Conserved roles of GATA4 and its target gene TBX2 in regulation of human cardiogenesis

Nicola Graham[1], Pavel Kirilenko[1], Ilya Patrushev[2], Ewan D. Fowler[1], Peter Kille[1], Michael Gilchrist[2], Nick D.L. Owens[3] and Branko Latinkic[1, *]

1 School of Biosciences, Cardiff University, Cardiff, United Kingdom

2 The Francis Crick Institute, London, United Kingdom

3 Department of Clinical and Biomedical Sciences, Exeter University, United Kingdom

*Corresponding author: latinkicb@cardiff.ac.uk

Abstract

The transcription factor (TF) GATA4 is a key mediator of cardiogenesis. GATA4 regulates cardiogenesis through the expression of its target genes, only some of which have been identified. We have used a gain of function model based on pluripotent embryonic ectoderm explants from *Xenopus* embryos expressing GATA4, to identify a set of downstream targets of GATA4 which are also regulated by Nodal, a known cardiogenic signal. GATA4 was shown to be required for the expression of target genes *tbx2* and *prdm1* in vivo, likely acting in a direct fashion by interacting with their regulatory regions. In addition, *tbx2* and *prdm1* are shown to have roles of their own in vivo, as downregulation of *tbx2*, a positive target, and overexpression of *prdm1*, a negative target, interferes with cardiac development in *Xenopus* embryos.

The conservation of the *GATA4-TBX2-PRDM1* regulatory relationship was shown in human iPSC-derived cardiomyocytes. Loss of function of GATA4 lead to downregulation of *TBX2*, upregulation of *PRDM1* expression and failure of cardiogenesis. GATA4-deficient cells failed to form normal cardiomyocytes, with most cells adopting alternative fates and only a small minority expressing an aberrant cardiomyocyte phenotype. Genome-wide transcriptomic analysis documented severe reduction of cardiomyocyte and endothelial cell transcriptomes and upregulation of transcriptional profiles of smooth muscle cells and fibroblasts. Disruption of TBX2 function did not alter cardiomyocyte differentiation efficiency but led to the formation of hypertrophic cardiomyocytes characterised by defective sarcomeres and deficient calcium signalling. In addition, we show that whilst *PRDM1* is not essential for formation of cardiomyocytes it is implicated in suppression of alternative cell fates.

The results presented establish a conserved regulatory relationship between GATA4 and its target genes *TBX2* and *PRDM1* and roles for these genes in the modulation of cardiomyocyte development, expanding the cardiac gene regulatory network and providing further insight into how cardiogenesis proceeds.

Introduction

Heart development and homeostasis is controlled by the collective action of numerous transcription factors (TFs) which come together to form complex and dynamic gene regulatory networks (GRN) that define and maintain cellular identity. Disruption of the cardiac GRN is associated with the development of congenital heart defects (CHDs), and cardiovascular disease (CVD) in later life (Fahed et al., 2014; Tan et al., 2002). The established core network of TFs that orchestrate cardiac development, includes GATA4/5/ 6, NKX2-5, HAND1, 2, TBX2/ 5/ 20 and proteins belonging to the MADS domain family such as MEF2 and SRF (Akerberg and Pu, 2020; Bruneau, 2013; Olson, 2006; Paige et al., 2015). Evidence from mouse knockout models has shown that the absence of any one of these factors results in an array of severe cardiac defects that almost invariably lead to embryonic lethality (Bruneau et al., 2001; Firulli et al., 1998; Koutsourakis et al., 1999; Kuo et al., 1997; Srivastava et al., 1997; Stennard et al., 2005; Watt et al., 2004) highlighting vulnerability of the cardiac GRN to perturbation.

GATA4 is a core member of the cardiac GRN with well documented roles in all stages of heart development as well as in post-natal life. Early experiments in mice identified an indirect role for GATA4 in visceral endoderm which is essential for cardiac morphogenesis in the overlying mesoderm (Kuo et al., 1997; Molkentin et al., 1997). Subsequent studies have identified cell autonomous roles for GATA4 in the developing heart (Watt et al., 2004; Zeisberg et al., 2005). Evidence for the importance of GATA4 in human heart development is exemplified by the characterisation of a heterozygous mutation, G296S, linked with congenital heart disease (CHD) (Garg et al., 2003). Cardiomyocytes generated from CHD patient-derived induced pluripotent stem cells (iPSCs) show defects in contractility and metabolism (Ang et al., 2016). The molecular basis for this phenotype was shown to be the inability of the mutant GATA4 protein to recruit TBX5 to common targets in chromatin, leading to downregulation of cardiac targets and to aberrant de-repression of endothelial targets (Ang et al., 2016). It is worth noting that this heterozygous mutation in patients is associated with the occurrence of a

range of atrial and ventricular septal defects of a severity often requiring surgical correction (Garg et al., 2003). In contrast, heterozygous mice carrying the corresponding mutation appear grossly normal, with a proportion showing mild defects in post-natal atrial septal closure (Misra et al., 2012), highlighting the limitations of mouse models of CHD.

In addition to being required for heart development, GATA4 has been shown to be sufficient to drive cardiomyocyte differentiation of permissive embryonic cells either on its own in pluripotent ectodermal explants from *Xenopus* embryos (Latinkić et al., 2003), or together with Baf60c in mouse early embryonic mesoderm (Takeuchi and Bruneau, 2009). Furthermore, GATA4 is used in most cocktails of cardiac TFs for reprogramming of non-cardiac cells to cardiomyocytes (Garry et al., 2022). For example, *GATA4* along with additional cardiac GRN members *MEF2* and *TBX5* is sufficient for reprogramming of fibroblasts to cardiomyocytes (Garry et al., 2022; Xie et al., 2022). The requirement for GATA4 for normal cardiac development, as outlined above, and the ability of GATA4 to steer the genomic program of non-cardiomyocytes towards a cardiomyocyte fate reaffirms it as a key component of the cardiac GRN.

Mechanistic studies focused on mapping of GATA4 occupancy sites in chromatin in different models such as mouse embryos, cardiomyocytes derived from human iPSCs and adult hypertrophic hearts, are beginning to identify the roles of GATA4 at the chromatin level by defining its stage-, signal- and cell type-specific targets. Besides the core cardiac GRN, numerous additional transcription factors regulate multiple aspects of cardiac development such as cardiac cell type diversification and morphogenesis. The emerging expanded cardiac GRN includes genes such as *TBX2* which has been shown to regulate the formation of the Atrio-Ventricular Canal (AVC) by repressing the transcriptional program for chamber cardiomyocytes. To describe an expanded cardiac GRN, it will be necessary to obtain and integrate a better knowledge of regulatory relationships between its nodes.

A simple model for studying the cardiac GRN is offered by overexpression of GATA4 in pluripotent ectodermal explants from blastula stage *Xenopus* embryos. GATA4 is sufficient to reprogram explant cells fated to become epidermis towards cardiomyocytes in a process which mimics normal heart development by producing beating cardiac tissue (Latinkić et al.,

2003). In this model, in addition to cardiomyocytes GATA4 induces endothelial cells as well as anterior endoderm, in keeping with its known roles in vivo. GATA4-driven cardiogenesis was shown to be enhanced by Wnt antagonism (Latinkić et al., 2003), in agreement with the well documented role of Wnt antagonism in heart development.
Besides GATA4 expression, pluripotent explants from *Xenopus* embryos can be directed towards cardiomyocyte fate by Nodal/Activin signalling, the principal inducer of mesoderm and endoderm in vertebrate embryos (Logan and Mohun, 1993; Takahashi et al., 2000). In *Xenopus* explants, Nodal signalling acts in a concentration dependent manner, with a high dose required for induction of endoderm and cardiomyocytes and a lower dose for skeletal muscle (Logan and Mohun, 1993; Takahashi et al., 2000).

As both GATA4 and Nodal induce cardiomyocytes, we used both cardiogenic triggers in this report to identify and characterise putative components of the cardiac GRNs as co-regulated differentially expressed genes (DEGs). A subset of DEGs were validated as GATA4 targets in vivo as well as in explants. Interference with normal expression of two such targets, *tbx2* and *prdm1* is shown to affect heart development in *Xenopus* embryos.
The wider relevance of the data obtained in *Xenopus* is shown by our results which demonstrate the essential roles of GATA4 and its target gene *TBX2* in differentiation of human cardiomyocytes from iPS cells.

Results

Identification of GATA4 targets under cardiogenic conditions

To dissect the genomic program for cardiogenesis regulated by GATA4 we employed a well-established gain of function model using pluripotent ectodermal explants from blastula stage *Xenopus* embryos (Latinkić et al., 2003). As GATA4 plays key roles in the development of multiple cell types of mesodermal and endodermal origin, and is sufficient to specify endodermal, endothelial and blood cells in *Xenopus* animal cap explants, we designed an experimental approach for selection of cardiogenesis-related genes regulated by GATA4. This consisted of defining a set of genes regulated by cardiogenic conditions: GATA4 (G4), GATA4 in the presence of Wnt antagonist Dkk-1 (G4D), which enhances cardiogenesis, and high level of Nodal/Activin signalling, known to induce cardiogenesis (Fig. 1, S1). At the same time, these genes were not regulated by a lower level of Nodal/Activin which induces skeletal but not cardiac muscle, and by Dkk-1 alone (Fig. 1, S1). In this model it takes approximately 2 days for GATA4 to induce cardiac differentiation as assessed by expression of cardiomyocyte markers (Fig. 1A, S1A).

We focussed on the early stages of the genomic program which was expected to include immediate-early and delayed-early targets, by sampling at 2.5h, 5h and 7.5h after excision of animal cap explants and activation of inducible GATA4-GR fusion by addition of dexamethasone (DEX). We performed polyA+ RNA-seq and calculated DEGs at each stage relative to control. The three conditions produced a highly concordant impact on gene expression for each time point (Fig. S1B-D); therefore, we combined these to define DEGs per condition. We found 1,092 DEGs (690 activated, 402 repressed) for GATA4 (G4), 1,067 (602 activated and 465 repressed) for GATA4 + Dkk-1 (G4D), and 1,696 (829 activated and 867 repressed) for Nodal. Comparing between conditions we found exceptional overlap of 914 genes differentially expressed in G4 and G4D, 83.7% and 85.7% of their genes respectively (Fig. 1B). The overlap with Nodal DEGs was more limited with 403 genes shared between the three cardiogenic treatment groups (Fig. 1B). We performed gene set enrichment using Enrichr (Kuleshov et al., 2016) (Fig. 1D, S2; Supplemental Table 1), and we calculated enrichments for all intersection groups indicated in the upset plot (Fig. 1B). We found a clear induction of cardiac gene expression programmes in the intersection of GATA4 and Nodal DEGs, with terms such as Heart Development (FDR < 10^{-7}, intersection of

G4/G4D/Nodal) and Cardiac Progenitor Differentiation (FDR < 10^{-5}, intersection of G4/G4D/Nodal). We further found an enrichment in Gene Ontology (GO) terms related to transcription factor driven regulation of gene expression in the intersection of GATA4 and Nodal conditions, with enrichment of terms such as Regulation Of Transcription By RNA Polymerase II (FDR < 10^{-23}) (Fig. S2).

Validation confirmed DEGs which satisfied the selection criteria (regulated by G4, G4D and high Nodal but not low Nodal/Activin) at all time points such as *fst*, *map3k12*, *rgs2* (Fig. 1E). Other DEGs were validated at some but not all time points.

Upregulated DEGs include known Nodal target *lefty1* (Cheng et al., 2000); at 2.5h *lefty1.S* is strongly induced by Nodal, as expected from a direct target of Nodal signalling (Fig. 1E). However, as the gene is also induced by low level of activin/Nodal signalling at that time, it is not validated as a DEG induced by cardiogenic conditions. The expression of *lefty1.S* induced by cardiogenic but not non-cardiogenic conditions is sustained beyond 2.5h, thus identifying it as a DEG (Fig. 1E). Additional upregulated DEGs include *igf2* and *ephA4* (Fig. 1F), Tbx family transcription factors *tbx2* and *tbx5* (Fig. 2D) and *gata2, hoxc13* and *lmo2* (Fig. S3).

In addition to upregulated DEGs, we have identified DEGs which are downregulated in response to cardiac inducing conditions (Fig. 1G). They were most consistently observed at 2.5h, the proximal time point at which direct action of G4, G4D and Nodal is most likely. Downregulated DEGs include transcription factors *prdm1* and *id3*, signalling molecules such as *pllk2*, *tnfsf10l, rgl2*, E3 ubiquitin ligase *trim29* and ubiquitin peptidase *usp43* (Fig. 1G). Since *id3* is a member of a small family of transcription factors (*id2-4* in *Xenopus*) with described developmental roles including the heart (Cunningham et al., 2017), we have also examined the expression of *id2* and *id4* under the conditions of our screen. As shown in Fig. 2A, in contrast to *id3, id2* and *id4* are not found to be differentially expressed.

