculated by randomizing tag positions and repeating the tag clustering procedure to determine the expected number of clusters exceeding each tag threshold. (2) Adjacent peaks must be greater than 500 bp away from one another. (3) Peaks must have at least 4-fold more tags (normalized to total tag count) than the input control sample from the same cell type to eliminate background artifacts. (4) Peaks must have 4-fold

more tags per bp in the peak region (200 bp) relative to the surrounding region (10 kb) to avoid identifying regions with genomic duplications or peaks without localized binding.

ATAC-seq data processing

Read alignment.: ATAC-seq reads were processed as described for ChIP-seq.

ATAC-seq peak calling.: ATAC-seq peak calls were generated as for ChIP-seq peaks except that the "factor" option was used in HOMER's findPeaks tool to account for the narrower peaks expected from ATAC-seq data compared to ChIP-seq data.

RNA-seq data processing

Read alignment and quantification.: Briefly, reads were aligned to hg19 (GRC37) using rnaSTAR[89]. Only reads aligning uniquely to a single location were used for downstream analysis (MAPQ > 10). PCR duplicates were removed using Picard.

Differential gene expression analysis.: Gene expression (FPKM, fragments per kilobase of transcript per million mapped reads) was determined by counting the number of reads overlapping exons for all genes using HOMER's analyzeRepeats.pl tool and the transcriptome annotation from GENCODE (version 34). DESeq2[90] was used to identify differentially expressed genes from unnormalized read counts comparing replicates from control and *ZIC3* KO mesoderm cells.

Bulk genomic assays signal visualization.: BigWig files for ChIP-, ATAC-, and RNA-seq were generated and normalized using HOMER[85].

snATAC-seq data processing

Demultiplexing of snATAC-seq reads.: We obtained FASTQ files for the DNA reads as well as for the combinatorial indexes i5 (768 different PCR indices) and T7 (96 different tagmentation indices) for each sequenced snATAC-seq library. We selected all reads with ≤2 mistakes per individual index (Hamming distance between each pair of indices is 4) and subsequently integrated the full barcode at the beginning of the read name in the demultiplexed FASTQ files using the ATACdemultilex tool (https://gitlab.com/Grouumf/ATACdemultiplex/). For the first replicate, SnapTools (https://github.com/r3fang/SnapTools) was used with bwa aligner to align the reads to the hg38 human reference genome assembly after indexing. After alignment, we converted the reads into fragments using Sinto (https://github.com/timoast/sinto) to be used as input for ArchR. For the other two replicates, the reads were aligned and filtered with Cell Ranger ATAC (https://github.com/10XGenomics/cellranger-atac). The barcoded and aligned fragment files were used as input for ArchR.

snATAC analysis.: The ArchR package[59] was used to process, analyze and visualize the snATAC-seq data. Three biological replicates were merged and batch-corrected using Harmony after an iterative LSI dimensionality reduction. ArchR uses Seurat's[91] graph clustering with "Seurat::FindClusters" for identifying high fidelity clusters at a resolution of 0.79. Clusters were visualized using the "uwot" package in R to generate UMAP analysis. ArchR uses MACS2 (v2.1.2) to call peaks for each cluster and adds a reproducible peak set.

The peak calling and reproducible peak set were run with default parameters. ArchR then computes the peak count matrix per cell after peak calling.

Identification of cluster–specific cis-regulatory regions (CREs).: ArchR uses the computed peak count matrix and wilcoxon test to identify unique differential peaks across the clusters. We considered peaks with p value < 0.05 and log2 fold change (FC) > 0 as cluster-specific.

snATAC CRE motif enrichment analysis.: To identify differentially enriched motifs in CREs for each cluster, we used ArchR to add the cisbp motif set in order to annotate the peak set. We identified differential marker peaks for each cluster, then ran motif enrichment on each set of peaks using all peaks as the background. We kept motifs with p value < 0.05 and log2 FC > 0 cutoffs.

snATAC cluster annotation.: We annotated snATAC-seq clusters using the gene score as a surrogate for gene expression of known marker genes[61,92]. ArchR[59] derives a gene score from chromatin accessibility at promoter regions and further considers CREs within 100 kb of the promoter. We annotated the nine identified clusters as follows: PSC using *POU5F1*, *SOX2*, *TFAP2C* and *KLF4*; neural using *OLIG3*, *SOX1*, *GBX2* and *HES1*; epiblast using *PRICKLE1*, *DPPA4*, *NANOG* and *NODAL*; extra-embryonic using *CDX2*, *GATA2*, *KRT7*; anterior primitive streak/mesendoderm using *EOMES*, *GSC*, *CER1* and *MIXL1*; lateral plate mesoderm using *LEFTY1*, *LEFTY2* and *KDR*; pre-cardiac mesoderm using *PDGFRA*, *GATA4* and *RBM20*; paraxial mesoderm using *TBX6*, *HES7*, *MSGN1* and *DLL1*; and presomitic mesoderm using *RIPPLY2*, *MESP2* and *HEYL*.

Comparative genomic ChIP-seq analysis.: To identify regulatory signals that are species-specific yet are associated with DNA that is homologous between species, we followed a previously published method[22]. We only compared H3K27ac read density between species in regions that are present in each species and have a one-to-one mapping in the genome Multiz alignments. To identify homologous regions between human and mouse, H3K27ac peaks in each species were first mapped to the other using the UCSC Genome Browser's LiftOver tool, requiring that at least 50% of the region was alignable. These regions were then mapped back to their original genome using the same LiftOver tool to ensure there was a one-to-one correspondence. Similarly, ChIP-seq reads from mouse were mapped to the human genome using the LiftOver tool to visualize in the human genome browser. Human- or mouse-gained enhancers were identified by quantifying the H3K27ac ChIP-seq signal (FPKM) in either human or mouse experiments at homologous H3K27ac peaks. Human- or mouse-gained enhancers were then defined as regions with a fold change of 4 after adding a pseudo count of 0.25 to avoid low read density variability in the fold change calculation.

Entropy analysis.: For each of the genes associated with stable enhancers or human gained enhancers, we calculated the Shannon Entropy[39] for the gene expression values (read per million reads) across all differentiation time points. Lower entropy means the expression of that gene is more stage-specific across the differentiation time course.

Super-enhancer identification.: Super-enhancers (SEs) were identified in a similar manner to the method outlined in Whyte et al.[38] First, SEs were found by running HOMER findPeaks using the "-style super" option. This process first identifies traditional ChIP-seq peaks by treating the experiment and input as normal ChIP-seq experiments and then merges identified peaks found within 12.5 kb into continuous regions. The "enhancer score" for each region is defined by the number of experimental reads minus the input reads normalized for sequencing depth. All regions are then sorted by their score and plotted by their relative rank and score (0–1). Enhancer regions past which the slope of the line reaches 1 are considered SEs, and all remaining enhancers are considered as "typical enhancers."

Super-enhancer domains/HGE overlap.: Replicates for topologically associated domain (TAD) calls from our previous data[27] were merged together, sorted, and filtered for duplicates. BEDtools was used to intersect TADs with genes previously identified to be potential key regulators based on chromatin interaction frequency (genes with at least 4 degrees of interactions)[27]. The TADs were then filtered for domains that contain only one coding gene to identify TADs that may act as super-enhancer domains[41]. BEDtools was used to intersect the filtered TADs with gained and stable enhancers for multiple timepoints with a minimum overlap of 0.1 kb.

VISTA validated enhancer/HGE overlap.: VISTA enhancer coordinates were downloaded from the VISTA Enhancer Browser[42] (https://enhancer.lbl.gov/cgi-bin/imagedb3.pl?form=search&show=1&search.form=no&search.result=yes) and filtered for human sequences. Three versions of VISTA regions were created: all regions, regions filtered for positive hits, and regions filtered for positive hits with activity in heart regions. BEDtools was used to intersect the VISTA regions with the gained and stable enhancers for multiple timepoints with a minimum overlap of 0.1 kb.

GWAS and HGE analysis.: Risk variants from cardiac-associated genome-wide association studies (GWAS) were identified from the GWAS Catalog[93]. For each study of interest, the locations of all trait-associated variants and the variants in linkage-disequilibrium with them ($R > 0.9$) were scored for their enrichment in the set of human-gained enhancers using RELI[94].

ENCODE H3K27ac ChIP-seq data comparison.: H3K27ac ChIP-seq data sets performed in five ENCODE cell-lines (Gm12878, HepG2, HUVEC, K562, NHEK) were downloaded from the UCSC Genome Browser (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeBroadHistone/). We used BEDTools merge to generate a union set of ChIP-seq peaks for the ENCODE and hPSC cardiac differentiation H3K27ac datasets. The RPKM values from each H3K27ac dataset were calculated for each peak and normalized by subtracting by the mean and dividing by the standard deviation across all H3K27ac datasets. We then clustered the normalized peak values for all peaks using k-means clustering and randomly downsampled to 10,000 peaks for visualization.

Gene Ontology (GO) enrichment.: To identify biological processes associated with differentially expressed gene datasets, the online Gene Ontology tool was used (http://geneontology.org/) using default settings. To identify biological processes/molecular

functions for genes associated with differentially accessible CRE datasets, the Genomic Regions Enrichment of Annotations Tool (GREAT[46]; http://great.stanford.edu/public/html/index.php) was used with default settings for indicated CREs in ".bed" format. *P* values shown for enrichment are Bonferroni-corrected binomial *P* values.

Motif analysis.: Known motif enrichment and *de novo* motif discovery were performed using HOMER's findMotifsGenome.pl tool[85] using 200 bp sequences centered on ATAC-seq peaks for human data. For mouse data and mouse-human comparisons H3K27ac peaks were used for motif analysis. When performing de novo motif discovery, ATAC-seq peak sequences were compared to a background set of 50,000 random genomic regions matched for overall GC-content. Nucleotide frequency and motif density plots were created using HOMER's annotatePeaks.pl tool[85].

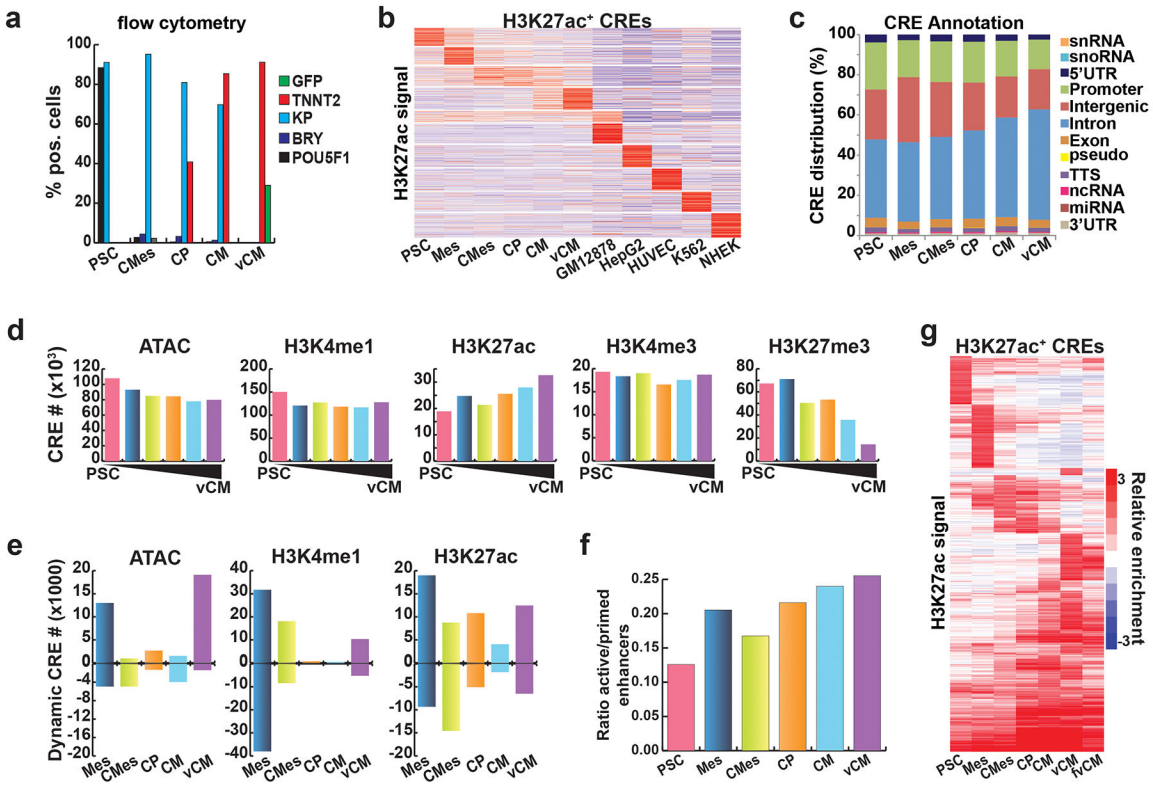
CRE genomic annotation.: HOMER's[85] annotatePeak.pl tool was used to annotate (e.g., intergenic, intronic etc.) CREs based on UCSC's genome annotation.