Whilst the screen conditions induced a programme of cardiogenesis, none of the three treatments is strictly cardiogenic. Indeed, our GO analysis has indicated enrichment of terms such as Endoderm Differentiation (Fig. 1D) and we found early endodermal genes among DEGs at 2.5h, transcription factors *osr1*, *osr2*, *sox17a*, *sox17b* as well as *hhex* (Fig. 1H).

Regulation of *tbx2*, *id3* and *prdm1* by GATA4

To validate identified DEGs as physiologically relevant targets of GATA4 we examined their regulation by GATA4 in vivo. As shown in Fig2B, downregulation of GATA4 using morpholino oligonucleotides (MOs) leads to upregulation of *id3* at gastrula and neurula stages; interestingly, even though *id2* and *id4* were not confirmed as DEGs, their expression was also upregulated by GATA4 MOs, suggesting an indirect mode of action. Downregulation of GATA4 and most closely related cardiogenic GATA factor, GATA5, led to upregulation of *prdm1*, and to downregulation of *tbx2* (Fig. 2E). This analysis has shown that GATA4 is required for normal expression of *id3*, *prdm1* and *tbx2* in vivo.

We next asked if GATA4 directly regulates *id3*, *prdm1* and *tbx2* through association with their regulatory regions. For this analysis we identified consensus GATA binding sites within 10 kb upstream and 2 kb downstream of their translation start sites and tested their presence in genomic DNA associated with exogenous GATA4. For all 3 genes, the minority of tested sites were found to associate with GATA4 (Fig. 2C, F and Fig S3B-C), suggesting that GATA4 is directly and specifically regulating their expression.

We subsequently tested if *tbx2* and *prdm1* have roles of their own in *Xenopus* heart development. The expression of *tbx2*, a GATA4 target upregulated by cardiogenic conditions, was downregulated by MOs, and this led to reduced and deformed heart formation in majority of embryos (Fig. 2G and Fig. S3D). The role of *prdm1* as a target of GATA4-mediated downregulation was tested by mRNA-mediated overexpression, which caused morphological defects such as cardia bifida and linear heart tube (Fig. 2G and Fig. S3E). These results indicate that *tbx2* and *prdm1* are targets GATA4 in cardiogenesis and that their normal expression is required for early heart development in *Xenopus* embryos.

GATA4 is required for normal expression for *PRDM1* and *TBX2* during human iPSC-CM differentiation

We next sought to establish whether these findings are translatable to human development using iPSC-cardiomyocyte (iPSC-CM) differentiation to model human cardiomyogenesis. Firstly, the expression of *GATA4, PRDM1* and *TBX2* was examined during normal iPSC-CM

differentiation. Consistent with previous studies *GATA4* expression is observed from the onset of differentiation with a noticeable increase in expression at day 4 (Fig. 3A), coincident with the establishment of the early cardiac progenitor (CP) pool (Churko et al., 2018). Its expression rises throughout differentiation peaking approximately when beating becomes apparent (day 9-10 of differentiation). Expression of *PRDM1* peaks early at day 3 and decreases when *GATA4* expression initially peaks at day 4 (Fig. 3A). Conversely *TBX2* expression becomes prominent only after the peak in *GATA4* expression. The expression patterns of these genes support the notion that the regulatory relationship between GATA4, *PRDM1*, and *TBX2* may be a conserved feature of cardiogenesis.

To establish whether the expression pattern of *PRDM1* and *TBX2* depends on GATA4, CRISPR-Cas9 gene editing was used to inactivate *GATA4* (Fig. S4). Knockout of the protein was confirmed by western blot (Fig.3B). RNA-sequencing and RT-PCR was then used to examine the expression of the target genes in the GATA4 null line. As shown in Fig. 3C-D, the expression of *PRDM1* and *TBX2* is disrupted in the absence of GATA4, with *PRDM1* expression being 2.1-fold higher than normal, and *TBX2* expression being 2.7-fold lower than normal. These results are consistent with the hypothesis that GATA4 regulates *TBX2* and *PRDM1* in cardiogenesis in humans as well as in *Xenopus*. Additional DEGs identified in the *Xenopus* screen showed dysregulation in the GATA4 null iPSC-CMs (Fig. 5H), suggesting that GATA4-regulated GRN has more extensive conservation besides TBX2 and PRDM1.

GATA4 deficient iPS cells fail to form functional cardiomyocytes

The GATA4 null iPSC lines were differentiated using Wnt modulation protocols (Burridge et al., 2014; Lian et al., 2012) and were examined up to Day 32 of differentiation. No beating was observed in 12 independent differentiation experiments. Staining for TNNT2 by immunofluorescence was used to detect cardiomyocytes in iPSC-CM differentiation of WT and GATA4 null lines. On average there was a 4-8 fold decrease in the number of TNNT2 positive cells (TNNT2+) formed from GATA4 -/- iPSCs in comparison to WT (Fig. 4C). In the few mutant cells that expressed TNNT2 the signal also appeared weaker. Mean fluorescence per cell was used to quantify this and the average signal in the GATA4-/- TNNT2+ cells was found to be 2.5 times lower than for WT TNNT2+ cells (Fig. 4D). Examination of the cells at a higher magnification reveals that in the TNNT2+ GATA4-/- cells myofibrils lack a normal

striation pattern (Fig. 4E, S6A). Consistent with the results seen for TNNT2, staining for additional sarcomeric markers MYBPC3 (Fig. 4E, S6B) and filamentous actin (Fig. S6C) was also disrupted in the GATA4 null lines. In contrast, staining for general cytoskeletal protein α-tubulin (TUBA1A) appears normal (Fig. S6D).

The physiological activity of the WT and GATA4-/- cells was examined at Day 32 of differentiation using calcium imaging. Figure 4F shows that calcium flux can be observed across a large proportion of WT cells; conversely in the GATA4-/- cells this activity is restricted to a few small regions (Fig. 4F). When regions of interest for each are plotted individually it is apparent that calcium cycling in the WT population is frequent and regular (Fig. 4G). In the GATA4 null cells the transients detected were on average 6 times slower than their WT counterparts. This together with the low differentiation efficiency, rare and low TNNT2 expression, and lack of sarcomere formation in the cells lacking GATA4, accounts for the failure to form beating cardiomyocytes.

GATA4 null cells likely adopt alternative cardiac cell fates

The vast reduction in the number of cardiomyocytes (TNNT2 and MYBPC3 positive cells) in the mutants raises the question of the identity of the remaining cell types.

To address how GATA4 null lines diverge from the WT path of differentiation, RNA-sequencing was carried out on cells at day 2, 5, and 10 of differentiation. At day 2 expression of GATA4 is low in the WT line therefore a large effect on the transcriptome at this time was not expected. Indeed, the GATA4 null cell transcriptome is comparable to that of the WT cells (Fig. S5A,B) as is the expression of early mesodermal and cardiac mesoderm markers *T* and *MESP1,* respectively (Fig. S5F).

At day 5 in the WT population the expression of cardiac progenitor genes including GATA4 rises. The levels of these genes are lower in the GATA4 null population and at day 10 the discrepancy between the WT and GATA4 null line becomes exaggerated. At day 10 the most affected cardiac GRN member is *TBX5*, being almost undetectable in the GATA4-/- cells (Fig. 5A,B)**.** However, not all members of the cardiac GRN are downregulated: for example, *MEF2C* expression is expressed at a similar level in both lines and *ISL1* expression is increased, strongly suggesting that members of cardiac GRN display differential requirements for GATA4 function (Fig. 5A,B). Analysis performed at additional timepoints demonstrates that *TBX5* expression is undetectable during cardiomyocyte differentiation of

the GATA4 null line (Fig. 5B. In contrast, expression of *NKX2-5* is readily detectable and expression of *ISL1* is higher after Day 6 (Fig. 5B).

During directed differentiation of iPS cells to cardiomyocytes other cardiovascular cell types are formed such as fibroblasts, endothelial, and smooth muscle cells (Grancharova et al., 2021; Jiang et al., 2022). Therefore, these cell types are the most likely candidates for the observed non-cardiomyocytes in GATA4 null mutant cell population.

In agreement with previous report that has identified an essential requirement for GATA4 in the formation of the epicardium (Watt et al., 2004) a decrease in the expression of epicardial marker genes (*TBX18* and *WT1)* was observed (Fig. 5C). Furthermore, the expression of endothelial marker genes *CDH5* and *PECAM1* was decreased to a similar degree as what was seen for the cardiomyocyte specific gene set, suggesting that cardiomyocytes, epicardial cells and endothelial cells all require GATA4 for their development in this model (Fig. 5C-E). The expression of markers which are highly but not exclusively expressed in smooth muscle cells such as *VIM, TAGLN,* and *ACTA2* was upregulated in the GATA4-/- mutant cells (Fig. 5C-E). Marker of activated fibroblasts *POSTN* (Fig. 5C) was also upregulated, as well as numerous ECM components (Fig. 5C and Supplemental Table 2). Our gene enrichment analysis indicates significant over-representation of multiple ECM terms, consistent with upregulation of smooth muscle and fibroblast markers. Among the genes upregulated in GATA4 -/- cells at Day 10 is *BNC2*, a recently described master regulator of myofibroblast differentiation (Fig. 5C) (Bobowski-Gerard et al., 2022). Closer inspection of cellular morphologies of mutant TNNT2+ cells revealed an appearance of highly convoluted cells suggestive of a myofibroblast phenotype (Fig. 5F).

Together, these results suggest that the loss of GATA4 adversely affects the development of the cardiomyocytes and endothelial lineages, whilst simultaneously promoting the development of related mesenchymal cells- activated fibroblasts, myofibroblasts and smooth muscle cells.

To see if GATA4 could directly regulate DEGs, we analysed the high resolution GATA4 ChIP-seq data set for adult and foetal mouse heart (Akerberg et al., 2019). From our dataset at day 5, 159 DEGs were found to be consistently differentially regulated by a log2FC value ≥2. Of these genes 80% were associated with ≥1 GATA4 binding sites. At day 10 by these criteria 453 genes were found to be differentially regulated and of these 77% were found to be

associated with ≥1 GATA4 binding sites (figure 5I-J). Importantly these included genes differentially expressed in both *Xenopus* and human RNA-seq datasets (Fig. 5H). Both up- and down-regulated DEGs were identified as direct targets based on GATA4 binding, in agreement with known activities of GATA4 in both transcriptional repression and activation.

TBX2 is required for normal cardiomyocyte differentiation

We next examined whether the GATA4 target genes identified and validated in the *Xenopus* screen are themselves required for normal iPSC-CM differentiation. CRISPR-Cas9 gene editing of the TBX2 locus caused an in-frame deletion of DNA binding domain, likely resulting in a loss of function mutation (Fig. 6A,B, Fig. S8). The *TBX2* mutant iPS cell line produces cardiomyocytes at a similar rate to WT iPS cells (Fig. 6C). However, the cells produced are on average 1.6-fold larger than WT cells at day 12 of differentiation. This difference is accentuated at day 32, with the size of the mutant cells around 2.3-fold larger (Fig. 6D). Additionally, examination of the myofibril and sarcomeric structure of the *TBX2* mutant cardiomyocytes by immunofluorescence (Fig. 6F-H) reveals them to be significantly more disorganised (74%) than their WT counterparts (40%).

In the *TBX2* mutants the amplitude of the calcium peaks recorded was notably higher, and there was a larger range of amplitudes observed indicating that there is a wider heterogeneity in the Ca^{++} handling properties of the mutant cardiomyocytes (Fig. 6 J-K). *TBX2* mutants also displayed increased propensity for formation of spontaneous calcium sparks, indicating defective calcium handling (Fig. 6L). Taken together these results indicate that interfering with TBX2 function promotes formation of maladaptive hypertrophic cardiomyocytes.

To further explore the phenotype of the cells, RNA-seq was conducted on samples taken at days 7 and 10. These time points correspond with the peak of *TBX2* expression during iPSC-CM differentiation and the onset of beating, respectively. Notable findings include increased expression at day 10 of *NPPA* and *NPPB* (Fig. 6N), genes whose elevated expression is known to be associated with heart failure (Tan et al., 2002). TBX2 is known to repress *NPPA* in AVC (Habets et al., 2002), thus increased *NPPA* expression in *TBX2* mutant iPSC-CMs strongly suggests that the mutation leads to loss of function. *VCAM1* is another gene that has been associated with heart failure and immune cell infiltration of the myocardium (Troncoso et al., 2021; Wang et al., 2021) and was also upregulated at day 10 (Fig. 6N). Several genes

involved in calcium regulation in cardiomyocytes were upregulated in *TBX2* mutant iPSC-CMs (Fig. 6M), potentially contributing to defective calcium handling in these cells. The increased expression of several sarcomeric genes e.g. *TNNT2*, *MYBPC3*, and *MYH7* in *TBX2* mutant cells is consistent with the hypertrophic phenotype (Fig. 6N). In agreement with this finding, GO terms such as Cardiac Muscle Cell Development were enriched (Fig. 6O, Supplemental Table 2).

PRDM1 deficient iPSC-CMs appear structurally unaltered but show premature beating

We next examined the requirement for *PRDM1* in iPSC-CM differentiation through the creation of a *PRDM1* null line using CRISPR-Cas9 gene editing (Fig. S10). Knockout of the protein was confirmed by western blotting (Fig. 7A). The *PRDM1*$^{-/-}$ line produced CMs at a similar efficiency to WT and the phenotype of these cells when examined by IF resembles that of WT cells (Fig. 7 C-D). In the PRDM1-/- line beating started around 2 days earlier on average than in the WT line (Fig. 7B). Gene expression analysis of several sarcomeric genes such as *MYL7, MYH6, MYH7,* and *TNNI3* indicates subtle mis-regulation, early increase at day 7 and subsequent decrease relative to the WT cells (Fig. 7E). RNA-seq results show that at day 3 when the level of PRDM1 is highest, upregulated DEGs greatly outnumber downregulated DEGs (Fig. 7H) and include genes associated with neural development and pluripotency (Supplemental Table 2) suggesting that transcriptional repressor PRDM1 might be fine-tuning the timing and levels of gene expression during cardiomyocyte differentiation and repressing early alternative gene programs.

Discussion

Cardiac development relies on the cooperative interaction of numerous transcription factors within the cardiac gene regulatory network. The knowledge of the composition of this network and how it operates is still incomplete. Here, using cardiogenic factors GATA4 and Nodal in pluripotent explants from *Xenopus* embryos we have examined the cardiogenic gene program. We identified a common gene set on which these cardiac inducing factors converge in order to regulate cardiogenesis. We focused on transcription factors *prdm1* and *tbx2* whose expression in X*enopus* embryos is regulated by GATA4, possibly directly. Using iPSC-CM differentiation and CRISPR-Cas9 gene editing we find this regulatory relationship to be conserved in human cardiomyogenesis and that target genes *PRDM1* and *TBX2* can modulate the cardiac gene program.

Our results demonstrate that GATA4 is essential for differentiation of human cardiomyocytes from iPS cells. Previous work in mouse, *Xenopus* and zebrafish models has established distinct stage-dependent roles for GATA4 during cardiomyocyte development, with defective heart still present in null mutant embryos (Borok et al., 2016; Crispino et al., 2001; Haworth et al., 2008; Holtzinger and Evans, 2005; Watt et al., 2004; Zeisberg et al., 2005). In the current study we show that human GATA4 -/- iPS cells lines have a greatly reduced capacity to form cardiomyocytes and do not show beating activity, a phenotype which is stronger than the phenotypes gained from GATA4 null mutation in animal models. Furthermore, a recent report by (Gonzalez-Teran et al., 2022) presented results for a human *GATA4* -/- iPS cell line that can form beating cardiomyocytes, albeit at a greatly reduced frequency. The reasons for this apparent discrepancy are not clear but likely include experimental details such as genetic background, which is known to be a major determinant of phenotypic variation in human iPSCs (Kilpinen et al., 2017).

The failure of GATA4 -/- cells to form beating cardiomyocytes in this study is underlined by a greatly reduced differentiation efficiency and defective sarcomere formation. Genome wide transcriptomic analysis has documented the downregulation of cardiomyocyte and endothelial cell gene sets, and an upregulation of genes associated with smooth muscle and activated fibroblast cell phenotype. These include a marker of activated fibroblasts *POSTN* (Shimazaki et al., 2008) and recently described master regulator of myofibroblasts *BNC2* (Bobowski-Gerard et al., 2022). A subpopulation of GATA4 null cells displayed myofibroblast-

like convoluted shape but was also found to be positive for TNNT2. Expression of cardiomyocyte-specific sarcomeric genes *TNNT2* and *MYBPC3* has been identified in a subpopulation of human cardiac fibroblasts (Litviňuková et al., 2020). It will be of interest to investigate if loss of GATA4 promotes transition of those fibroblasts to myofibroblast phenotype.

Loss of GATA4 also upregulates numerous genes involved in fibrosis and inflammation such as *COL3A1, COL8A1,* and *COL11A1*, *IL11* as well as *IGFBP7*, which is associated with senescence and heart failure (López et al., 2021; Moschen et al., 2011; O'Reilly, 2023; Xiang et al., 2017; Zhang et al., 2022).

The impact of GATA4 deficiency on cardiac GRN was not uniform: whilst the expression of *TBX5* was undetectable, the expression of *ISL1* (at Day 10) and *NKX2-5* (at Day 5) was upregulated in GATA4 -/- cells. This finding is consistent with the proposed repression of *NKX2-5* by GATA4/5/6 in vivo (Jiang et al., 1999). Of note, *NKX2-5* is known to be associated with ECM remodelling and pathological smooth muscle cell phenotype (Papaioannou et al., 2024). It will be of interest to determine if *NKX2-5* has a similar role in regulating pro-fibrotic phenotype in GATA4 -/- iPSC-CM cells.

The downregulation of *TBX18* and *WT1* observed in GATA4 -/- iPSC-CM cells suggests defective epicardial development, in agreement with an in vivo requirement of GATA4 for development of the epicardium in the mouse embryo (Borok et al., 2016; Watt et al., 2004). This is an apparently paradoxical observation as select markers of epicardial derivatives, smooth muscle and fibroblasts, are upregulated in GATA4 mutant human iPSC-derived cells. One possibility is that diminished epicardial development still permits formation of abnormal smooth muscle cells and fibroblasts.

GATA4 is known to be expressed in cardiac fibroblasts. Our results show that deletion of GATA4 leads to pro-fibrotic and myofibroblast gene expression signature, suggesting that it is required for the normal fibroblast phenotype. In agreement with this hypothesis, overexpression of GATA4 in cardiac fibroblasts was recently shown to reduce heart failure-induced fibrosis (Yamada et al., 2024). The role of GATA4 in attenuating fibrosis is not restricted to cardiac fibroblasts, as it also induces regression of liver fibrosis by deactivating hepatic stellate cells (Arroyo et al., 2021).

A potential mechanistic basis for the observed features of smooth muscle and activated fibroblasts in GATA4 null cells is offered by a recent study by Robbe *et al.* (Robbe et al., 2022) which demonstrated that GATA4 physically interacts with repressive chromatin modifier CHD4 recruiting it to non-cardiomyocyte gene regulatory sites in mouse myocardium. The deletion of a GATA site in the modulatory region of smooth muscle gene Myh11, led to its upregulation in cardiomyocytes (Robbe et al., 2022). Misexpression of Myh11 in cardiomyocytes has previously been shown to result in sarcomere disarray, and results in a significant decrease in ventricular output (Wilczewski et al., 2018). Thus, it is possible that removal of GATA4 leads to de-repression of non-CM genes which based on our RNA-seq data are part of gene expression signatures of smooth muscle cells and activated fibroblasts. Future work will determine the precise nature of non-cardiomyocytes which express these genes.

Our results show that besides being a conserved target of GATA4 during cardiogenesis, *TBX2* is required for normal development of human iPSC-CMs. Mutant *TBX2* iPSC-CMs show hallmarks of pathological hypertrophy at the cellular and molecular levels. In addition, whole genome transcriptome analysis shows upregulation of cardiomyocyte stress genes as well as of cardiomyocyte differentiation gene sets. These features are consistent with the described role of TBX2 as transcriptional repressor of the myocardial gene program: in the absence of WT TBX2 activity, tight regulation of cardiomyocyte differentiation might be disrupted to result in upregulation of cardiomyocyte gene expression, leading to hypertrophy and stress phenotype.

The actions of TBX2 in cardiac development have been well characterised in the context of AVC and OFT formation, where it represses the myocardial gene program (Aanhaanen et al., 2011; Christoffels et al., 2004; Harrelson et al., 2004; Singh et al., 2012). However, *TBX2* is broadly expressed in the linear heart tube before becoming restricted to these specialised tissues, indicating a potential for a broader role in early cardiac development outside of the OFT and AVC. In mouse *TBX2* KO models AVC and OFT development is defective, whilst there was no effect on the development of myocardium (Harrelson et al., 2004; Singh et al., 2012). It is possible that the apparent discrepancy is caused by species-specific difference or by the difference between in vivo and in vitro models. The Wnt modulation method of cardiac differentiation used in this study primarily produces ventricular-like cardiomyocytes, with

some atrial and node-like cells also present (Lian et al. 2013; Burridge et al. 2014b; Galdos et al. 2022). With differentiation protocols for the derivation of additional cardiomyocytes subtypes such as atrial, node-like, or AVC and conduction-like (Cyganek et al., 2018; Lyra-Leite et al., 2022; Prodan et al., 2022; Wiesinger et al., 2022; Ye et al., 2024) it will be possible to more precisely dissect any cardiac cell type-specific roles of human TBX2.

Loss of *PRDM1* expression does not have a major impact on the outcome of cardiac differentiation in human iPSC-CMs, as efficiency of differentiation and cardiomyocyte morphology are unchanged; instead, it is characterised by subtle differences in beating onset. In the absence of *PRDM1,* beating is accelerated by on average 2 days. In mutant cells, most differentially expressed genes are upregulated at Day 3 of differentiation and include genes associated with neural development. There are multiple examples of PRDM1 repressing differentiation in other developmental contexts, for example in mammary luminal and sebaceous gland stem cells, in migrating primordial germ cells, and enterocyte maturation (Ahmed et al., 2016; Harper et al., 2011; Horsley et al., 2006; Yamashiro et al., 2016). Therefore, in this scenario it is possible that the absence of PRDM1 leads to earlier cardiomyocyte beating through a lack of repressive activity, which might be required for tightly regulated progression through iPSC-CM differentiation.

In conclusion, we have established a conserved cardiac regulatory module of GATA4 and its target genes *PRDM1* and *TBX2*, identified in *Xenopus*, which also operates in the development of human cardiomyocytes.

Limitations

A clear limitation of the current study is the lack of high-resolution information on cell fate trajectories in GATA4 -/- iPSC-CM cultures, which will be essential for better insight into the roles of GATA4 in specification and differentiation of cardiovascular cell types.

Materials and Methods

Xenopus embryos and explants

All work with *Xenopus laevis* was approved by Cardiff University's Animal Welfare and Ethical Review Board and was undertaken under a license from the UK Home Office. *Xenopus laevis*

embryos were obtained by mating of frogs primed with human chorionic gonadotrophin (Sigma) or by in vitro fertilisation (Sive, 2000). Jelly membrane was removed with 2% cysteine-HCl, pH 7.8 (Sive, 2000). Embryos were grown in 10% Normal Amphibian Media (NAM) and staged as described (Sive, 2000). Whole embryos (WE) or explants were cultured until age match control siblings had reached desired stage. Microinjections were carried using an IM 300 Micro-injector (Narishige Scientific), in 75% NAM containing 3% Ficoll (Sigma). Morpholino Oligonucleotides (MOs) were supplied from Gene Tools (http://www.gene-tools.com/) and injected at 10 nL/embryo. Tbx2 MO was described (Cho et al., 2011). GATA4 and GATA5 translation- and -splicing blocking MOs were described (Haworth et al., 2008). mMESSAGE mMACHINE kit (Ambion) was used for capped mRNA synthesis. Previously described templates for making capped mRNA: GATA4-GR (Afouda et al., 2005; Latinkić et al., 2003), Gata4 (Gallagher et al., 2012), *Xenopus* n*odal5* (Takahashi et al., 2000). *Xenopus prdm1* (de Souza et al., 1999) was obtained from the European *Xenopus* Resource Centre (Portsmouth, UK), and coding sequence was amplified by PCR and cloned into BamHI-XbaI sites of pCS2+ by PCR. WT and mutant human TBX2 CDS was cloned into pCS2+ using PCR. Correct sequence was confirmed by sequencing. Injection solutions included lineage tracers biotin- and rhodamine-dextran (Invitrogen; (Latinkić et al., 2003)). For cardiogenic Nodal signalling we injected 50-110 pg of *nodal5* mRNA and for lower, non-cardiogenic Nodal signalling we used 16 units of activin protein provided as XTC cell line conditioned media (Smith et al., 1988). Cardiogenic nodal5 and non-cardiogenic but muscle-inducing activin doses were confirmed (Fig. S1A).