Clustered heatmaps.: Data was clustered using Gene Cluster[95] and then visualized as heatmaps using Java TreeView[96]. The rows in the heatmap for Fig. 1b represent a set of loci as indicated by the numbers on the right of the heatmap.

Statistics and reproducibility.: Statistical analyses were performed using GraphPad Prism, Python and R. Specific tests are indicated in the figure legends. No statistical methods were used to predetermine sample size. Experiments did not employ randomization nor investigator blinding.

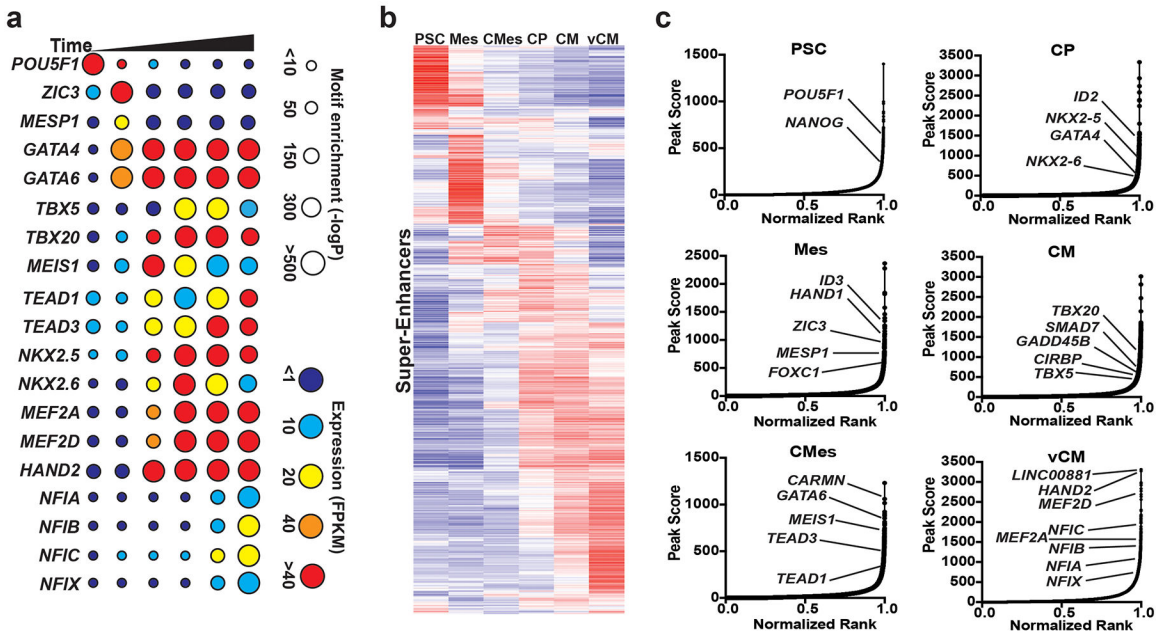
Data and code availability.: Sequencing data is available from the GEO database (snATAC: GSE192500, reviewer token gxwhcgkyvhsxpex; bulk sequencing data: GSE192365, reviewer token qbirqwesxxyhfuz). A list of all the used sequencing datasets (new and public) and their accession numbers are available in Supplemental Table 13. Bulk sequencing data can be viewed using the UCSC genome browser (https://genome.ucsc.edu/s/Cbenner/Chi-211218-ReviewerTracks). HOMER analysis package is available online (homer.ucsd.edu).

Extended Data



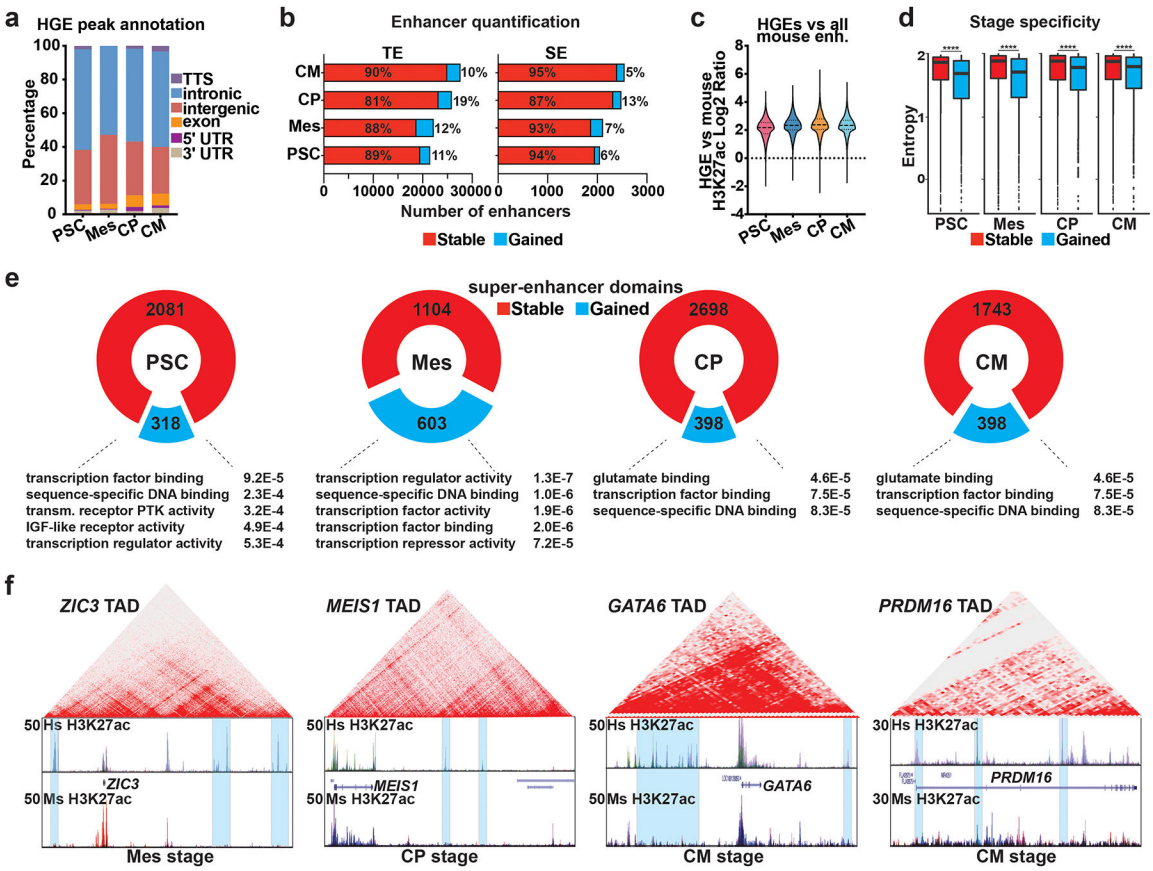
Extended Data Fig. 1: Gene regulatory networks dynamically remodel during human cardiomyocyte differentiation.

a, Flow cytometry analysis shows the percentage of cell types generated at each cardiac developmental stage as assessed by the indicated markers. **b**, Heatmap shows cell-type cardiac enhancers through comparison with ENCODE-derived cell line enhancer data. **c**, CREs are enriched in intergenic and intronic regions during cardiomyocyte differentiation. **d**, Number of CREs for different epigenomic assays are shown across the six stages of human vCM differentiation. **e**, Bar plot shows the number of CREs that are gained or lost compared to the preceding stage during human ventricular cardiomyocyte differentiation. **f**, Analyzing the ratio of active enhancers ($H3K27ac^{+}/H3K4me1^{+}$) to all enhancers ($H3K4me1^{+}$) reveals increased enhancer utilization during human vCM differentiation. **g**, Heatmap displays enhancer activity as assessed by H3K27ac signal during *in vitro* vCM differentiation and in purified human fetal CMs. PSC, human pluripotent stem cells; Mes, mesoderm; CMes, cardiac mesoderm; CP, cardiac progenitor; CM, cardiomyocyte; vCM, ventricular cardiomyocyte; fvCM, fetal ventricular cardiomyocyte; CRE, cis-regulatory region; pos., positive; snRNA, small nuclear RNA; snoRNA, small nucleolar RNA; UTR, untranslated region; pseudo, pseudogene; TTS, transcription termination site; ncRNA; non-coding RNA; miRNA, microRNA; HUVEC, human umbilical vein endothelial cells; NHEK, normal human epidermal keratinocytes; fvCM, fetal ventricular cardiomyocyte.