Chromatin Immunoprecipitation

Chromatin Immunoprecipitation was performed as described by Blythe et al. (Blythe et al., 2009) with minor changes: 2-300 embryos were injected at 2-cell stage with 0.4 ng of capped mRNA coding for HA tagged rat GATA4 (Gallagher et al., 2012). Stage 9-9.5 embryos were fixed in 50 mL falcon tube for 30min in 25 mL of 10% NAM, 1% formaldehyde on ice without agitation, briefly washed in 10% NAM and quenched for 10 min. in 25 mL 10% NAM, 0.125 M glycine followed by 3 washes in 10% NAM. Embryos were transferred into 1.5 mL microcentrifuge tubes (50/tube) and frozen on dry ice before storage at -80 ^{0}C.
The crosslinked 50 embryos were thawed on ice for 15 min and homogenised in 600 mL of RIPA buff + Protease Inhibitor tablets (complete EDTA-free, Roche) and incubated for 15

minutes beginning from the time of initial homogenization. Homogenates were centrifuged for 5 min 3000 rpm, 4^0C. Supernatant was removed and the inside wall of the tube above the pellet wiped with fine paper to remove lipid residue. The pellet was resuspended in 400 mL of RIPA buff +PI to get ~450 mL. Sonication was performed with Soniprep 150 at 70% power: 8 cycles (10 sec. on – 50 sec. off) in 2 mL round bottom Eppendorf tubes on ice water bath. Shear median DNA size is ~500 bp. The volume was adjusted to 650 mL with RIPA+PI and samples centrifuged for 10 min at 13000g, 4^0C. Supernatant: 600 mL taken for IP, 5 mL taken for the INPUT, 20 mL for shear control. Anti-HA antibody magnetic beads were used according to manufacturer recommendations (Pierce). 50 µL of beads per 50 embryos, prewashed 3x 5min in TBS + 0.01% BSA using a magnet stand. Supernatant 600 mL taken for IP mixed with prewashed beads and incubated overnight rotating at 4^0C. Supernatants were removed and beads stored at -20^0C. Beads were washed 3x for 5 min in TBS-Tw (0.01% Tw20). Reversing crosslinks: beads (with chromatin), input and shear control samples were supplemented with 200 mL of TE buffer + 1/20 vol. of 5M NaCl and incubated overnight at 65^0C. 1 mL RNAse A 10mg/mL was added to each sample and incubated for 30 min at 37^0 C. + 1 mL Proteinase K (Roche), incubation 1h at 55^0C. Proteins were removed by phenol-chloroform extraction and DNA precipitated with isopropanol. Samples were dissolved in 50 mL of ultrapure distilled water. PCR reactions were performed as for RT-PCR (below). Amplicons were chosen within 10kb upstream and 2kb downstream of translation start site based on presence of at least one consensus GATA binding site

RNA extraction and RT-PCR analysis

RNA was extracted using the acid-guanidinium thiocyanate phenol chloroform method (Chomczynski and Sacchi, 1987) and cDNA was synthesised using the RevertAid reverse transcriptase kit (ThermoFisher). All primer sequences used were designed using NCBI's primer blast program (Ye et al., 2012) and are listed in Supplemental Table 4. Samples were normalised using *odc1* primers for *Xenopus* samples and *GAPDH* primers for human samples. To find the appropriate amount of cDNA to use in further PCR reactions. PCR reactions were conducted in a volume of 25 µL, to each was added 0.1 µM of forward and reverse primer (Eurofins), 1X MyTaq red reaction buffer (Bioline), 0.5U KAPA Taq Polymerase (both KAPA Biosytems), and the appropriate volume of cDNA. An MJ Mini Thermal Cycler Machine

(Biorad) was used for cycling. The amplicons generated were visualized using agarose gel electrophoresis. All primer sequences are presented in Supplemental Table 4.

Whole mount in situ hybridisation to detect expression of *myl7* in *Xenopus* embryos was performed as described (Gallagher et al., 2012).

Human iPS cell culture and gene editing

All cell lines produced in this study were derived from Rebl Pat iPSCs kindly provided by Prof. Chris Denning, Nottingham University, and described in (Hammad et al., 2016).

iPSC culture

iPSCs were cultured on Corning™ cell culture plates coated with Geltrex™ (ThermoFisher) at 30-45 µg/mL in DMEM/F12 Media (ThermoFisher) and maintained with B8 media formulated as described by (Kuo et al., 2020) . Passaging was carried out when cultures reached 75-80% confluency. ReLeSR™ was used to dissociate iPSC colonies as per the manufacturer's instructions (StemCell Technologies), following this the cells were replated and fed with B8 2.5 µM Y27632 (HelloBio) for 24 hours following passaging before being switched to B8 for regular feeding every other day. All cells were maintained at 37°C in a humidified atmosphere with 5% CO_2.

CRISPR-Cas9 gene editing of iPSCs

All crRNAs were designed as pairs using CCTop (Stemmer et al., 2015). Each pair was intended to be used together to induce a frameshifting mutation. To produce gRNAs each crRNA (Integrated DNA Technologies) was combined with tracrRNA (Integrated DNA Technologies) at an equimolar concentration in IDTE Nuclease free buffer (30 mM HEPES (pH 7.5), 100 mM potassium acetate), then heated to 95°C for 2 minutes followed by gradual cooling to RT. Each gRNA was then combined with an equal volume of 6.2 µg/µL Alt-R S.p. HiFi Cas9 Nuclease V3 diluted in Cas9 storage buffer (10 mM Tris-HCl (pH 7.4), 300 mM NaCl, 0.1 mM EDTA, 1 mM DTT) and incubated together for 10 minutes at RT to form a ribonucleoprotein complex (RNP). Transfection of the RNPs into iPSCs was carried out in suspension using the P3 Primary Cell 4D-Nucleofector™ X Kit L (Lonza) and a Lonza 4D Nucleofector with program CA137. Single clones were selected and screened using gDNA

PCR followed by gel electrophoresis. All mutations were confirmed using Sanger sequencing (Eurofins) and SyntheGo ICE (Conant et al., 2022) signal deconvolution. Sequences of crRNAs and genotyping PCR primers are provided in Supplemental Table 4. Following gene editing the pluripotency of the lines was assessed by confirming expression of *NANOG*, *POU5F1* and *SOX2* (Fig. S4, S5, S7, and S10).

Karyotyping

Genomic DNA from the parental Rebl Pat human iPSC line and gene edited lines derived from it was extracted from >2 million cells using the Blood and Tissue DNA Extraction Kit (Qiagen), following the manufacturers guidance. The gDNA was karyotyped using the Infinium Global Screening Array-24, covering 654,027 markers across the genome at the MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University. The results are summarised in Supplemental Table 5.

CM differentiation and maintenance

Differentiation of iPSCs into CMs was carried out using two commonly utilised Wnt modulation protocols (Burridge et al., 2014; Lian et al., 2012). For differentiation cells were seeded at a density of 14 x 10^4 – 17.5 x 10^4 cells per cm^2, in wells coated with Geltrex™ as above and expanded to 75-80% confluency in B8 medium before differentiation was initiated. For the CDM3 protocol published by Burridge *et al.* the cells were changed to CDM3 (RPMI 1640, 500 µg/mL, 213 µg/mL L-Ascorbic acid-2-phophate, Penicillin – Streptomycin 50 U/mL) with 6 µM CHIR99021 (HelloBio) for 48 hours. Following this media was changed to CDM3 with Wnt 2 µM WntC59 (HelloBio) for 48 hours. After these treatment stages cells were maintained in CDM3 media and refreshed every other day. Then switched to CDM3-L (RPMI 1640 no glucose, 500 µg/ml Fraction V BSA, 213 µg/ml L-Ascorbic acid-2-phosphate, 5 mM Sodium DL-Lactate) from days 10-14 to metabolically enrich the cultures for CMs.

For the GiWi protocol by Lian *et al.* the cells were plated as above and changed to RPMI $B27^{-insulin}$ (ThermoFisher) with 6 µM CHIR99021 for 24 hours to initiate differentiation. This media was then removed and replaced with RPMI $B27^{-insulin}$ for 48 hours. Following this the spent media was removed and mixed 1:1 with fresh RPMI $B27^{-insulin}$ to make 'conditioned media'.

To the conditioned media was added 2 µM WntC59 and the cells were maintained in this for a further 48 hours. After this the media was refreshed every 2 days with RMPI B27$^{-insulin}$, and then with RPMI B27 from day 8-10. For cells differentiated using the GiWi protocol RPMI B27-L (RPMI 1640 no glucose, 1x B27, 5 mM Sodium D-L-Lactate) was used to metabolically select cells from day 10-14.

Upon finishing selection CMs derived using the CDM3 or GiWi protocols were maintained in CDM3 with media changes every 2 days.

CM dissociation

The cell layer was washed with Hanks Buffered Saline Solution without Ca and Mg (HBSS, Merck). This was then removed and pre-warmed dissociation solution (HBSS, 200 U/mL Collagen Type II (Worthington Biotech), 1 mM HEPES pH 7.4, 10 µM Y37632 (HelloBio)) was added and the cells incubated at 37 ℃ for 3-3.5 hours. The cell solution was then gently triturated combined with an equal volume of media and then filtered through a 100 µm filter. The cells in suspension were spun down at 400 rcf for 5 minutes, and the supernatant removed. The cell pellets were resuspended at an appropriate concentration in CDM3 with 10% heat inactivated FBS for re-plating. The following day the media was changed to CDM3.

Immunofluorescence

Any cells to be analysed were dissociated as described above and then plated onto 13 mm glass coverslips coated with 30-45 µg/mL Geltrex™. The cells were allowed to attach and recover for a minimum of 48 hours before any further processing. The fixation and permeabilization method used was dependent on the antigen being detected (see Supplemental Table 3): either being carried out using 4% PFA for 10 minutes at RT, followed by permeabilization with 0.25% Triton-X-100 DPBS $^{-Ca\ -Mg}$ for 10 minutes at RT, or using 100% ice cold methanol for 15 minutes at 4°C. Regardless of fixation and permeabilization method used blocking was carried out with DPBS $^{-Ca\ -Mg}$ 10% serum for 1-hour at RT, with serum type matched to the secondary antibody used. Primary antibody incubation was carried out overnight at 4°C in blocking buffer, see Supplemental Table 3 for dilutions. The samples were then washed 5 times for 15-minutes at RT with PBS- 0.1% Triton-X100, then incubated with

1:1000 Goat Anti-Mouse or anti-rabbit Alexa fluor conjugated antibodies (ThermoFisher) in blocking buffer overnight.

Hoechst 33342 at 1 µg/mL was used to counterstain nuclei by incubation for 15 minutes, followed by 3 five-minute washes in PBS $^{-Ca2+\ Mg2+.}$ This was followed by mounting in Vectashield hard set mounting media with Phalloidin TRITC. Alternatively, HCS Cell Mask Orange (Invitrogen) was used at 1x to stain cell membranes and nuclei for 30 minutes, followed by three five-minute washes PBS $^{-Ca\ -Mg}$. Mounting was then carried out using Vectashield with DAPI (Vector Laboratories) to stain nuclei. A Zeiss LSM880 Confocal Microscope was used for imaging and FIJI/Image J Software (Schneider et al., 2012) for image analysis and editing. Following IF staining and imaging the TNNT2+ cells were analysed against criteria adapted from (Ang et al., 2016).

Calcium Imaging

Any cells to be analysed were dissociated as described above and then plated onto 13 mm glass coverslips coated with 30-45 µg/mL Geltrex™. The cells were allowed to attach and recover for a minimum of 48 hours before any further processing. For analysis of the TBX2 lines and their WT counterparts' cells were loaded with the Ca^{2+}-sensitive indicator Fluo-4-AM (5 µmol/L) for 30 min at room temperature, followed by at least 10 min for de-esterification. Experiments were performed in a modified Tyrode's solution containing, in mM: 133 NaCl, 5 KCl, 1 NaH_2PO_4, 10 4-(2-hydroxyethyl) -1piperazineethanesulfonic acid (HEPES), 10 glucose, 1.8 $CaCl_2$, 1 $MgCl_2$), pH 7.4 with NaOH. Confocal line-scanning microscopy was performed using an inverted Leica SP5 confocal microscope with a 63x 1.2 NA water immersion objective. Spontaneous Ca^{++} transients were recorded using the 488 nm line of an argon laser in line-scanning mode at 400 lines per second, with fluorescence emission collected between 500-650 nm. The confocal pinhole set to <2 Airy units. For the GATA4 lines and WT controls imaging was conducted using the widefield setting on the Olympus IX71 microscope. Before imaging the cells were loaded with 1 µM Fluo8-AM or 1 µM Fluo4-AM in cell culture media for 30 minutes at 37 ̊C. This media was then removed and Tyrode's solution with 1mM HEPES pH 7.4 (Merck, T2397) was added. The cells were then imaged using a 488 nm laser. For both, temperature was maintained at 37 °C during Ca^{++} imaging experiments using a heated microscope enclosure. The change in fluorescence over time was plotted using FIJI/Image J Software (Schneider et al., 2012), and Ca^{++} transient

analysis was performed using custom written MATLAB scripts. As part of this analysis non-cell background fluorescence was subtracted from any readings, then average cell fluorescence (F) was divided by the diastolic fluorescence (F_0) to account for possible variability in dye loading. This was used to calculate an average for Ca^{++} transient amplitude value (reported as F/F_0) and full duration at half maximum (FDHM).

RNA-seq

The *Xenopus* RNA samples were converted to cDNA libraries using a Bioo NextFlex directional v.2 kit and were sequenced at Bristol University's sequencing facility. The human RNA samples were converted to cDNA libraries using Illumina Truseq stranded mRNA library kit and sequenced at the Genome Research Hub, School of Biosciences, Cardiff University.

RNA-seq Quantification and Differential expression

Xenopus laevis sequence analysis*:* paired end reads were aligned to *X. laevis* gene models v7.2 by JGI with Bowtie (Langmead et al., 2009) allowing for up to 3 mismatches in seed and leaving the other parameters at the default value. Gene hits were counted using RSEM (Li and Dewey, 2011)with default parameters. Expected read counts reported by RSEM were used to call differentially expressed genes which were called using DESeq (Anders and Huber, 2010). The dispersion was estimated with "per-condition" method and "gene-est-only" sharing mode. Each treatment was separately compared to control at each timepoint, genes with padj < 0.1 were defined as differentially expressed. To determine a gene's differential expression irrespective of timepoint, we took the timepoint with the minimum padj.

Human sequence analysis: RNA-seq data (75bp single-end) data generated from Human iPSC cell lines (40-45 M per sample) were quality trimmed using fastp (Chen, 2023) and trimmed sequences aligned against the annotated Ensembl GRCh38.p14 human genome using STAR (Dobin et al., 2013). Quantification against gene objects was performed by using Rsubread (Liao et al., 2019) and differentially expressed genes (DEGs) derived using DESeq2 using the SARTools wrapper using default parameters (Varet et al., 2016). As with *Xenopus* data, each time point was compared to time and batch-matched controls and genes with p-adj <0.1 were defined as differentially expressed.

Gene set enrichment

Xenopus laevis: to assess the composition of each group of genes we performed gene set enrichment using Enrichr. We took genes from each cluster with a known *Xenopus laevis* gene symbol and converted these to human symbols, by removing any ".N," "-a," or "-b" suffix for an integer N and converting to uppercase. We then made the following substitutions to convert certain known *Xenopus* gene symbols to human where the name of the ortholog has diverged or only a paralog exists: pou5f3 → POU5F1, mix1 → mixl1, dppa2 → DPPA4, lefty → lefty2, ventx1-3 → NANOG, mespb → MESP1, sox17a/b → SOX17. We remove any duplicate names that arose in this process. Gene nomenclature for *Xenopus laevis* we used is described at xenbase.org (Fisher et al., 2023) . For simplicity, in figures we omitted the information on homeologues (.L or .S) with the exception of *lefty1.S*, which was validated as DEG, unlike *lefty1.L*. We calculated enrichments for the following gene sets: KEGG_2021_Human, WikiPathway_2023_Human, BioPlanet_2019, GO_Biological_Process_2023, GO_Molecular_Function_2023, GO_Cellular_Component_2023. All enrichments are found in Supplemental Table 1. Human Gene Ontology enrichment analysis was carried out using ShinyGO v0.77 and STRING v11.5 for network analysis (Ge et al., 2020; Szklarczyk et al., 2019) and is presented in Supplemental Table 2. Normalised counts for DEGs identified in human iPSC samples are presented in Supplemental Table 3.

Western blotting

Cells were lysed in RIPA Buffer (50 mM Tris-HCl pH8, 150 mM NaCl, 0.1 % SDS, 0.5% Na Deoxycholate, 1 % NP-40, and 1x Protease Inhibitor Complete Mini (Roche)). Total protein extracts were mixed with 2x Laemmli Loading Buffer (BioRad) and boiled for 10 minutes before loading, 20 µg of total protein was used per well. Bands were resolved via polyacrylamide gel electrophoresis; the samples were then transferred to polyvinylidene difluoride membrane (Millipore). 5% Skimmed Milk in TBS Tween 20 (TBS-Tw: 5 mM Tris-Base pH 7.4, 20 mM NaCl, 0.1% Tween 20) was used to block the membranes for 1 hour. All antibodies used are described in Supplemental Table 4. For detection of TBX2 skimmed milk concentration was reduced to 1 % to improve signal intensity. Incubation with primary antibodies was carried out overnight in blocking buffer at 4°C with constant rotation. Following this the membranes were washed with TBS-Tw 3 times for 10 minutes each time. Secondary antibodies were then applied to the membranes diluted in 5% skimmed milk TBS-

Tw or for detection of TBX2 1% skimmed milk TBS-Tw. The membranes were 3 times for 10 minutes each time in TBS-Tw to remove excess antibody. Clarity Western ECL (Biorad) was used for development followed by exposure of the membranes to Amersham Hyperfilm (Merck).

Data Availability

The sequence data will be made publicly available in NCBI's Gene Expression Omnibus through GEO Series accession numbers GSE308990 (human) and GSE308764 (*Xenopus*) as of the date of publication.

Acknowledgements

This work was supported by grants from British Heart Foundation (PG/11/115/29287 and FS/18/42/33827) and from Heart Research Wales to BL. EDF is supported by a British Heart Foundation Intermediate Basic Science Research Fellowship (FS/IBSRF/21/25071). NO is supported by a Wellcome Career Development Award (227357/Z/23/Z). We are grateful to Prof. Chris Denning and his group (Nottingham University) for generously providing training in iPSC-CM techniques and for providing iPSC lines. We thank Prof. Nick Allen and his group for local support for iPSC culture work. Expert technical support by Alexandra Evans (MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University) and Angela Marchbank (Genome Research Hub, School of Biosciences, Cardiff University) is gratefully acknowledged.

Aanhaanen WTJ, Boukens BJD, Sizarov A, Wakker V, de Gier-de Vries C, van Ginneken AC, Moorman AFM, Coronel R, Christoffels VM. 2011. Defective Tbx2-dependent patterning of the atrioventricular canal myocardium causes accessory pathway formation in mice. *J Clin Invest* **121**:534–544. doi:10.1172/JCI44350

Afouda BA, Ciau-Uitz A, Patient R. 2005. GATA4, 5 and 6 mediate TGFbeta maintenance of endodermal gene expression in Xenopus embryos. *Development* **132**:763–774. doi:10.1242/dev.01647

Ahmed MI, Elias S, Mould AW, Bikoff EK, Robertson EJ. 2016. The transcriptional repressor Blimp1 is expressed in rare luminal progenitors and is essential for mammary gland development. *Development* **143**:1663–1673. doi:10.1242/dev.136358

Akerberg BN, Gu F, VanDusen NJ, Zhang X, Dong R, Li K, Zhang B, Zhou B, Sethi I, Ma Q, Wasson L, Wen T, Liu J, Dong K, Conlon FL, Zhou J, Yuan G-C, Zhou P, Pu WT. 2019. A reference map of murine cardiac transcription factor chromatin occupancy identifies dynamic and conserved enhancers. *Nat Commun* **10**:4907. doi:10.1038/s41467-019-12812-3

Akerberg BN, Pu WT. 2020. Genetic and epigenetic control of heart development. *Cold Spring Harb Perspect Biol* **12**. doi:10.1101/cshperspect.a036756

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biol* **11**:R106. doi:10.1186/gb-2010-11-10-r106

Ang Y-S, Rivas RN, Ribeiro AJS, Srivas R, Rivera J, Stone NR, Pratt K, Mohamed TMA, Fu J-D, Spencer CI, Tippens ND, Li M, Narasimha A, Radzinsky E, Moon-Grady AJ, Yu H, Pruitt BL, Snyder MP, Srivastava D. 2016. Disease model of GATA4 mutation reveals transcription factor cooperativity in human cardiogenesis. *Cell* **167**:1734-1749.e22. doi:10.1016/j.cell.2016.11.033

Arroyo N, Villamayor L, Díaz I, Carmona R, Ramos-Rodríguez M, Muñoz-Chápuli R, Pasquali L, Toscano MG, Martín F, Cano DA, Rojas A. 2021. GATA4 induces liver fibrosis regression by deactivating hepatic stellate cells. *JCI Insight* **6**. doi:10.1172/jci.insight.150059

Blythe SA, Reid CD, Kessler DS, Klein PS. 2009. Chromatin immunoprecipitation in early Xenopus laevis embryos. *Dev Dyn* **238**:1422–1432. doi:10.1002/dvdy.21931

Bobowski-Gerard M, Boulet C, Zummo FP, Dubois-Chevalier J, Gheeraert C, Bou Saleh M, Strub J-M, Farce A, Ploton M, Guille L, Vandel J, Bongiovanni A, Very N, Woitrain E, Deprince A, Lalloyer F, Bauge E, Ferri L, Ntandja-Wandji L-C, Cotte AK, Grangette C, Vallez E, Cianférani S, Raverdy V, Caiazzo R, Gnemmi V, Leteurtre E, Pourcet B, Paumelle R, Ravnskjaer K, Lassailly G, Haas JT, Mathurin P, Pattou F, Dubuquoy L, Staels B, Lefebvre P, Eeckhoute J. 2022. Functional genomics uncovers the transcription factor BNC2 as required for myofibroblastic activation in fibrosis. *Nat Commun* **13**:5324. doi:10.1038/s41467-022-33063-9

Borok MJ, Papaioannou VE, Sussel L. 2016. Unique functions of Gata4 in mouse liver induction and heart development. *Dev Biol* **410**:213–222. doi:10.1016/j.ydbio.2015.12.007

Bruneau BG, Nemer G, Schmitt JP, Charron F, Robitaille L, Caron S, Conner DA, Gessler M, Nemer M, Seidman CE, Seidman JG. 2001. A murine model of Holt-Oram syndrome defines roles of the T-box transcription factor Tbx5 in cardiogenesis and disease. *Cell* **106**:709–721. doi:10.1016/s0092-8674(01)00493-7

Bruneau BG. 2013. Signaling and transcriptional networks in heart development and regeneration. *Cold Spring Harb Perspect Biol* **5**:a008292. doi:10.1101/cshperspect.a008292

Burridge PW, Matsa E, Shukla P, Lin ZC, Churko JM, Ebert AD, Lan F, Diecke S, Huber B, Mordwinkin NM, Plews JR, Abilez OJ, Cui B, Gold JD, Wu JC. 2014. Chemically defined generation of human cardiomyocytes. *Nat Methods* **11**:855–860. doi:10.1038/nmeth.2999

Cheng AM, Thisse B, Thisse C, Wright CV. 2000. The lefty-related factor Xatv acts as a feedback inhibitor of nodal signaling in mesoderm induction and L-R axis development in xenopus. *Development* **127**:1049–1061. doi:10.1242/dev.127.5.1049

Chen S. 2023. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *iMeta* **2**:e107. doi:10.1002/imt2.107

Chomczynski P, Sacchi N. 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* **162**:156–159. doi:10.1006/abio.1987.9999

Cho G-S, Choi S-C, Park EC, Han J-K. 2011. Role of Tbx2 in defining the territory of the pronephric nephron. *Development* **138**:465–474. doi:10.1242/dev.061234

Christoffels VM, Hoogaars WMH, Tessari A, Clout DEW, Moorman AFM, Campione M. 2004. T-box transcription factor Tbx2 represses differentiation and formation of the cardiac chambers. *Dev Dyn* **229**:763–770. doi:10.1002/dvdy.10487

Churko JM, Garg P, Treutlein B, Venkatasubramanian M, Wu H, Lee J, Wessells QN, Chen S-Y, Chen W-Y, Chetal K, Mantalas G, Neff N, Jabart E, Sharma A, Nolan GP, Salomonis N, Wu JC. 2018. Defining human cardiac transcription factor hierarchies using integrated single-cell heterogeneity analysis. *Nat Commun* **9**:4906. doi:10.1038/s41467-018-07333-4

Conant D, Hsiau T, Rossi N, Oki J, Maures T, Waite K, Yang J, Joshi S, Kelso R, Holden K, Enzmann BL, Stoner R. 2022. Inference of CRISPR Edits from Sanger Trace Data. *The CRISPR Journal* **5**:123–130. doi:10.1089/crispr.2021.0113

Crispino JD, Lodish MB, Thurberg BL, Litovsky SH, Collins T, Molkentin JD, Orkin SH. 2001. Proper coronary vascular development and heart morphogenesis depend on interaction of GATA-4 with FOG cofactors. *Genes Dev* **15**:839–844. doi:10.1101/gad.875201

Cunningham TJ, Yu MS, McKeithan WL, Spiering S, Carrette F, Huang C-T, Bushway PJ, Tierney M, Albini S, Giacca M, Mano M, Puri PL, Sacco A, Ruiz-Lozano P, Riou J-F, Umbhauer M, Duester G, Mercola M, Colas AR. 2017. Id genes are essential for early heart formation. *Genes Dev* **31**:1325–1338. doi:10.1101/gad.300400.117

Cyganek L, Tiburcy M, Sekeres K, Gerstenberg K, Bohnenberger H, Lenz C, Henze S, Stauske M, Salinas G, Zimmermann W-H, Hasenfuss G, Guan K. 2018. Deep phenotyping of human

induced pluripotent stem cell-derived atrial and ventricular cardiomyocytes. *JCI Insight* **3**. doi:10.1172/jci.insight.99941

de Souza FS, Gawantka V, Gómez AP, Delius H, Ang SL, Niehrs C. 1999. The zinc finger gene Xblimp1 controls anterior endomesodermal cell fate in Spemann's organizer. *EMBO J* **18**:6062–6072. doi:10.1093/emboj/18.21.6062

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**:15–21. doi:10.1093/bioinformatics/bts635