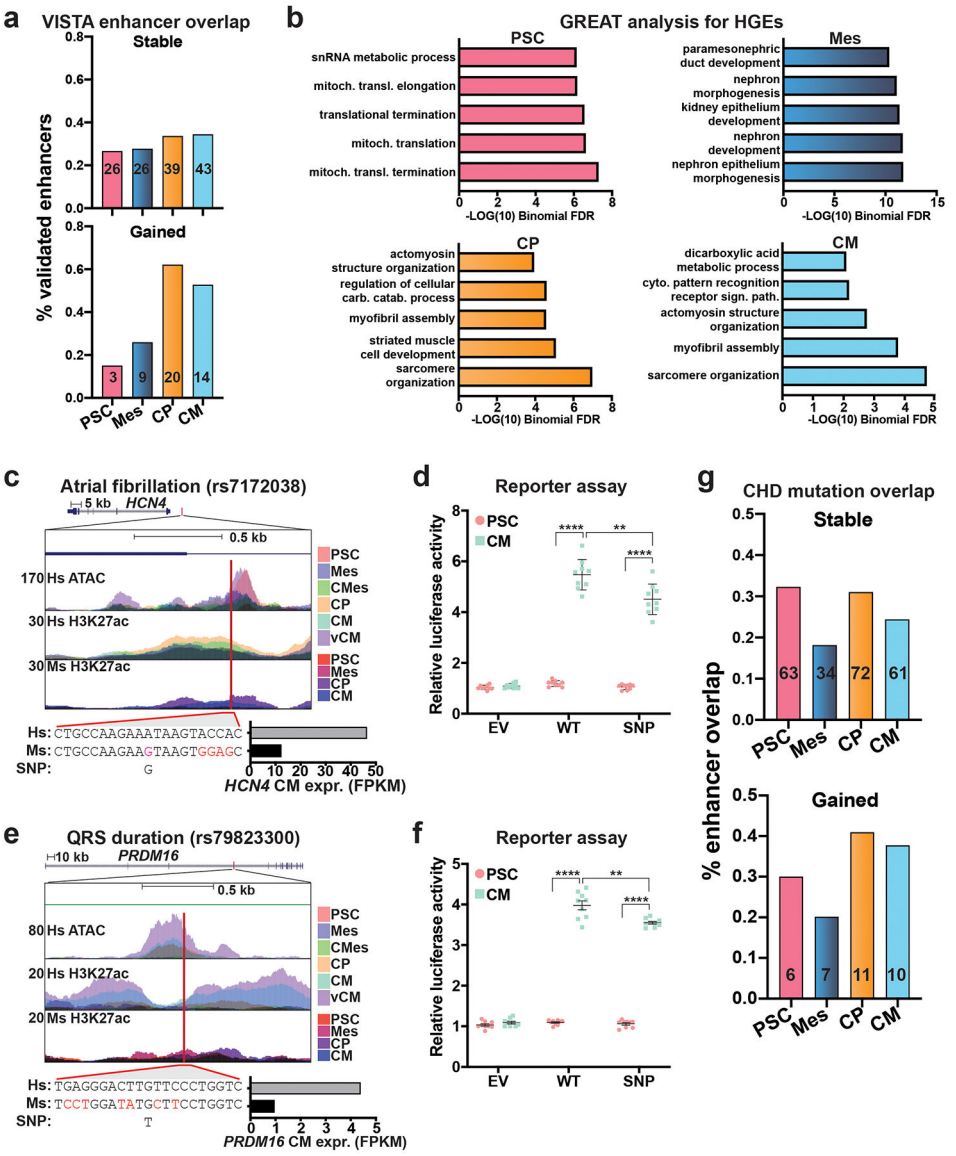
Extended Data Fig. 2: Integrative transcriptional and chromatin accessibility analysis identifies putative key cardiac regulators.

a, Dot plot shows expression (circle color) and transcription factor motif enrichment (circle size) of known and potential novel regulators of human ventricle cardiomyocyte differentiation. **b**, Heatmap shows differentially regulated super-enhancers during cardiomyocyte differentiation. **c**, Distribution of H3K27ac signal for each of the indicated stages shows super-enhancer loci with associated genes of interest based on proximity. PSC, pluripotent stem cells; Mes, mesoderm; Cmes, cardiac mesoderm; CP, cardiac progenitor; CM, cardiomyocyte; vCM, ventricular cardiomyocyte; FPKM, fragments per kilobase of transcript per million mapped reads.



Extended Data Fig. 3: Human gained enhancers consist of traditional and super-enhancers during cardiomyocyte differentiation.

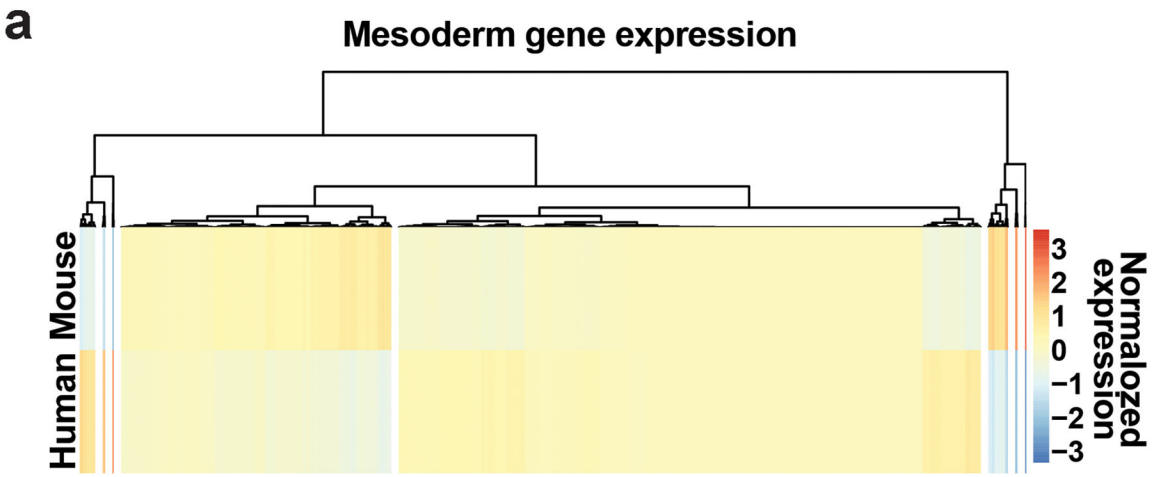
a, HGEs are enriched in intergenic and intronic regions during cardiomyocyte differentiation. **b**, HGEs consist of both traditional (TE) and super-enhancers (SE). **c**, Violin plots reveal comparable activity of HGEs compared to all other mouse stages combined. Dashed line in violin plots indicates the mean and dotted lines indicate quartiles. **d**, Shannon entropy calculations show stage-specific expression of genes near gained or stable enhancers. Lower entropy indicates more stage-specific gene expression. Horizontal line in boxplots represents the median, the box indicates the interquartile range and the dots represent outliers. Data was compared using Wilcoxon rank sum test (****p value < 0.0001). **e**, Stable and gained super-enhancer domains are primarily associated with transcription factors based on GO-term analysis. **f**, Representative super-enhancer TADs with HGEs (blue boxes) are shown at various cardiomyocyte developmental stages. HGE, human-gained enhancer; PSC, pluripotent stem cells; Mes, mesoderm; CP, cardiac progenitor; CM, cardiomyocyte; UTR, untranslated region; pseudo, pseudogene; TTS, transcription termination site; enh., enhancers; TAD, topologically associated domain; hs, human; ms, mouse.



Extended Data Fig. 4: HGEs are associated with cardiac developmental specific activity and function.

a, Bar graph shows the percentage and number of VISTA validated enhancers driving cardiac LacZ reporter activity *in vivo* that overlap with stable and gained enhancers at different cardiomyocyte developmental stages. **b**, GREAT analysis reveals biological processes associated with HGEs at the examined cardiac developmental stages. **c**-**f**, ATAC-seq/ChIP-seq profiles show that SNPs associated with (**c**) atrial fibrillation and (**e**) QRS duration overlap with HGEs, and luciferase reporter assays (**d, f**) confirm that the SNPs affect enhancer activity. Sequence comparison surrounding the SNP is shown on the bottom left of (**c**) and (**e**) with nucleotides that are not conserved between mouse and human indicated in red. Gene expression in human (grey) and mouse (black) cardiomyocytes is shown on the bottom right of (**c**) and (**e**). n = 9 for the luciferase reporter assays. Data was compared using a two-tailed Student's t-test. *p < 0.05; **p < 0.01; ***p < 0.001; ****p < 0.0001. Data represented as the mean ± SEM with individual data points. **g**,

Bar graph shows the percentage and absolute number of noncoding mutations found in CHD patients that overlap with stable and gained enhancers at different cardiomyocyte developmental stages. HGE, human-gained enhancer; PSC, pluripotent stem cells; Mes, mesoderm; CMes, cardiac mesoderm; CP, cardiac progenitor; CM, cardiomyocyte; vCM, ventricular cardiomyocyte; snRNA, small nuclear RNA; mitoch., mitochondrial; transl., translation; carb., carbohydrate; catab., catabolism; cyto., cytoplasmic; FDR, false discovery rate; Hs, human; Ms, mouse; expr, expression; FKPM, fragments per kilobase per million reads; EV, empty vector; WT, wild-type; SNP; single nucleotide polymorphism; CHD, congenital heart disease.



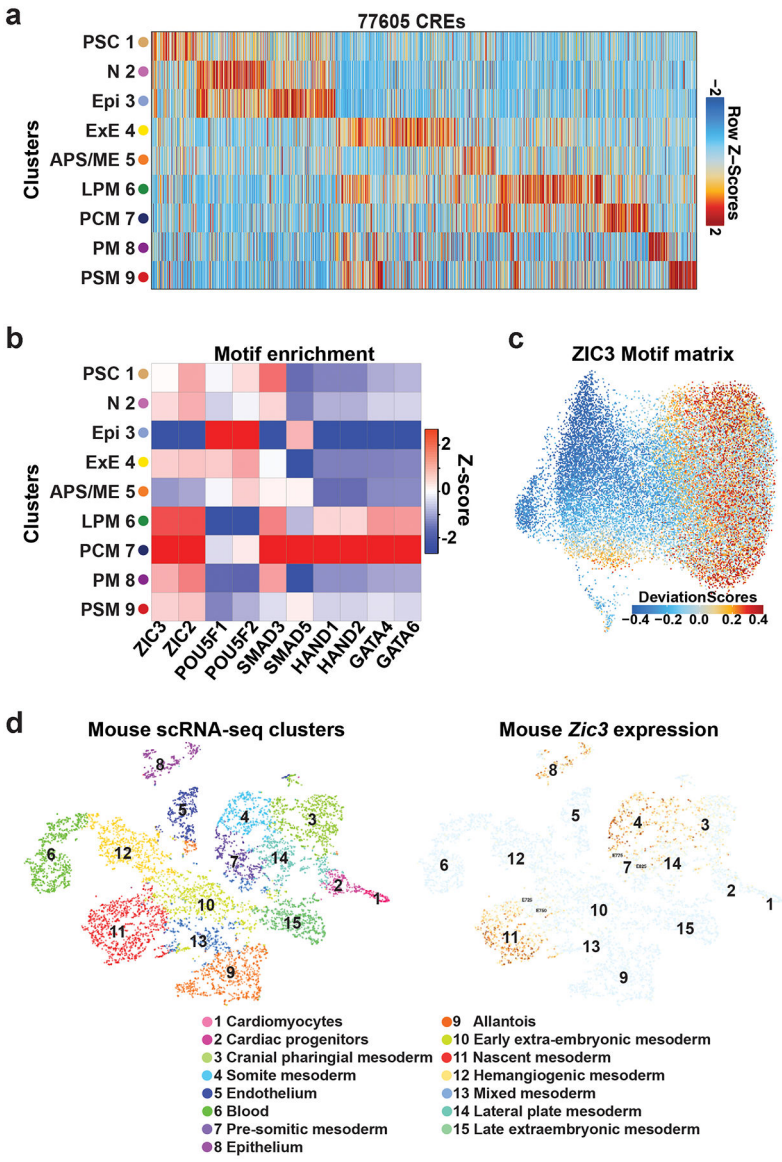
b

Top 10 motifs Mes stage

Human			Mouse		
Motif	Match	p-val.	Motif	Match	p-val.
	Gata (Zf)	10^{-473}		KLF (Zf)	10^{-178}
	ZIC (Zf)	10^{-361}		ETS	10^{-147}
	T-box	10^{-267}		HMG	10^{-143}
	Homeobox	10^{-243}		Gata (Zf)	10^{-125}
	Forkhead	10^{-225}		Homeobox	10^{-123}
	bHLH	10^{-167}		Forkhead	10^{-82}
	ETS	10^{-156}		T-box	10^{-76}
	HMG	10^{-143}		THAP	10^{-75}
	GLIS3(Zf)	10^{-140}		STAT	10^{-69}
	TEAD	10^{-136}		HTH	10^{-66}

Extended Data Fig. 5: Gene expression and transcription factor motif analyses reveal distinct differences between human and mouse mesoderm.