Fahed AC, Roberts AE, Mital S, Lakdawala NK. 2014. Heart failure in congenital heart disease: a confluence of acquired and congenital. *Heart Fail Clin* **10**:219–227. doi:10.1016/j.hfc.2013.09.017

Firulli AB, McFadden DG, Lin Q, Srivastava D, Olson EN. 1998. Heart and extra-embryonic mesodermal defects in mouse embryos lacking the bHLH transcription factor Hand1. *Nat Genet* **18**:266–270. doi:10.1038/ng0398-266

Fisher M, James-Zorn C, Ponferrada V, Bell AJ, Sundararaj N, Segerdell E, Chaturvedi P, Bayyari N, Chu S, Pells T, Lotay V, Agalakov S, Wang DZ, Arshinoff BI, Foley S, Karimi K, Vize PD, Zorn AM. 2023. Xenbase: key features and resources of the Xenopus model organism knowledgebase. *Genetics* **224**. doi:10.1093/genetics/iyad018

Gallagher JM, Komati H, Roy E, Nemer M, Latinkić BV. 2012. Dissociation of cardiogenic and postnatal myocardial activities of GATA4. *Mol Cell Biol* **32**:2214–2223. doi:10.1128/MCB.00218-12

Garg V, Kathiriya IS, Barnes R, Schluterman MK, King IN, Butler CA, Rothrock CR, Eapen RS, Hirayama-Yamada K, Joo K, Matsuoka R, Cohen JC, Srivastava D. 2003. GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5. *Nature* **424**:443–447. doi:10.1038/nature01827

Garry GA, Bassel-Duby R, Olson EN. 2022. Direct reprogramming as a route to cardiac repair. *Semin Cell Dev Biol* **122**:3–13. doi:10.1016/j.semcdb.2021.05.019

Ge SX, Jung D, Yao R. 2020. ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics* **36**:2628–2629. doi:10.1093/bioinformatics/btz931

Gonzalez-Teran B, Pittman M, Felix F, Thomas R, Richmond-Buccola D, Hüttenhain R, Choudhary K, Moroni E, Costa MW, Huang Y, Padmanabhan A, Alexanian M, Lee CY, Maven BEJ, Samse-Knapp K, Morton SU, McGregor M, Gifford CA, Seidman JG, Seidman CE, Gelb BD, Colombo G, Conklin BR, Black BL, Bruneau BG, Krogan NJ, Pollard KS, Srivastava D. 2022. Transcription factor protein interactomes reveal genetic determinants in heart disease. *Cell* **185**:794-814.e30. doi:10.1016/j.cell.2022.01.021

Grancharova T, Gerbin KA, Rosenberg AB, Roco CM, Arakaki JE, DeLizo CM, Dinh SQ, Donovan-Maiye RM, Hirano M, Nelson AM, Tang J, Theriot JA, Yan C, Menon V, Palecek

SP, Seelig G, Gunawardane RN. 2021. A comprehensive analysis of gene expression changes in a high replicate and open-source dataset of differentiating hiPSC-derived cardiomyocytes. *Sci Rep* **11**:15845. doi:10.1038/s41598-021-94732-1

Habets PEMH, Moorman AFM, Clout DEW, van Roon MA, Lingbeek M, van Lohuizen M, Campione M, Christoffels VM. 2002. Cooperative action of Tbx2 and Nkx2.5 inhibits ANF expression in the atrioventricular canal: implications for cardiac chamber formation. *Genes Dev* **16**:1234–1246. doi:10.1101/gad.222902

Hammad M, Rao W, Smith JGW, Anderson DG, Langer R, Young LE, Barrett DA, Davies MC, Denning C, Alexander MR. 2016. Identification of polymer surface adsorbed proteins implicated in pluripotent human embryonic stem cell expansion. *Biomater Sci* **4**:1381–1391. doi:10.1039/c6bm00214e

Harper J, Mould A, Andrews RM, Bikoff EK, Robertson EJ. 2011. The transcriptional repressor Blimp1/Prdm1 regulates postnatal reprogramming of intestinal enterocytes. *Proc Natl Acad Sci USA* **108**:10585–10590. doi:10.1073/pnas.1105852108

Harrelson Z, Kelly RG, Goldin SN, Gibson-Brown JJ, Bollag RJ, Silver LM, Papaioannou VE. 2004. Tbx2 is essential for patterning the atrioventricular canal and for morphogenesis of the outflow tract during heart development. *Development* **131**:5041–5052. doi:10.1242/dev.01378

Haworth KE, Kotecha S, Mohun TJ, Latinkic BV. 2008. GATA4 and GATA5 are essential for heart and liver development in Xenopus embryos. *BMC Dev Biol* **8**:74. doi:10.1186/1471-213X-8-74

Holtzinger A, Evans T. 2005. Gata4 regulates the formation of multiple organs. *Development* **132**:4005–4014. doi:10.1242/dev.01978

Horsley V, O'Carroll D, Tooze R, Ohinata Y, Saitou M, Obukhanych T, Nussenzweig M, Tarakhovsky A, Fuchs E. 2006. Blimp1 defines a progenitor population that governs cellular input to the sebaceous gland. *Cell* **126**:597–609. doi:10.1016/j.cell.2006.06.048

Jiang CL, Goyal Y, Jain N, Wang Q, Truitt RE, Coté AJ, Emert B, Mellis IA, Kiani K, Yang W, Jain R, Raj A. 2022. Cell type determination for cardiac differentiation occurs soon after seeding of human-induced pluripotent stem cells. *Genome Biol* **23**:90. doi:10.1186/s13059-022-02654-6

Jiang Y, Drysdale TA, Evans T. 1999. A role for GATA-4/5/6 in the regulation of Nkx2.5 expression with implications for patterning of the precardiac field. *Dev Biol* **216**:57–71. doi:10.1006/dbio.1999.9469

Kilpinen H, Goncalves A, Leha A, Afzal V, Alasoo K, Ashford S, Bala S, Bensaddek D, Casale FP, Culley OJ, Danecek P, Faulconbridge A, Harrison PW, Kathuria A, McCarthy D, McCarthy SA, Meleckyte R, Memari Y, Moens N, Soares F, Mann A, Streeter I, Agu CA, Alderton A, Nelson R, Harper S, Patel M, White A, Patel SR, Clarke L, Halai R, Kirton CM, Kolb-Kokocinski A, Beales P, Birney E, Danovi D, Lamond AI, Ouwehand WH, Vallier L, Watt FM,

Durbin R, Stegle O, Gaffney DJ. 2017. Common genetic variation drives molecular heterogeneity in human iPSCs. *Nature* **546**:370–375. doi:10.1038/nature22403

Koutsourakis M, Langeveld A, Patient R, Beddington R, Grosveld F. 1999. The transcription factor GATA6 is essential for early extraembryonic development. *Development* **126**:723–732. doi:10.1242/dev.126.9.723

Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, McDermott MG, Monteiro CD, Gundersen GW, Ma'ayan A. 2016. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* **44**:W90-7. doi:10.1093/nar/gkw377

Kuo CT, Morrisey EE, Anandappa R, Sigrist K, Lu MM, Parmacek MS, Soudais C, Leiden JM. 1997. GATA4 transcription factor is required for ventral morphogenesis and heart tube formation. *Genes Dev* **11**:1048–1060. doi:10.1101/gad.11.8.1048

Kuo H-H, Gao X, DeKeyser J-M, Fetterman KA, Pinheiro EA, Weddle CJ, Fonoudi H, Orman MV, Romero-Tejeda M, Jouni M, Blancard M, Magdy T, Epting CL, George AL, Burridge PW. 2020. Negligible-Cost and Weekend-Free Chemically Defined Human iPSC Culture. *Stem Cell Reports* **14**:256–270. doi:10.1016/j.stemcr.2019.12.007

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**:R25. doi:10.1186/gb-2009-10-3-r25

Latinkić BV, Kotecha S, Mohun TJ. 2003. Induction of cardiomyocytes by GATA4 in Xenopus ectodermal explants. *Development* **130**:3865–3876. doi:10.1242/dev.00599

Lian X, Hsiao C, Wilson G, Zhu K, Hazeltine LB, Azarin SM, Raval KK, Zhang J, Kamp TJ, Palecek SP. 2012. Robust cardiomyocyte differentiation from human pluripotent stem cells via temporal modulation of canonical Wnt signaling. *Proc Natl Acad Sci USA* **109**:E1848-57. doi:10.1073/pnas.1200250109

Liao Y, Smyth GK, Shi W. 2019. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res* **47**:e47. doi:10.1093/nar/gkz114

Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**:323. doi:10.1186/1471-2105-12-323

Litviňuková M, Talavera-López C, Maatz H, Reichart D, Worth CL, Lindberg EL, Kanda M, Polanski K, Heinig M, Lee M, Nadelmann ER, Roberts K, Tuck L, Fasouli ES, DeLaughter DM, McDonough B, Wakimoto H, Gorham JM, Samari S, Mahbubani KT, Saeb-Parsy K, Patone G, Boyle JJ, Zhang Hongbo, Zhang Hao, Viveiros A, Oudit GY, Bayraktar OA, Seidman JG, Seidman CE, Noseda M, Hubner N, Teichmann SA. 2020. Cells of the adult human heart. *Nature* **588**:466–472. doi:10.1038/s41586-020-2797-4

Logan M, Mohun T. 1993. Induction of cardiac muscle differentiation in isolated animal pole explants of Xenopus laevis embryos. *Development* **118**:865–875. doi:10.1242/dev.118.3.865

López B, Ravassa S, Moreno MU, José GS, Beaumont J, González A, Díez J. 2021. Diffuse myocardial fibrosis: mechanisms, diagnosis and therapeutic approaches. *Nat Rev Cardiol* **18**:479–498. doi:10.1038/s41569-020-00504-1

Lyra-Leite DM, Gutiérrez-Gutiérrez Ó, Wang M, Zhou Y, Cyganek L, Burridge PW. 2022. A review of protocols for human iPSC culture, cardiac differentiation, subtype-specification, maturation, and direct reprogramming. *STAR Protocols* **3**:101560. doi:10.1016/j.xpro.2022.101560

Misra C, Sachan N, McNally CR, Koenig SN, Nichols HA, Guggilam A, Lucchesi PA, Pu WT, Srivastava D, Garg V. 2012. Congenital heart disease-causing Gata4 mutation displays functional deficits in vivo. *PLoS Genet* **8**:e1002690. doi:10.1371/journal.pgen.1002690

Molkentin JD, Lin Q, Duncan SA, Olson EN. 1997. Requirement of the transcription factor GATA4 for heart tube formation and ventral morphogenesis. *Genes Dev* **11**:1061–1072. doi:10.1101/gad.11.8.1061

Moschen AR, Fritz T, Clouston AD, Rebhan I, Bauhofer O, Barrie HD, Powell EE, Kim S-H, Dinarello CA, Bartenschlager R, Jonsson JR, Tilg H. 2011. Interleukin-32: a new proinflammatory cytokine involved in hepatitis C virus-related liver inflammation and fibrosis. *Hepatology* **53**:1819–1829. doi:10.1002/hep.24285

O'Reilly S. 2023. Interleukin-11 and its eminent role in tissue fibrosis: a possible therapeutic target. *Clin Exp Immunol* **214**:154–161. doi:10.1093/cei/uxad108

Olson EN. 2006. Gene regulatory networks in the evolution and development of the heart. *Science* **313**:1922–1927. doi:10.1126/science.1132292

Paige SL, Plonowska K, Xu A, Wu SM. 2015. Molecular regulation of cardiomyocyte differentiation. *Circ Res* **116**:341–353. doi:10.1161/CIRCRESAHA.116.302752

Papaioannou I, Dritsoula A, Kang P, Baliga RS, Trinder SL, Cook E, Shiwen X, Hobbs AJ, Denton CP, Abraham DJ, Ponticos M. 2024. NKX2-5 regulates vessel remodeling in scleroderma-associated pulmonary arterial hypertension. *JCI Insight* **9**. doi:10.1172/jci.insight.164191

Prodan N, Ershad F, Reyes-Alcaraz A, Li L, Mistretta B, Gonzalez L, Rao Z, Yu C, Gunaratne PH, Li N, Schwartz RJ, McConnell BK. 2022. Direct reprogramming of cardiomyocytes into cardiac Purkinje-like cells. *iScience* **25**:105402. doi:10.1016/j.isci.2022.105402

Robbe ZL, Shi W, Wasson LK, Scialdone AP, Wilczewski CM, Sheng X, Hepperla AJ, Akerberg BN, Pu WT, Cristea IM, Davis IJ, Conlon FL. 2022. CHD4 is recruited by GATA4 and NKX2-5 to repress noncardiac gene programs in the developing heart. *Genes Dev* **36**:468–482. doi:10.1101/gad.349154.121

Schneider CA, Rasband WS, Eliceiri KW. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9**:671–675. doi:10.1038/nmeth.2089

Shimazaki M, Nakamura K, Kii I, Kashima T, Amizuka N, Li M, Saito M, Fukuda K, Nishiyama T, Kitajima S, Saga Y, Fukayama M, Sata M, Kudo A. 2008. Periostin is essential for cardiac healing after acute myocardial infarction. *J Exp Med* **205**:295–303. doi:10.1084/jem.20071297

Singh R, Hoogaars WM, Barnett P, Grieskamp T, Rana MS, Buermans H, Farin HF, Petry M, Heallen T, Martin JF, Moorman AFM, ’t Hoen PAC, Kispert A, Christoffels VM. 2012. Tbx2 and Tbx3 induce atrioventricular myocardial development and endocardial cushion formation. *Cell Mol Life Sci* **69**:1377–1389. doi:10.1007/s00018-011-0884-2

Sive HL. 2000. Early Development of Xenopus laevis: A Laboratory Manual. Cold Spring Harbor Laboratory Press.