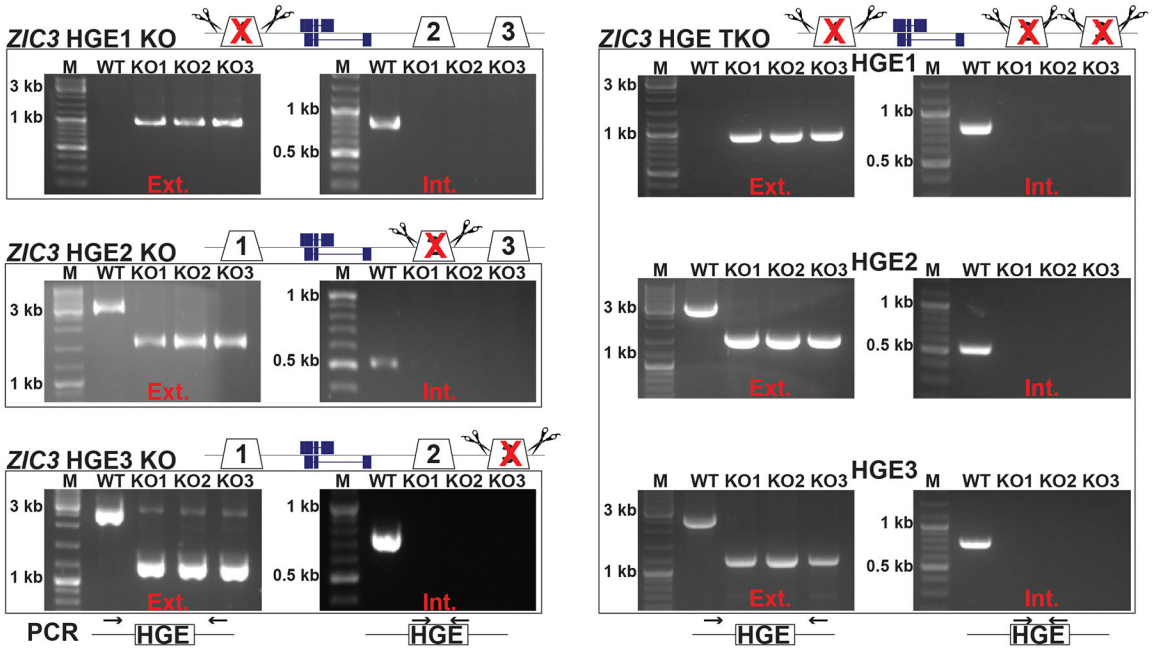
a, Clustered heatmap of human and mouse mesoderm gene expression profiles shows both conserved and divergent gene expression clusters. **b**, Table shows top transcription factor motifs for human and mouse mesoderm active enhancers. Mes, mesoderm; val., value.



Extended Data Fig. 6: CREs within mesoderm cell subpopulations are associated with distinct transcription factor binding activity.

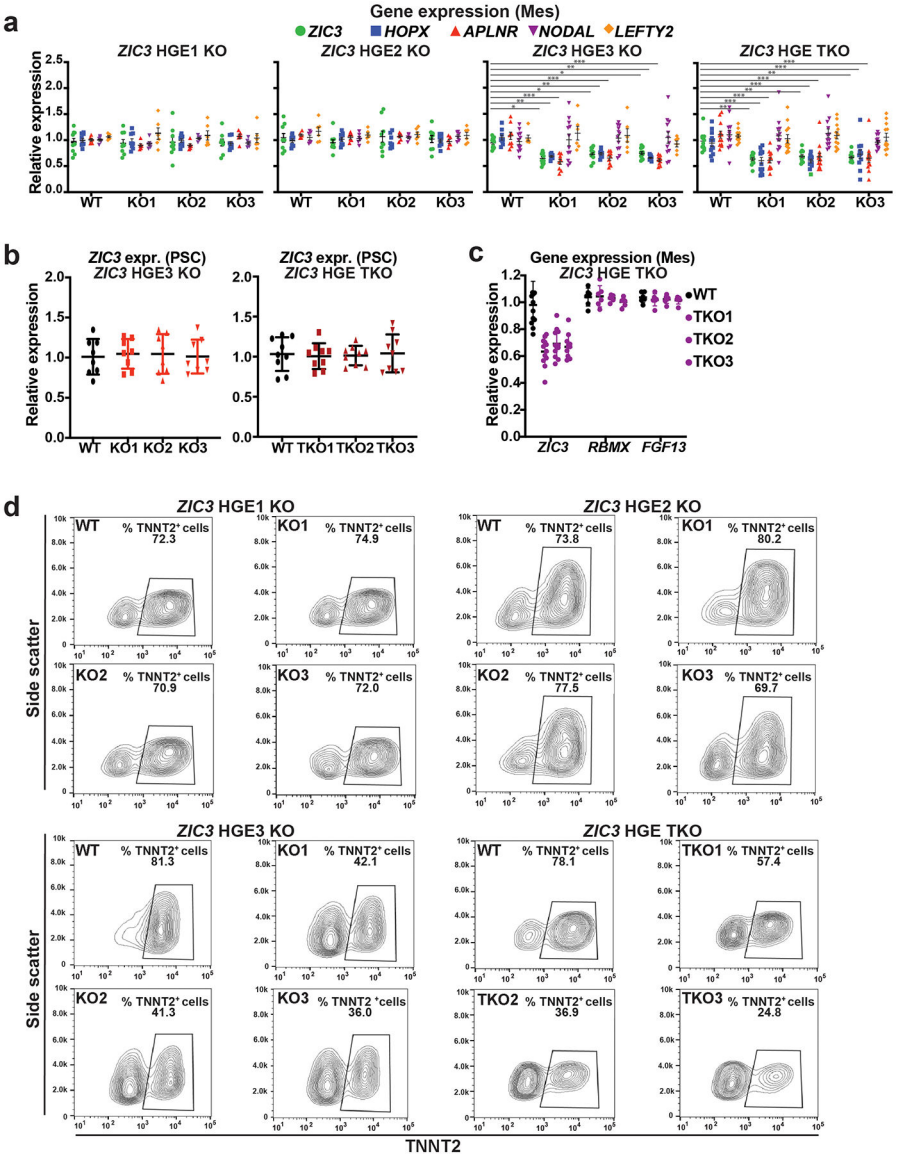
a, Clustered heatmap shows differentially (FDR < 0.05) accessible CREs between identified snATAC mesoderm cell subpopulations. **b**, Heatmap of transcription factor binding motif analysis of cluster-specific CREs from human mesodermal snATAC-seq data reveals that *ZIC3* along with other cardiac developmental transcription factor motifs are enriched in the pre-cardiac mesoderm subpopulation. **c**, ZIC3 motif enrichment is projected onto the UMAP of human PSC-derived mesoderm snATAC-seq data. **d**, (Left) Mouse scRNA-seq data from E7.25-E8.5 *Mesp1-Cre; Rosa26-tdTomato* embryos are displayed by tSNE (t-distributed stochastic neighbor embedding) plots for all developmental stages combined. (Right) *Zic3*

expression is projected on to the tSNE plot. PSC, pluripotent stem cell; N, neural; Epi, epiblast; ExE, extra-embryonic; APS/ME, anterior primitive streak/mesendoderm; LPM, lateral-plate mesoderm; PCM, pre-cardiac mesoderm; PM; paraxial mesoderm, PSM, pre-somitic mesoderm; CRE, cis-regulatory region; adj., adjusted.



Extended Data Fig. 7: Generation of clonal ZIC3 HGE deletion lines.

Human PSC *ZIC3* HGE1–3 individual knockout (KO) and triple KO (TKO) lines were generated through CRISPR/Cas9-mediated deletion of *ZIC3* HGE1–3 enhancer regions (9.2 kb, 1.8 kb, and 1.5 kb, respectively). PCR genotyping using external (ext., i.e., outside deleted regions) and internal (int., i.e., inside deleted regions) primers confirmed the individual human PSC *ZIC3* HGE KOs (left) and the human PSC *ZIC3* HGE TKO (right). HGE TKO was confirmed by individual HGE1–3 genotyping. Wild-type (WT) cells were used as control. Note that the WT HGE1 band is 9.2 kb and thus difficult to PCR amplify. M, marker, ext., external; int., internal.



Extended Data Fig. 8: HGE3 KO and HGE TKO human pluripotent stem cell knockout lines display reduced cardiomyocyte differentiation and mesoderm-specific *ZIC3* expression defects.

a, All individual PCR data for *ZIC3* HGE knockout (KO) and HGE triple knockout (TKO) lines (n = 6–12) are shown with statistical significance indicated. *TBP* gene expression was used for gene expression normalization. **b**, qPCR analysis of *ZIC3* HGE3 KO and HGE TKO human PSC lines reveals no significant changes in *ZIC3* expression at the PSC stage compared to WT cells (n = 9). **c**, qPCR analysis reveals significant changes in the expression of *ZIC3* but not neighboring genes, *RBMX* and *FGF13*, in HGE TKO human PSC lines when compared to WT human PSC lines at the mesoderm stage (n = 6). **d**, Representative examples of flow cytometry analyses for TNNT2 staining (x-axis) show that *ZIC3* HGE3 and TKO lines (but not HGE1 nor HGE2) display reduced cardiac differentiation. Data is represented as the mean ± SEM with individual data points. Data was analyzed using a two-tailed Student's t-test. *p < 0.05; **p < 0.01; ***p < 0.001; ****p < 0.0001. HGE, human-gained enhancer; expr., expression; PSC, pluripotent stem cells; Mes, mesoderm.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

We thank M. Swim for help with analysis of flow cytometry data and E. Ines for technical assistance. We thank K. Jepsen and the UCSD Institute for Genomic Medicine for help with sequencing and the UCSD Human Embryonic Stem Cell Core Facility for help with cell sorting. This work was supported in part by grants from the NIH to B.R., S.M.E., N.C.C.; CIRM to B.R., N.C.C.; AHA to J.B.