Smith JC, Yaqoob M, Symes K. 1988. Purification, partial characterization and biological effects of the XTC mesoderm-inducing factor. *Development* **103**:591–600. doi:10.1242/dev.103.3.591

Srivastava D, Thomas T, Lin Q, Kirby ML, Brown D, Olson EN. 1997. Regulation of cardiac mesodermal and neural crest development by the bHLH transcription factor, dHAND. *Nat Genet* **16**:154–160. doi:10.1038/ng0697-154

Stemmer M, Thumberger T, Del Sol Keyer M, Wittbrodt J, Mateo JL. 2015. Cctop: an intuitive, flexible and reliable crispr/cas9 target prediction tool. *PLoS ONE* **10**:e0124633. doi:10.1371/journal.pone.0124633

Stennard FA, Costa MW, Lai D, Biben C, Furtado MB, Solloway MJ, McCulley DJ, Leimena C, Preis JI, Dunwoodie SL, Elliott DE, Prall OWJ, Black BL, Fatkin D, Harvey RP. 2005. Murine T-box transcription factor Tbx20 acts as a repressor during heart development, and is essential for adult heart integrity, function and adaptation. *Development* **132**:2451–2462. doi:10.1242/dev.01799

Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ, Mering C von. 2019. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* **47**:D607–D613. doi:10.1093/nar/gky1131

Takahashi S, Yokota C, Takano K, Tanegashima K, Onuma Y, Goto J, Asashima M. 2000. Two novel nodal-related genes initiate early inductive events in Xenopus Nieuwkoop center. *Development* **127**:5319–5329. doi:10.1242/dev.127.24.5319

Takeuchi JK, Bruneau BG. 2009. Directed transdifferentiation of mouse mesoderm to heart tissue by defined factors. *Nature* **459**:708–711. doi:10.1038/nature08039

Tan F-L, Moravec CS, Li J, Apperson-Hansen C, McCarthy PM, Young JB, Bond M. 2002. The gene expression fingerprint of human heart failure. *Proc Natl Acad Sci USA* **99**:11387–11392. doi:10.1073/pnas.162370099

Troncoso MF, Ortiz-Quintero J, Garrido-Moreno V, Sanhueza-Olivares F, Guerrero-Moncayo A, Chiong M, Castro PF, García L, Gabrielli L, Corbalán R, Garrido-Olivares L, Lavandero S. 2021. VCAM-1 as a predictor biomarker in cardiovascular disease. *Biochim Biophys Acta Mol Basis Dis* **1867**:166170. doi:10.1016/j.bbadis.2021.166170

Varet H, Brillet-Guéguen L, Coppée J-Y, Dillies M-A. 2016. SARTools: A DESeq2- and EdgeR-Based R Pipeline for Comprehensive Differential Analysis of RNA-Seq Data. *PLoS ONE* **11**:e0157022. doi:10.1371/journal.pone.0157022

Wang T, Tian J, Jin Y. 2021. VCAM1 expression in the myocardium is associated with the risk of heart failure and immune cell infiltration in myocardium. *Sci Rep* **11**:19488. doi:10.1038/s41598-021-98998-3

Watt AJ, Battle MA, Li J, Duncan SA. 2004. GATA4 is essential for formation of the proepicardium and regulates cardiogenesis. *Proc Natl Acad Sci U S A* **101**:12573–8.

Wiesinger A, Li J, Fokkert L, Bakker P, Verkerk AO, Christoffels VM, Boink GJJ, Devalla HD. 2022. A single cell transcriptional roadmap of human pacemaker cell differentiation. *eLife* **11**. doi:10.7554/eLife.76781

Wilczewski CM, Hepperla AJ, Shimbo T, Wasson L, Robbe ZL, Davis IJ, Wade PA, Conlon FL. 2018. CHD4 and the NuRD complex directly control cardiac sarcomere formation. *Proc Natl Acad Sci USA* **115**:6727–6732. doi:10.1073/pnas.1722219115

Xiang F-L, Fang M, Yutzey KE. 2017. Loss of β-catenin in resident cardiac fibroblasts attenuates fibrosis induced by pressure overload in mice. *Nat Commun* **8**:712. doi:10.1038/s41467-017-00840-w

Xie Y, Liu J, Qian L. 2022. Direct cardiac reprogramming comes of age: Recent advance and remaining challenges. *Semin Cell Dev Biol* **122**:37–43. doi:10.1016/j.semcdb.2021.07.010

Yamada Y, Sadahiro T, Nakano K, Honda S, Abe Y, Akiyama T, Fujita R, Nakamura M, Maeda T, Kuze Y, Onishi M, Seki M, Suzuki Y, Takeuchi C, Iwasaki YW, Murano K, Sakata-Yanagimoto M, Chiba S, Kato H, Sakamoto H, Hiramatsu Y, Ieda M. 2024. Cardiac reprogramming and gata4 overexpression reduce fibrosis and improve diastolic dysfunction in heart failure with preserved ejection fraction. *Circulation*. doi:10.1161/CIRCULATIONAHA.123.067504

Yamashiro C, Hirota T, Kurimoto K, Nakamura T, Yabuta Y, Nagaoka SI, Ohta H, Yamamoto T, Saitou M. 2016. Persistent requirement and alteration of the key targets of PRDM1 during primordial germ cell development in mice. *Biol Reprod* **94**:7. doi:10.1095/biolreprod.115.133256

Ye C, Yang C, Zhang H, Gao R, Liao Y, Zhang Yali, Jie L, Zhang Yanhui, Cheng T, Wang Y, Ren J. 2024. Canonical Wnt signaling directs the generation of functional human PSC-derived atrioventricular canal cardiomyocytes in bioprinted cardiac tissues. *Cell Stem Cell* **31**:398-409.e5. doi:10.1016/j.stem.2024.01.008

Zeisberg EM, Ma Q, Juraszek AL, Moses K, Schwartz RJ, Izumo S, Pu WT. 2005. Morphogenesis of the right ventricle requires myocardial expression of Gata4. *J Clin Invest* **115**:1522–1531. doi:10.1172/JCI23769

Zhang L, Smyth D, Al-Khalaf M, Blet A, Du Q, Bernick J, Gong M, Chi X, Oh Y, Roba-Oshin M, Coletta E, Feletou M, Gramolini AO, Kim K-H, Coutinho T, Januzzi JL, Tyl B, Ziegler A, Liu PP. 2022. Insulin-like growth factor-binding protein-7 (IGFBP7) links senescence to heart failure. *Nat Cardiovasc Res* **1**:1195–1214. doi:10.1038/s44161-022-00181-y

Figure Legends

Figure 1. Identification and validation of cardiogenesis-related targets of GATA4. (A) Experimental outline. Animal cap explants from blastula stage embryos were excised and incubated for the indicated time before being harvested and subjected to RNA-seq analysis. The samples injected with mRNAs coding for proteins known to induce cardiac tissue are indicated in red. GATA4-GR was activated by incubation in 1 mM dexamethasone. The control sample were uninjected animal cap explants. Animal caps treated with a low dose of soluble activin, known to induce skeletal but not cardiac muscle (Fig. S1), was used as an additional control and a method to select genes responsive to cardiac inducing conditions. (B) Upset plot describing intersection of differentially expressed genes irrespective of timepoint compared to control. Bars indicate total number of genes differentially expressed in combination of conditions indicated in lower plot at any timepoint. Supplemental Figure 1E describes per timepoint differentially expressed genes. (C) Heatmap showing genes activated and repressed per condition irrespective of timepoint. Left grey heatmap indicates whether gene was significantly differentially expressed in indicated condition. Corresponding heatmap divided per time point shown in Supplemental Figure 1. (D) **.** Gene set enrichment calculated using Enrichr. Heatmap shows -log10 FDR for enrichment of indicated term for genesets defined by upset plot in (B)**,** within WikiPathway, BioPlanet and KEGG annotations. Heatmap divided into five classes of differing enrichment response. Geneset enrichments for Gene Ontology is provided in Supplemental Figure 2, and table of all tested terms provided in Supplemental Table 1. (E, F) RT-PCR validation of a subset of upregulated DEGs. Time points where each DEG was validated are indicated by an asterisk. o*dc1* expression is used as a loading control. The same *odc1* data is shown in all panels where the same cDNA samples were used. (G) RT-PCR validation of downregulated DEGs at 2.5h of incubation. All DEGs were validated as downregulated by GATA4-GR, GATA4-GR + Dkk-1 and Nodal 5 compared to the level of expression in uninjected or activin treated animal caps. (H) RT-PCR validation of DEGs linked with endoderm development at 2.5h and 24h. Time points where each DEG was validated are indicated by an asterisk.

Figure 2. *id3*, *tbx2* and *prdm1* are regulated by GATA4. (A) RT-PCR analysis for the expression of *id2-4* genes in control (uninjected), GATA4 injected, GATA4+Dkk-1 injected, Nodal 5 injected and activin treated animal cap explants at 2.5 and 5h after excision of animal cap explants and activation of inducible GATA4-GR by 1 mM dexamethasone. *id3* is downregulated by cardiogenic conditions at 2.5h and more strongly at 5h. This is in contrast to family members *id2* and *id4*, which were not picked up in the screen. *odc1* expression is used as loading control. (B) GATA4 is required for regulation of *id2-4* gene expression in vivo. Embryos injected with control Morpholino Oligonucleotides, GATA4 Morpholino Oligonucleotides (G4MO) or GATA5 Morpholino Oligonucleotides (G5MO) were collected at indicated stages and analysed for expression of *id2-4* and *odc1* by RT-PCR. Downregulation of GATA4 (G4MO samples) leads to upregulation of *id2-4* expression at st. 10 and st. 19 (indicated by red box), as assessed by RT-PCR. (C) GATA4 associates with *id3* genomic locus. 3 genomic regions containing GATA sites were amplified from genomic DNA (input) or from DNA immunoprecipitated (IP) by antibodies against HA-tagged GATA4. Region 2 but not 1 and 3 associate with GATA4. (D) Validation of *tbx2* and *tbx5* as GATA4 targets in animal cap explants. Samples collected at indicated time points were analysed by RT-PCR for expression of *tbx2*, *tbx5* and *odc1*. *tbx2* is strongly induced by GATA4 and GATA4+Dkk-1 and weakly by Nodal 5 at 5h; *tbx5* is induced by all 3 cardiogenic conditions at 7.5h and by GATA4 and GATA4+Dkk-1 at 5h and 24h. (E) GATA4 is required for regulation of *prdm1* and *tbx2* in vivo. *prdm1* expression in both GATA4 splice-blocking MO (G4SP) and GATA5 splice-blocking MO (G5SP) is upregulated compared to control MO-injected embryos at st.17, whereas *tbx2* expression is downregulated upon GATA4 and GATA5 knockdown (indicated by red boxes). (F) GATA4 binds discrete genomic regions of *prdm1* and *tbx2*. Two additional GATA4-binding regions were identified for *prdm1* and are shown in Fig. S3B. The region 2 is the same as the amplicon 12 in Fig. S3B. Sequences of all positive amplicons are shown in Fig. S3C. (G) Downregulation of *tbx2* and gain of function of *prdm1* lead to heart abnormalities in vivo. Ventral views of representative st. 34 embryos (anterior to the top) with the heart revealed by *myl7* WMISH. Injection of *tbx2* MO leads to cardia bifida or smaller heart region. Injection of *prdm1* mRNA causes defective development of the heart (linear heart tube or cardia bifida). Additional examples and phenotype classification are shown in Fig. S3D, E. Dashed line outlines cement gland in a sample where it is not clear.

Figure 3. GATA4 is required for regulation of its target genes PRDM1 and TBX2 in iPSC-CMs. (A) Analysis of the expression of GATA4 and its target genes during iPSC-CM differentiation examined using RT-PCR. Representative of 3 repeats. (B) A western blot showing the absence of GATA4 protein at day 4 in the GATA4 null lines created using CRISPR-Cas9 gene editing. Representative of 3 repeats. ERK2 is used as a loading control. (C) RT-PCR data showing the expression of GATA4 in mutant cells at day 4 is reduced, versus an increase in the expression of PRDM1 (arrowhead), when compared to WT cells. (D) RT-PCR for day 10 samples showing GATA4 expression remains lower in the mutant lines, as is TBX2 (arrowhead). Representative of 2 experiments. For all RT-PCRs GAPDH has been used as a normalisation control.