References

1. Long HK, Prescott SL & Wysocka J Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. Cell 167, 1170–1187 (2016). 10.1016/j.cell.2016.09.018 [PubMed: 27863239]
2. Ren B & Yue F Transcriptional Enhancers: Bridging the Genome and Phenome. Cold Spring Harb Symp Quant Biol 80, 17–26 (2015). 10.1101/sqb.2015.80.027219 [PubMed: 26582789]
3. Kathiriya IS, Nora EP & Bruneau BG Investigating the transcriptional control of cardiovascular development. Circ Res 116, 700–714 (2015). 10.1161/circresaha.116.302832 [PubMed: 25677518]
4. Bruneau BG The developmental genetics of congenital heart disease. Nature 451, 943–948 (2008). 10.1038/nature06801 [PubMed: 18288184]
5. Fahed AC, Gelb BD, Seidman JG & Seidman CE Genetics of congenital heart disease: the glass half empty. Circ Res 112, 707–720 (2013). 10.1161/circresaha.112.300853 [PubMed: 23410880]
6. Homsy J et al. De novo mutations in congenital heart disease with neurodevelopmental and other congenital anomalies. Science 350, 1262–1266 (2015). 10.1126/science.aac9396 [PubMed: 26785492]
7. Richter F et al. Genomic analyses implicate noncoding de novo variants in congenital heart disease. Nat Genet 52, 769–777 (2020). 10.1038/s41588-020-0652-z [PubMed: 32601476]
8. Arvanitis M et al. Genome-wide association and multi-omic analyses reveal ACTN2 as a gene linked to heart failure. Nat Commun 11, 1122 (2020). 10.1038/s41467-020-14843-7 [PubMed: 32111823]
9. Hocker JD et al. Cardiac cell type-specific gene regulatory programs and disease risk association. Sci Adv 7 (2021). 10.1126/sciadv.abf1444
10. Lahm H et al. Congenital heart disease risk loci identified by genome-wide association study in European patients. J Clin Invest 131 (2021). 10.1172/jci141837
11. Gilsbach R et al. Dynamic DNA methylation orchestrates cardiomyocyte development, maturation and disease. Nat Commun 5, 5288 (2014). 10.1038/ncomms6288 [PubMed: 25335909]
12. Gilsbach R et al. Distinct epigenetic programs regulate cardiac myocyte development and disease in the human heart in vivo. Nat Commun 9, 391 (2018). 10.1038/s41467-017-02762-z [PubMed: 29374152]
13. Nothjunge S et al. DNA methylation signatures follow preformed chromatin compartments in cardiac myocytes. Nat Commun 8, 1667 (2017). 10.1038/s41467-017-01724-9 [PubMed: 29162810]
14. Paige SL et al. A temporal chromatin signature in human embryonic stem cells identifies regulators of cardiac development. Cell 151, 221–232 (2012). [PubMed: 22981225]
15. Wamstad JA et al. Dynamic and coordinated epigenetic regulation of developmental transitions in the cardiac lineage. Cell 151, 206–220 (2012). 10.1016/j.cell.2012.07.035 [PubMed: 22981692]
16. Zhao MT et al. Cell Type-Specific Chromatin Signatures Underline Regulatory DNA Elements in Human Induced Pluripotent Stem Cells and Somatic Cells. Circ Res 121, 1237–1250 (2017). 10.1161/circresaha.117.311367 [PubMed: 29030344]
17. Doevendans PA, Daemen MJ, de Muinck ED & Smits JF Cardiovascular phenotyping in mice. Cardiovasc Res 39, 34–49 (1998). 10.1016/s0008-6363(98)00073-x [PubMed: 9764188]

18. Krishnan A et al. A detailed comparison of mouse and human cardiac development. Pediatr Res 76, 500–507 (2014). 10.1038/pr.2014.128 [PubMed: 25167202]
19. Reilly SK & Noonan JP Evolution of Gene Regulation in Humans. Annual review of genomics and human genetics 17, 45–67 (2016).
20. Carroll SB Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell 134, 25–36 (2008). 10.1016/j.cell.2008.06.030 [PubMed: 18614008]
21. Carroll SB et al. Pattern formation and eyespot determination in butterfly wings. Science 265, 109–114 (1994). 10.1126/science.7912449 [PubMed: 7912449]
22. Cotney J et al. The evolution of lineage-specific regulatory activities in the human embryonic limb. Cell 154, 185–196 (2013). [PubMed: 23827682]
23. Reilly SK et al. Evolutionary genomics. Evolutionary changes in promoter and enhancer activity during human corticogenesis. Science (New York, N.Y.) 347, 1155–1159 (2015). [PubMed: 25745175]
24. de la Torre-Ubieta L et al. The Dynamic Landscape of Open Chromatin during Human Cortical Neurogenesis. Cell 172, 289–304.e218 (2018). 10.1016/j.cell.2017.12.014 [PubMed: 29307494]
25. Van Vliet P, Wu SM, Zaffran S & Pucéat M Early cardiac development: a view from stem cells to embryos. Cardiovasc Res 96, 352–362 (2012). 10.1093/cvr/cvs270 [PubMed: 22893679]
26. Veevers J et al. Cell-Surface Marker Signature for Enrichment of Ventricular Cardiomyocytes Derived from Human Embryonic Stem Cells. Stem Cell Reports 11, 828–841 (2018). 10.1016/j.stemcr.2018.07.007 [PubMed: 30122443]
27. Zhang Y et al. Transcriptionally active HERV-H retrotransposons demarcate topologically associating domains in human pluripotent stem cells. Nat Genet 51, 1380–1388 (2019). 10.1038/s41588-019-0479-7 [PubMed: 31427791]
28. Zhu J et al. Genome-wide chromatin state transitions associated with developmental and environmental cues. Cell 152, 642–654 (2013). 10.1016/j.cell.2012.12.033 [PubMed: 23333102]
29. Stergachis AB et al. Developmental fate and cellular maturity encoded in human regulatory DNA landscapes. Cell 154, 888–903 (2013). 10.1016/j.cell.2013.07.020 [PubMed: 23953118]
30. Heintzman ND et al. Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature 459, 108–112 (2009). 10.1038/nature07829 [PubMed: 19295514]
31. Sahara M et al. Population and Single-Cell Analysis of Human Cardiogenesis Reveals Unique LGR5 Ventricular Progenitors in Embryonic Outflow Tract. Dev Cell 48, 475–490.e477 (2019). 10.1016/j.devcel.2019.01.005 [PubMed: 30713072]
32. Asp M et al. A Spatiotemporal Organ-Wide Gene Expression and Cell Atlas of the Developing Human Heart. Cell 179, 1647–1660.e1619 (2019). 10.1016/j.cell.2019.11.025 [PubMed: 31835037]
33. Mononen MM, Leung CY, Xu J & Chien KR Trajectory mapping of human embryonic stem cell cardiogenesis reveals lineage branch points and an ISL1 progenitor-derived cardiac fibroblast lineage. Stem Cells 38, 1267–1278 (2020). 10.1002/stem.3236 [PubMed: 32497389]
34. Grunert M, Dorn C & Rickert-Sperling S Cardiac Transcription Factors and Regulatory Networks.
35. Bertero A et al. Dynamics of genome reorganization during human cardiogenesis reveal an RBM20-dependent splicing factory. Nat Commun 10, 1538 (2019). 10.1038/s41467-019-09483-5 [PubMed: 30948719]
36. Liu Q et al. Genome-Wide Temporal Profiling of Transcriptome and Open Chromatin of Early Cardiomyocyte Differentiation Derived From hiPSCs and hESCs. Circulation research 121, 376–391 (2017). [PubMed: 28663367]
37. Hnisz D et al. Super-enhancers in the control of cell identity and disease. Cell 155, 934–947 (2013). 10.1016/j.cell.2013.09.053 [PubMed: 24119843]
38. Whyte WA et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell 153, 307–319 (2013). 10.1016/j.cell.2013.03.035 [PubMed: 23582322]
39. Schug J et al. Promoter features related to tissue specificity as measured by Shannon entropy. Genome Biol 6, R33 (2005). 10.1186/gb-2005-6-4-r33 [PubMed: 15833120]
40. Uebbing S et al. Massively parallel discovery of human-specific substitutions that alter enhancer activity. Proc Natl Acad Sci U S A 118 (2021). 10.1073/pnas.2007049118

41. Dowen JM et al. Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. Cell 159, 374–387 (2014). 10.1016/j.cell.2014.09.030 [PubMed: 25303531]
42. Visel A, Minovitsky S, Dubchak I & Pennacchio LA VISTA Enhancer Browser--a database of tissue-specific human enhancers. Nucleic Acids Res 35, D88–92 (2007). 10.1093/nar/gkl822 [PubMed: 17130149]
43. Aronsen JM, Swift F & Sejersted OM Cardiac sodium transport and excitation-contraction coupling. J Mol Cell Cardiol 61, 11–19 (2013). 10.1016/j.yjmcc.2013.06.003 [PubMed: 23774049]
44. George AL Jr. et al. Assignment of the human heart tetrodotoxin-resistant voltage-gated Na+ channel alpha-subunit gene (SCN5A) to band 3p21. Cytogenet Cell Genet 68, 67–70 (1995). 10.1159/000133892 [PubMed: 7956363]
45. Berthold J, Schenkova K & Rivero F Rho GTPases of the RhoBTB subfamily and tumorigenesis. Acta Pharmacol Sin 29, 285–295 (2008). 10.1111/j.1745-7254.2008.00773.x [PubMed: 18298893]
46. McLean CY et al. GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol 28, 495–501 (2010). 10.1038/nbt.1630 [PubMed: 20436461]
47. Kerr KF et al. Genome-wide association study of heart rate and its variability in Hispanic/Latino cohorts. Heart Rhythm 14, 1675–1684 (2017). 10.1016/j.hrthm.2017.06.018 [PubMed: 28610988]
48. Creutz CE et al. The copines, a novel class of C2 domain-containing, calcium-dependent, phospholipid-binding proteins conserved from Paramecium to humans. J Biol Chem 273, 1393–1402 (1998). 10.1074/jbc.273.3.1393 [PubMed: 9430674]
49. DiFrancesco D HCN4, Sinus Bradycardia and Atrial Fibrillation. Arrhythm Electrophysiol Rev 4, 9–13 (2015). 10.15420/aer.2015.4.1.9 [PubMed: 26835093]
50. Hennis K et al. Speeding Up the Heart? Traditional and New Perspectives on HCN4 Function. Front Physiol 12, 669029 (2021). 10.3389/fphys.2021.669029 [PubMed: 34122140]
51. Cibi DM et al. Prdm16 Deficiency Leads to Age-Dependent Cardiac Hypertrophy, Adverse Remodeling, Mitochondrial Dysfunction, and Heart Failure. Cell Rep 33, 108288 (2020). 10.1016/j.celrep.2020.108288 [PubMed: 33086060]
52. Nam JM, Lim JE, Ha TW, Oh B & Kang JO Cardiac-specific inactivation of Prdm16 effects cardiac conduction abnormalities and cardiomyopathy-associated phenotypes. Am J Physiol Heart Circ Physiol 318, H764–h777 (2020). 10.1152/ajpheart.00647.2019 [PubMed: 32083975]
53. Jiang Z et al. Zic3 is required in the extra-cardiac perinodal region of the lateral plate mesoderm for left-right patterning and heart development. Hum Mol Genet 22, 879–889 (2013). 10.1093/hmg/dds494 [PubMed: 23184148]
54. Sutherland MJ, Wang S, Quinn ME, Haaning A & Ware SM Zic3 is required in the migrating primitive streak for node morphogenesis and left-right patterning. Hum Mol Genet 22, 1913–1923 (2013). 10.1093/hmg/ddt001 [PubMed: 23303524]
55. Ware SM et al. Identification and functional analysis of ZIC3 mutations in heterotaxy and related congenital heart defects. Am J Hum Genet 74, 93–105 (2004). 10.1086/380998 [PubMed: 14681828]
56. Zhu L et al. Identification of a novel role of ZIC3 in regulating cardiac development. Hum Mol Genet 16, 1649–1660 (2007). 10.1093/hmg/ddm106 [PubMed: 17468179]
57. Preissl S et al. Single-nucleus analysis of accessible chromatin in developing mouse forebrain reveals cell-type-specific transcriptional regulation. Nat Neurosci 21, 432–439 (2018). 10.1038/s41593-018-0079-3 [PubMed: 29434377]
58. Cusanovich DA et al. Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. Science 348, 910–914 (2015). 10.1126/science.aab1601 [PubMed: 25953818]
59. Granja JM et al. ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. Nat Genet 53, 403–411 (2021). 10.1038/s41588-021-00790-6 [PubMed: 33633365]
60. Zhang Y et al. Model-based Analysis of ChIP-Seq (MACS). Genome Biology 9, R137 (2008). 10.1186/gb-2008-9-9-r137 [PubMed: 18798982]
61. Zhang Q et al. Unveiling Complexity and Multipotentiality of Early Heart Fields. Circ Res 129, 474–487 (2021). 10.1161/circresaha.121.318943 [PubMed: 34162224]

62. Kaestner KH The FoxA factors in organogenesis and differentiation. Curr Opin Genet Dev 20, 527–532 (2010). 10.1016/j.gde.2010.06.005 [PubMed: 20591647]
63. Bardot E et al. Foxa2 identifies a cardiac progenitor population with ventricular differentiation potential. Nat Commun 8, 14428 (2017). 10.1038/ncomms14428 [PubMed: 28195173]
64. Tsankov AM et al. Transcription factor binding dynamics during human ES cell differentiation. Nature 518, 344–349 (2015). 10.1038/nature14233 [PubMed: 25693565]
65. Bedard JE, Haaning AM & Ware SM Identification of a novel ZIC3 isoform and mutation screening in patients with heterotaxy and congenital heart disease. PloS one 6, e23755 (2011). [PubMed: 21858219]
66. Bellchambers HM & Ware SM Loss of Zic3 impairs planar cell polarity leading to abnormal left-right signaling, heart defects and neural tube defects. Hum Mol Genet (2021). 10.1093/hmg/ddab195
67. Cast AE, Gao C, Amack JD & Ware SM An essential and highly conserved role for Zic3 in left-right patterning, gastrulation and convergent extension morphogenesis. Dev Biol 364, 22–31 (2012). 10.1016/j.ydbio.2012.01.011 [PubMed: 22285814]
68. Kitaguchi T, Nagai T, Nakata K, Aruga J & Mikoshiba K Zic3 is involved in the left-right specification of the Xenopus embryo. Development 127, 4787–4795 (2000). [PubMed: 11044394]
69. Xu J et al. Genome-wide CRISPR screen identifies ZIC2 as an essential gene that controls the cell fate of early mesodermal precursors to human heart progenitors. Stem Cells 38, 741–755 (2020). 10.1002/stem.3168 [PubMed: 32129551]
70. Yang R et al. Amnion signals are essential for mesoderm formation in primates. Nat Commun 12, 5126 (2021). 10.1038/s41467-021-25186-2 [PubMed: 34446705]
71. Gebbia M et al. X-linked situs abnormalities result from mutations in ZIC3. Nat Genet 17, 305–308 (1997). 10.1038/ng1197-305 [PubMed: 9354794]
72. Belmont JW, Mohapatra B, Towbin JA & Ware SM Molecular genetics of heterotaxy syndromes. Curr Opin Cardiol 19, 216–220 (2004). 10.1097/00001573-200405000-00005 [PubMed: 15096953]
73. Chhin B et al. Elucidation of penetrance variability of a ZIC3 mutation in a family with complex heart defects and functional analysis of ZIC3 mutations in the first zinc finger domain. Hum Mutat 28, 563–570 (2007). 10.1002/humu.20480 [PubMed: 17295247]
74. Cowan J, Tariq M & Ware SM Genetic and functional analyses of ZIC3 variants in congenital heart disease. Hum Mutat 35, 66–75 (2014). 10.1002/humu.22457 [PubMed: 24123890]
75. Drakhlis L et al. Human heart-forming organoids recapitulate early heart and foregut development. Nat Biotechnol 39, 737–746 (2021). 10.1038/s41587-021-00815-9 [PubMed: 33558697]
76. Hofbauer P et al. Cardioids reveal self-organizing principles of human cardiogenesis. Cell 184, 3299–3317.e3222 (2021). 10.1016/j.cell.2021.04.034 [PubMed: 34019794]
77. Lewis-Israeli YR et al. Self-assembling human heart organoids for the modeling of cardiac development and congenital heart disease. Nat Commun 12, 5142 (2021). 10.1038/s41467-021-25329-5 [PubMed: 34446706]
78. Richards DJ et al. Human cardiac organoids for the modelling of myocardial infarction and drug cardiotoxicity. Nat Biomed Eng 4, 446–462 (2020). 10.1038/s41551-020-0539-4 [PubMed: 32284552]
79. Rossi G et al. Capturing Cardiogenesis in Gastruloids. Cell Stem Cell 28, 230–240.e236 (2021). 10.1016/j.stem.2020.10.013 [PubMed: 33176168]
80. Lian X et al. Directed cardiomyocyte differentiation from human pluripotent stem cells by modulating Wnt/β-catenin signaling under fully defined conditions. Nat Protoc 8, 162–175 (2013). 10.1038/nprot.2012.150 [PubMed: 23257984]
81. Concordet JP & Haeussler M CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. Nucleic Acids Res 46, W242–w245 (2018). 10.1093/nar/gky354 [PubMed: 29762716]
82. Buenrostro JD, Giresi PG, Zaba LC, Chang HY & Greenleaf WJ Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods 10, 1213–1218 (2013). 10.1038/nmeth.2688 [PubMed: 24097267]

83. Preissl S et al. Single-nucleus analysis of accessible chromatin in developing mouse forebrain reveals cell-type-specific transcriptional regulation. Nat Neurosci 21, 432–439 (2018). 10.1038/s41593-018-0079-3 [PubMed: 29434377]
84. Corces MR et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat Methods 14, 959–962 (2017). 10.1038/nmeth.4396 [PubMed: 28846090]
85. Heinz S et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell 38, 576–589 (2010). 10.1016/j.molcel.2010.05.004 [PubMed: 20513432]
86. Robinson JT et al. Juicebox.js Provides a Cloud-Based Visualization System for Hi-C Data. Cell Syst 6, 256–258.e251 (2018). 10.1016/j.cels.2018.01.001 [PubMed: 29428417]
87. Deacon DC et al. Combinatorial interactions of genetic variants in human cardiomyopathy. Nat Biomed Eng 3, 147–157 (2019). 10.1038/s41551-019-0348-9 [PubMed: 30923642]
88. Langmead B & Salzberg SL Fast gapped-read alignment with Bowtie 2. Nat Methods 9, 357–359 (2012). 10.1038/nmeth.1923 [PubMed: 22388286]
89. Dobin A et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21 (2013). 10.1093/bioinformatics/bts635 [PubMed: 23104886]
90. Love MI, Huber W & Anders S Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15, 550 (2014). 10.1186/s13059-014-0550-8 [PubMed: 25516281]
91. Hao Y et al. Integrated analysis of multimodal single-cell data. Cell 184, 3573–3587.e3529 (2021). 10.1016/j.cell.2021.04.048 [PubMed: 34062119]
92. Minn KT et al. High-resolution transcriptional and morphogenetic profiling of cells from micropatterned human ESC gastruloid cultures. Elife 9 (2020). 10.7554/eLife.59445
93. Buniello A et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res 47, D1005–d1012 (2019). 10.1093/nar/gky1120 [PubMed: 30445434]
94. Harley JB et al. Transcription factors operate across disease loci, with EBNA2 implicated in autoimmunity. Nat Genet 50, 699–707 (2018). 10.1038/s41588-018-0102-3 [PubMed: 29662164]
95. de Hoon MJ, Imoto S, Nolan J & Miyano S Open source clustering software. Bioinformatics 20, 1453–1454 (2004). 10.1093/bioinformatics/bth078 [PubMed: 14871861]
96. Saldanha AJ Java Treeview--extensible visualization of microarray data. Bioinformatics 20, 3246–3248 (2004). 10.1093/bioinformatics/bth349 [PubMed: 15180930]

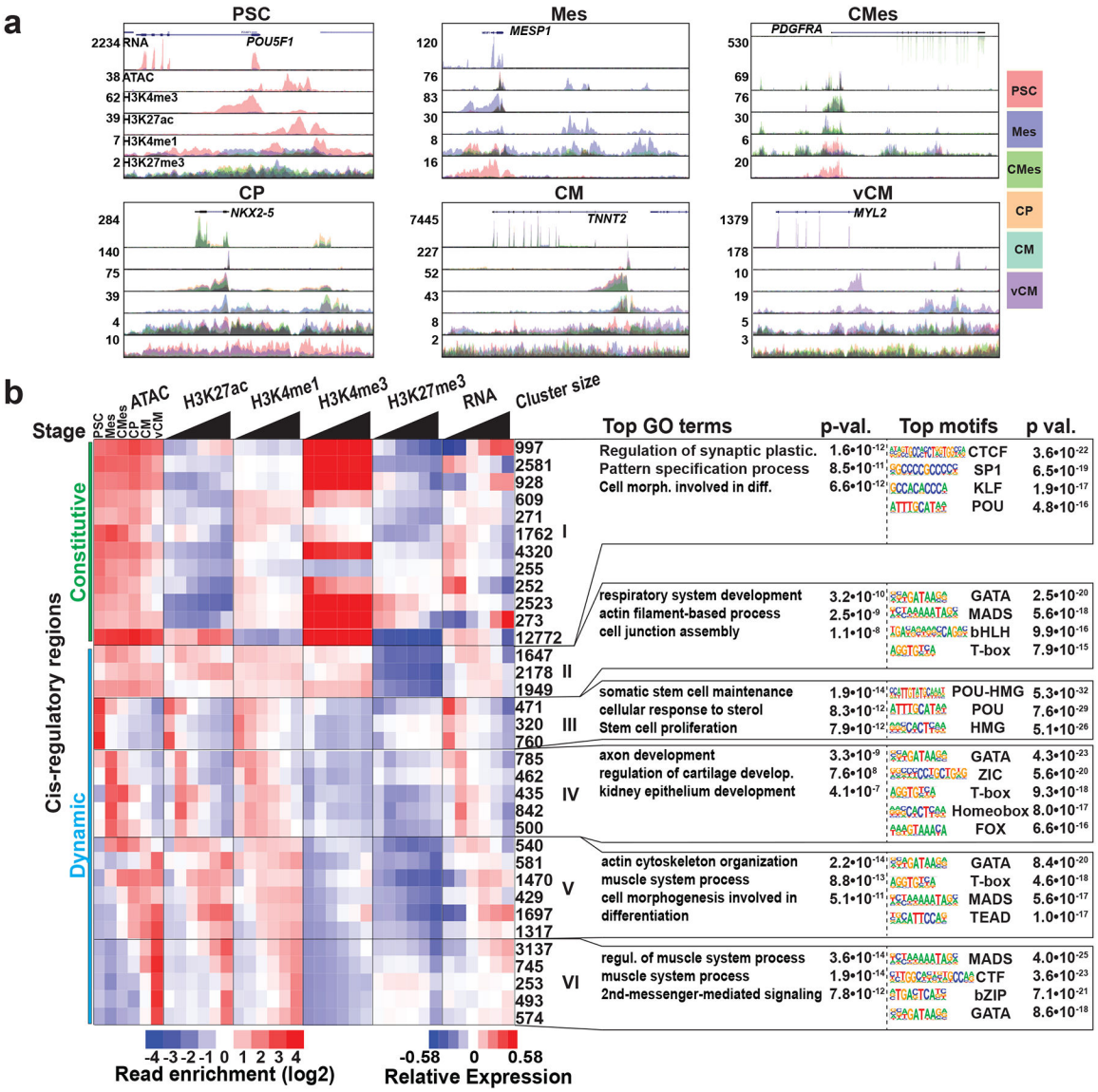


Fig. 1: Dynamic epigenomic changes correlate with cardiac developmental gene expression programs.

a, ChIP-, ATAC- and RNA-seq profiles are shown for representative marker genes for each cardiac developmental stage. **b**, Clustered heatmaps of ATAC-seq, ChIP-seq and RNA-seq reveal developmental stage-specific gene regulatory programs. Each row represents a set of related CREs. The number of CREs is indicated on the right. The RNA-seq heatmap is based on expression of genes that correlate to cognate CRE cluster. Top enriched GO terms and transcription factor motifs are displayed for each cluster. PSC, pluripotent stem cells; Mes, mesoderm; CMes, cardiac mesoderm; CP, cardiac progenitor; CM, cardiomyocyte; vCM, ventricular cardiomyocyte; val., value; GO, gene ontology; plastic., plasticity; morph., morphogenesis; diff., differentiation; develop., development; regul., regulation.

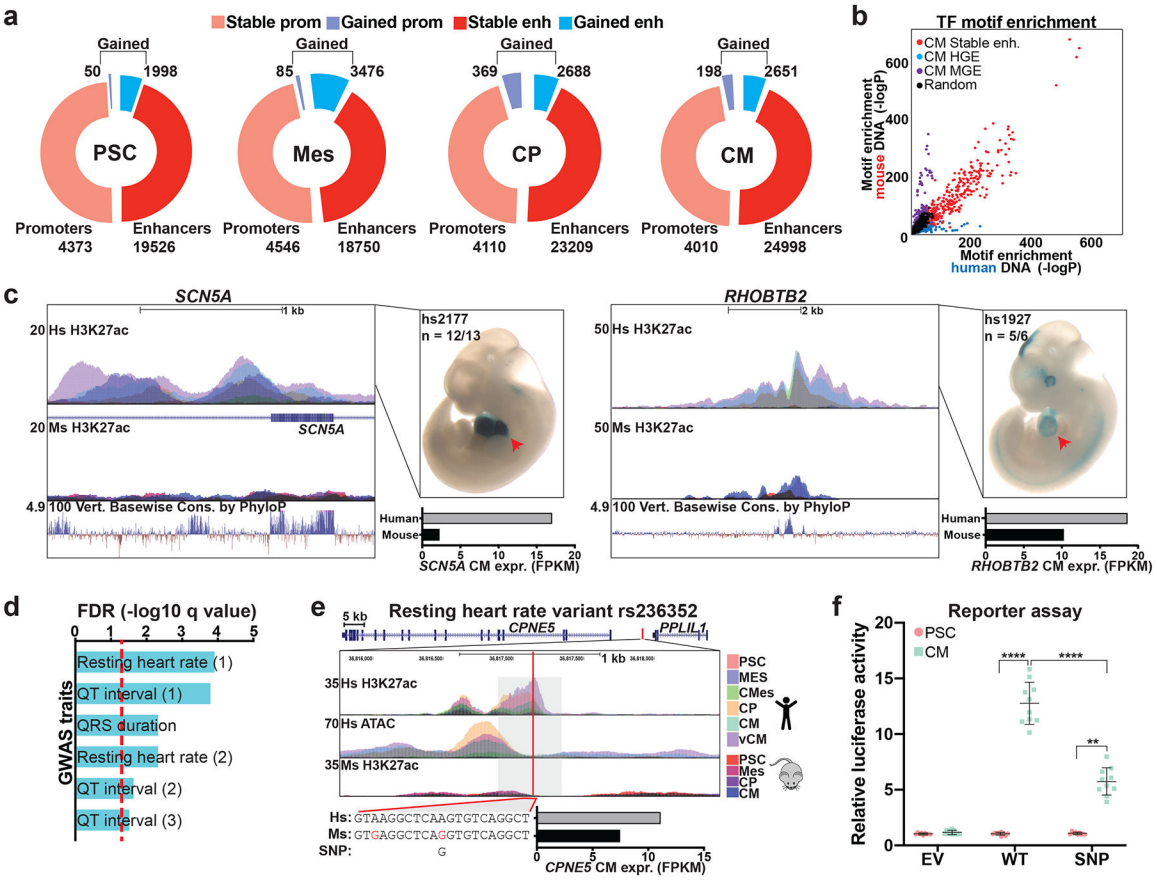


Fig. 2: Comparative human and mouse epigenomic analysis reveals human gained CREs that are associated with specific cardiac traits.

a, The number of stable and gained promoters and enhancers is displayed for each analyzed cardiac developmental stage. **b**, Scatterplot showing motif enrichment of TFs in human DNA (x-axis) and mouse DNA (y-axis) for the indicated enhancer groups and a comparable set of random regions as a control. Each dot represents a TF motif. Motif enrichment in one set of enhancers is compared to the enrichment in orthologous regions for that set of enhancers. **c**, VISTA database analysis reveals that HGEs near *SCN5A* and *RHOBTB2* (using VISTA elements hs2177 and hs1927, respectively) display cardiac *in vivo* reporter activity (number of positive embryos out of total number of injected embryos). RNA-seq shows that *SCN5A* and *RHOBTB2* exhibit greater expression in human cardiomyocytes than mouse cardiomyocytes. **d**, Bar plot shows significant enrichment of variants associated with species-specific cardiac traits within HGEs. The numbers in brackets indicate data sets from different GWAS studies for the same trait. Dashed red line indicates FDR q-value cutoff for significance ($p < 0.05$, Benjamini-Hochberg corrected). **e**, The resting heart rate associated rs236352 genetic variant resides in a cardiac HGE near *CPNE5*. Sequence comparison surrounding the SNP is shown on the bottom left with nucleotides that are not conserved between mouse and human indicated in red. *CPNE5* gene expression in human (grey) and mouse (black) cardiomyocytes is shown on the bottom right. **f**, Luciferase reporter assays show that this wild-type HGE (shaded area in **e**) can drive cardiomyocyte-specific enhancer activity compared to the control (EV) and that the rs236352-associated

risk allele (SNP) reduces this activity (n = 10). PSC, pluripotent stem cells; Mes, mesoderm; CMes, cardiac mesoderm; CP, cardiac progenitor; CM, cardiomyocyte; vCM, ventricular cardiomyocyte; prom, promoter, enh, enhancer; kb, kilobase; hs, human; ms; vert., vertebrate; cons., conservation; expr., expression; FPKM, fragments per kilobase of transcript per million mapped reads; FDR, false discovery rate; EV, empty vector; WT, wild-type; SNP, single nucleotide polymorphism. Data was analyzed using a two-tailed Student's t-test. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$. Data are represented as the mean ± standard error of the mean (SEM) with individual data points.

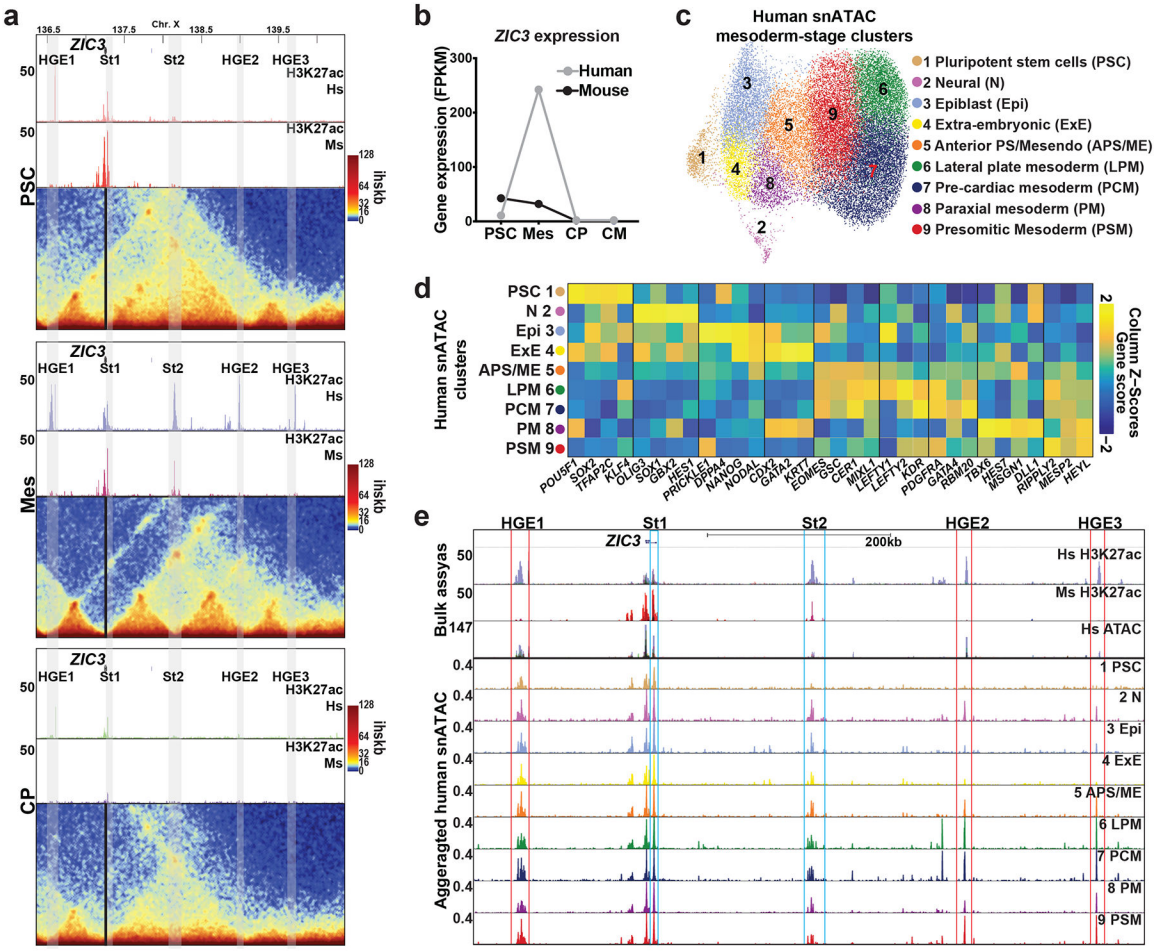


Fig. 3: Human-gained enhancers in *ZIC3* locus establish a potential *ZIC3* mesoderm-related gene regulatory network that directs elevated human *ZIC3* expression at mesoderm stage.

a, *in situ* Hi-C assays for human pluripotent stem cells (PSC), mesoderm progenitors (Mes) and cardiac progenitors (CP) reveal chromatin interactions between *ZIC3* promoter (black line) and stable (St) and human gained enhancers (HGE) shaded in grey. H3K27ac ChIP-seq from human data is compared to mouse H3K27ac ChIP-seq mapped onto the human genome. **b**, Human *ZIC3* gene expression is increased at mesoderm stage compared to mouse *Zic3* during CM differentiation. **c**, UMAP visualization of human PSC-derived mesoderm snATAC-seq data reveals distinct mesoderm clusters/subpopulations. **d**, Mesoderm clusters are annotated based on gene scores of known developmental markers. **e**, Bulk human (Hs) and mouse (Ms) H3K27Ac ChIP-seq, bulk human ATAC-seq and aggregated human snATAC-seq profiles across the *ZIC3* locus for each cluster show differential chromatin accessibility of *ZIC3* HGEs, in particular HGE2 and HGE3 (red outlined boxes), but not stable enhancers (blue outlined boxes) between mesoderm clusters. Chr., chromosome; HGE, human-gained enhancer; St, stable; hs, human; ms, mouse; Ihskb, interactions per hundred square kilobase per billion interactions. PSC, pluripotent stem cells; Mes, mesoderm; CP, cardiac progenitor; CM, cardiomyocyte; HGE, human-gained enhancer; FPKM, fragments per kilobase of transcript per million mapped reads; N, neural; Epi, epiblast; ExE, extra-embryonic; APS/ME, anterior primitive streak-/mesendoderm;

LPM, lateral-plate mesoderm; PCM, pre-cardiac mesoderm; PM, paraxial mesoderm; PSM, pre-somitic mesoderm.

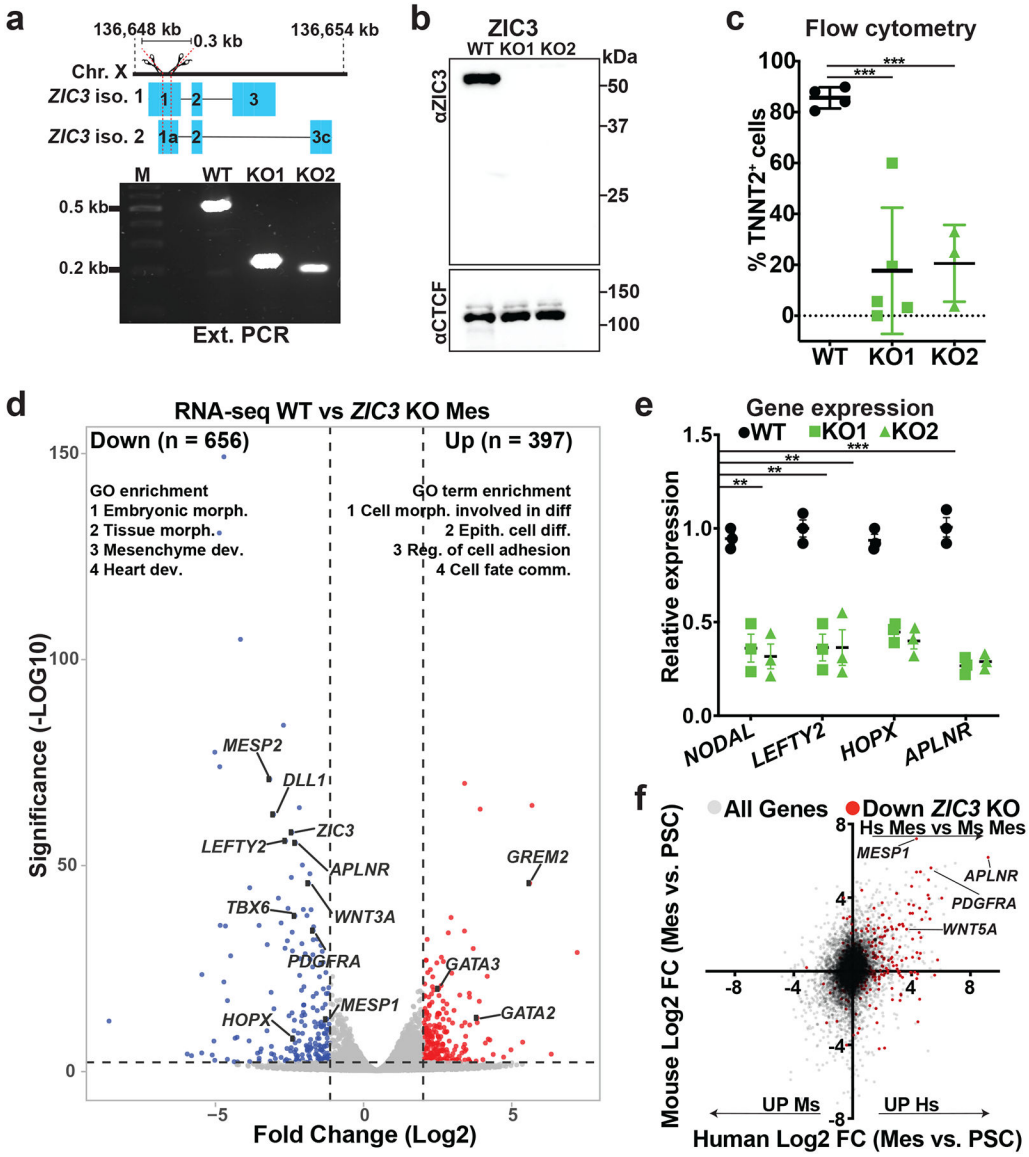


Fig. 4: *ZIC3* is required for human cardiomyocyte differentiation.

a, Top diagram displays the targeting strategy to generate human PSC *ZIC3* knockout (KO) lines through CRISPR/Cas9-mediated deletion of a 0.3 kb region in exon 1 of *ZIC3* that is common to both *ZIC3* isoforms. Bottom panel shows PCR genotyping that confirmed two human PSC *ZIC3* KO lines (versus wild-type/WT human PSCs). **b**, Western blot using a ZIC3 antibody further validated *ZIC3* KO clones. CTCF was used as a loading control. **c**, Human PSC *ZIC3* KO lines display reduced human cardiomyocyte differentiation as assessed by TNNT2 flow cytometry at D15 (n = 3–5). **d**, Volcano plot shows differentially expressed genes and associated GO-term enrichments between WT and *ZIC3* KO human mesoderm cells (adj. p-value < 0.05 and FC ≥ 2). **e**, qPCR validates downregulated NODAL signaling related genes (*NODAL* and *LEFTY2*), as well as early cardiac differentiation (*HOPX* and *APLNR*) genes in *ZIC3* KO human mesoderm cells (n = 3). **f**, Scatterplot shows enrichment of *ZIC3*-dependent genes (red dots) among genes upregulated in human (hs)

mesoderm compared to mouse (ms) mesoderm (upper right quadrant). Data was analyzed using a two-tailed Student's t-test. *p < 0.05; **p < 0.01; ***p < 0.001; ****p < 0.0001. Data are represented as the mean ± SEM with individual data points. Chr., chromosome; kb, kilobase; iso., isoform; M, marker; WT, wild-type; KO, knock-out; Ext., external; PSC, pluripotent stem cells; Mes, mesoderm; adj., adjusted; FC, fold change; GO, gene ontology; morph., morphogenesis; dev., development; epith., epithelial; diff., differentiation; reg., regulation; comm., commitment.

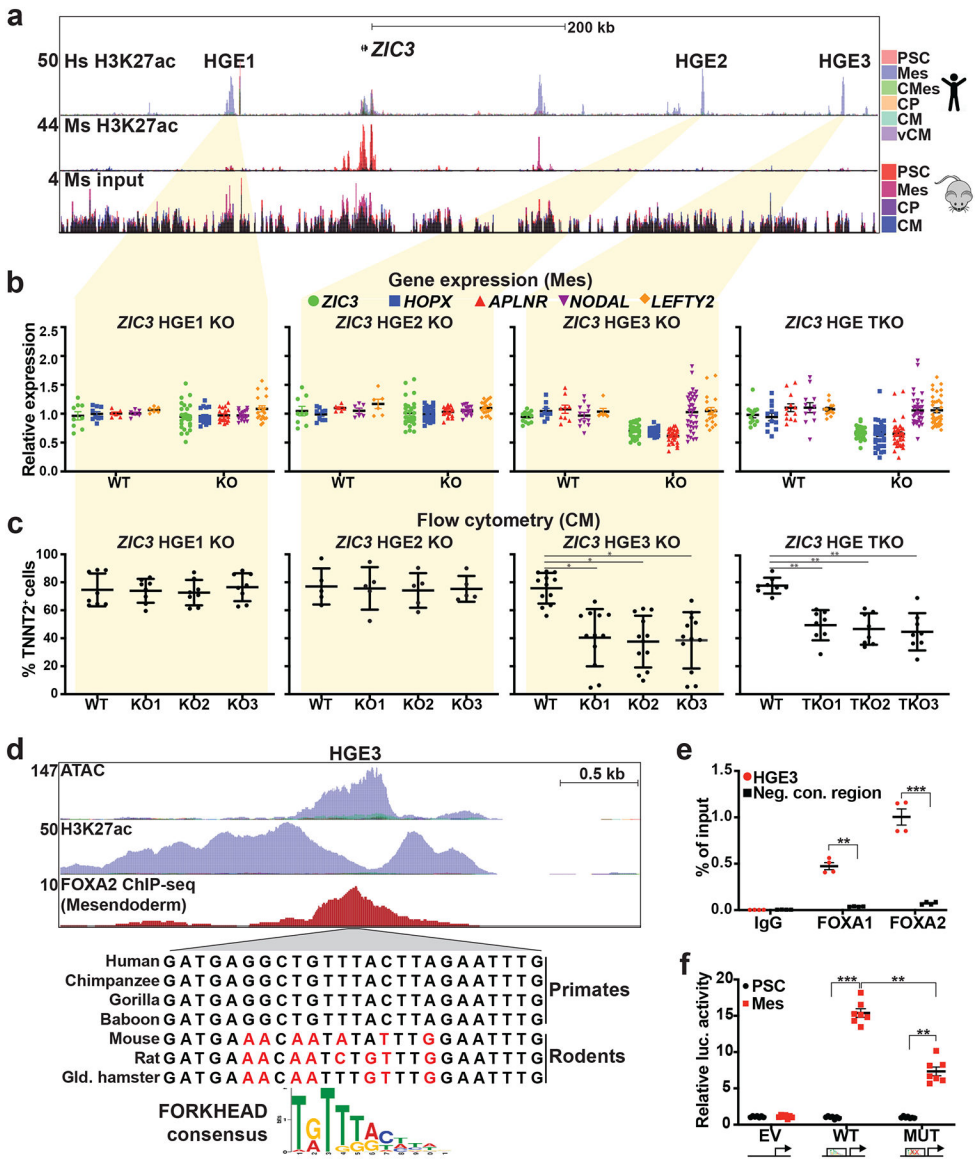


Fig. 5: A *ZIC3* human gained enhancer regulates mesoderm-specific *ZIC3* expression and cardiomyocyte differentiation.

a, Chromatin profiles of human (Hs) and mouse (ms; mapped onto the human genome) H3K27ac ChIP-seq profiles at the *ZIC3* locus show three human-gained enhancers (HGE1–3). Mouse input DNA mapped onto the human genome is shown to highlight mouse/human conserved DNA regions. Each track contains data for all examined stages. **b**, qPCR analysis reveals that *ZIC3* HGE3 knockout (KO) and HGE triple knockout (TKO) lines specifically exhibit reduced expression of early cardiac differentiation genes (*HOPX* and *APLNR*) but not NODAL signaling related (*NODAL*, *LEFTY2*) genes. For each KO, data from individual clones were pooled. For plots with all individual data points and statistics, see Extended Data Fig. 8a. *TBP* gene expression was used for gene expression normalization. **c**, *ZIC3* HGE3 knockout and HGE triple knockout human PSC lines display reduced CM differentiation as assessed by TNNT2 flow cytometry at D15 (n = 6–12). **d**, *ZIC3* HGE3 contains a highly conserved FORKHEAD motif within an ATAC-seq peak that is

bound by FOXA2 according to FOXA2 ChIP-seq and surrounded by a H3K27ac mark. **e**, ChIP-qPCR analysis reveals FOXA1/2 binding to *ZIC3* HGE3 in mesoderm cells (n = 4, two technical replicates for each biological replicate and two biological replicates for each condition). Negative Control (Neg. Con.) is a region that has no identified enhancer activity. **f**, Luciferase reporter assays reveal that the HGE3 region (wild-type/WT), compared to a control (EV, empty vector), can drive mesoderm but not human PSC reporter activity. This activity was reduced when the FORKHEAD motif was replaced with corresponding mouse sequences as denoted by red nucleotides in **d** (Mutant/MUT). Luc, luciferase. n = 7. Data was compared using a two-tailed Student's t-test. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$. Data represented as the mean ± SEM with individual data points. HGE, human-gained enhancer; kb, kilobase; PSC, human pluripotent stem cells; Mes, mesoderm; CMes, cardiac mesoderm; CP, cardiac progenitor; CM, cardiomyocyte; vCM, ventricular cardiomyocyte; Luc, luciferase.