Figure 4. GATA4 null iPSCs fail to form normal cardiomyocytes. (A) Immunofluorescent (IF) staining for TNNT2 (green) in WT and GATA4$^{-/-}$ 1 mutant cells at Day 32 of cardiac directed differentiation. DAPI (blue) was used to counterstain nuclei. Representative examples from 4 experiments are shown. (B) A comparison of the proportion of beating wells detected per experiment between the WT and GATA4 null line. Based on 12 independent differentiation experiments. (C) A comparison of the proportion of TNNT2 positive cells identified by IF between the cell lines. TNNT2+ ive WT cells = 64.25% ± 10.87 and GATA4 -/- 1 = 14.50% ± 7.14. (D) TNNT2 mean fluorescence per cell for WT and GATA4$^{-/-}$ 1, n = 3. For panels B and C a one-way ANOVA was used for statistical analysis. For panel D, a student's t-test has been used. P-values: * ≤ 0.05, and **** ≤ 0.0001. (E) Representative images of individual WT and GATA4$^{-/-}$ 1 cells at 40x magnification stained for TNNT2 or MYBPC3 (green), DAPI (blue) was used to counterstain nuclei. The enlarged images outlined in red box provide a closer look at individual myofibrils. In the WT line >90% of the CMs present have clear sarcomeres compared to <7% for the GATA4-/-1 line. Further examples are presented in supplemental figure S6. (F) Fluctuations in calcium levels were measured over time in a field of WT and GATA4$^{-/-}$ 1 cells at Day 32 of differentiation using calcium indicator Fluo-8AM. The fluctuation in signal is shown as an RGB pseudo-coloured projection. Red areas being pixels with the highest signal variance over time, and blue being the lowest, n = 3. (G) In addition to these overviews, calcium flux was plotted over time from selected regions of interest for each cell line. For each cells line n= 9, with 3 wells analysed across 3 independent experiments.

Figure 5. In GATA4 -/- cells, cardiomyocyte and endothelial cell transcriptomes are suppressed whereas smooth muscle cell and fibroblast genes are upregulated. (A) A

heatmap showing the normalised expression of a selection of marker genes indicative of progression through iPSC-CM differentiation (B) Analysis of a number of cardiac progenitor marker genes by RT-PCR, expression of GAPDH was used for normalisation. n =2. (C) A heatmap showing the normalised expression of a selection of cardiomyocyte (CM), myofibroblast (MyoF), endocardial (EC), extracellular matrix (ECM) and smooth muscle cell (SMC) genes at Day 10. A selection of the genes from C were validated in an additional RNA sets at days 10 (D) and 32 (E) of differentiation using RT-PCR (n = 2). (F) Immunofluorescent images showing WT and GATA4 -/-1 cell at 40x magnification at day 32 of differentiation. Stained for TNNT2 (green), cell mask orange (Orange), and nuclei stained with DAPI (blue). (G) Cell membrane morphology was classified for WT and GATA4 -/- 1 cells into smooth, convoluted, and highly convoluted. Those belonging to the convoluted category have membrane extensions, and those in the highly convoluted category displaying many of these branching extensions. WT cells analysed = 88, GATA4-/-1 = 55. (H) A table showing the direction of change for a selection of prospective GATA4 targets identified in the *Xenopus* screen (GATA4 overexpression) and their expression change in the GATA4 null iPSC line. Upregulated genes in the *Xenopus* column would be expected to be downregulated in human samples, and vice versa. (I-J) The pie charts displayed show the number of DEGs in the GATA4 mutants, at days 5 and 10 respectively, with a Log2FC value of 2 or more that are bound by GATA4 in foetal or adult mouse hearts, according to datasets generated by Akerberg *et al.* 2019. The accompanying tables list a selection of these bound genes and their log2FC value from our RNA-seq analysis of the GATA4-/- 1 null line.

Figure 6. TBX2 mutant cells show pathological hypertrophy and upregulation of CM genes.

(A) A western blot showing TBX2 expression in *Xenopus* embryos injected uniformly with WT TBX2 or TBX2$^{c.760_761del}$ mRNA at 2-4 cell stage. The embryos were harvested for protein analysis at stages 8.5 (blastula) and 10.5 (early gastrula). ERK2 was used as a loading control. The mutant protein is readily detected at st. 8.5 but is subsequently degraded at st. 10.5. (B) The TBX2 protein structure is shown with the DNA binding domain indicated in green. In TBX2$^{c.700_761del}$ mutants, exon 3 is skipped. The region of the TBX2 DNA binding domain affected by this is indicated. Image produced using BioRender. (C) The proportion of TNNT2 positive cells in the total cell population was recorded. The values are given as a percentage

± SEM and are as follows; WT 62% ± 7.00 and TBX2$^{c.700_761del}$ 66% ± 9.23 (n=3). (D) The surface area of the CMs was measured across 3 differentiations at day 32 and 1 differentiation at day 12. (E) Representative fields of WT and TBX2 mutant day 32 iPSC-CMs are shown at 40x magnification. The cells were stained for TNNT2 (green) to identify CMs, and counterstained with DAPI (blue) to identify nuclei. (F) Enlarged pictures of WT and TBX2 mutant CMs at day 32 showing disarray of myofibrils and sarcomeres. (G) Representative images of cells belonging to classes I-IV, with class I cells showing the highest level of order and class IV the most severe disorganisation. (H) Quantification of cardiomyocyte disorganisation in the WT and TBX2 mutant line through categorisation into one of the four categories described in (G), n =3. (I) Representative Ca^{++} transients for WT and TBX2$^{c.700_761del}$ mutants at ~ day 32 of differentiation for each cell line. (J) The amplitude of individual Ca cell transients. (K) Calcium imaging time to peak measurements. A One-way ANOVA has been used for all statistical comparisons shown. P-values; * ≤ 0.05, *** ≤ 0.001, and **** ≤ 0.0001. (L) (*upper panel*) Exemplar confocal Ca^{++} line scan recordings of Ca^{++} transients in a WT and TBX2 mutant cell. Spontaneous Ca^{++} sparks are visible during diastole between beats. (*lower panel*) TBX2 mutant cells tended to exhibit more diastolic Ca^{++} sparks than WT cells, although this was not statistically significant (P = 0.09; X^2 test) (M) The heatmap displays the normalised RNA-seq data for a selection of cardiac progenitor and myocardium differentiation markers, n = 2. (N) A selection of GO terms found to be upregulated in the TBX2$^{c.700_761del}$ mutant at day 10 of differentiation. Figure produced using ShinyGo.

Figure 7. Accelerated formation of CMs from PRDM1$^{-/-}$ iPS cells

(A) Western blot analysis of PRDM1 protein in WT and mutant iPS cells at indicated days of iPSC-CM differentiation. ERK2 is used as a loading control. (B) Beating onset was recorded for WT (average onset = day 9.9) and PRDM1$^{-/-}$ (average onset = day 7.8) lines, n = 8. (C) CM differentiation efficiency, assessed by staining for TNNT2 (green) at day 10, is not affected in PRDM1$^{-/-}$ cells. The values are given as a percentage ± SEM and are as follows; WT 73% ± 16.3 n=576; PRDM1$^{-/-}$ 84% ± 12.3 n=367, across n = 3 experiments. (D) Representative images of WT and PRDM1$^{-/-}$ cells stained for TNNT2 (green) and counterstained with DAPI (blue) for nuclei detection at day 32 of differentiation. Shown at 20x magnification. (E) Samples collected at day 7 of differentiation were subjected to RNA-seq. A heatmap showing

the expression of cardiac progenitor and cardiomyocyte specific genes in the WT and PRDM1$^{-/-}$ lines at day 7 of differentiation, n = 1. (F) Expression of a selection of sarcomere genes analysed by RT-PCR in a WT and PRDM1$^{-/-}$ differentiation. *GAPDH* is used as a loading control. Days of iPSC-CM differentiation and number of PCR cycles are indicated. (G) An upset plot showing the number of DEGs per time point and their direction of change. (H) Comparable sarcomeric structure was seen between the WT and PRDM1 mutant cardiomyocytes. Staining for sarcomeric protein TNNT2 (green) in day 32 hCMs formed from WT and PRDM1 mutant cells, counterstained with DAPI (blue); enlarged images are shown on the right-hand side. (G) The proportion of cells in each class, WT n = 180 and PRDM1-/-, n = 150, over n = 3 experiments. The classification system utilised is as described in Figure 6.

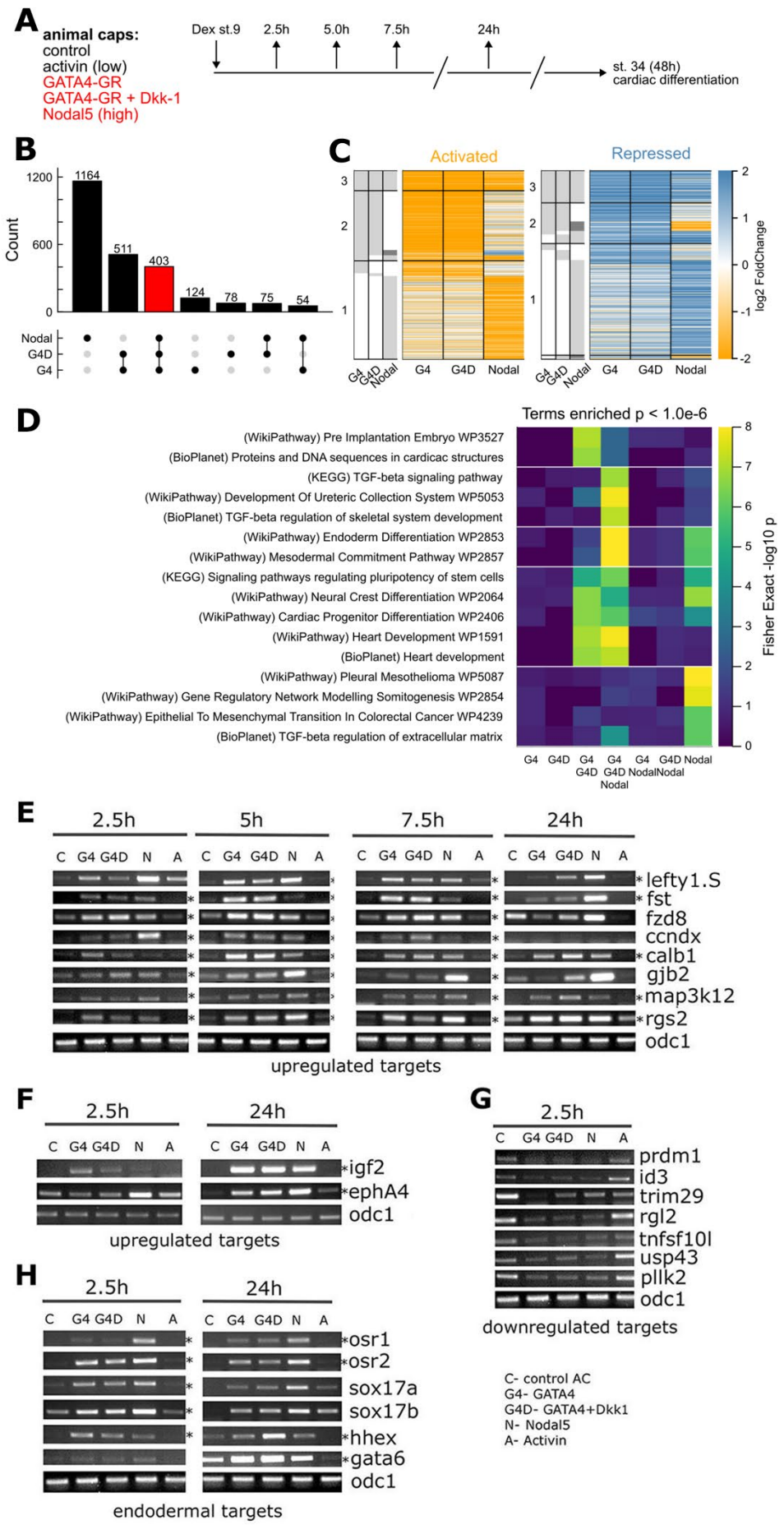
A
animal caps:
control
activin (low)
GATA4-GR
GATA4-GR + Dkk-1
Nodal5 (high)
Dex st.9
2.5h
5.0h
7.5h
24h
st. 34 (48h)
cardiac differentiation
B
Count
1164
511
403
124
78
75
54
Nodal
G4D
G4
C
Activated
Repressed
log2 FoldChange
D
Terms enriched p < 1.0e-6
(WikiPathway) Pre Implantation Embryo WP3527
(BioPlanet) Proteins and DNA sequences in cardiac structures
(KEGG) TGF-beta signaling pathway
(WikiPathway) Development Of Ureteric Collection System WP5053
(BioPlanet) TGF-beta regulation of skeletal system development
(WikiPathway) Endoderm Differentiation WP2853
(WikiPathway) Mesodermal Commitment Pathway WP2857
(KEGG) Signaling pathways regulating pluripotency of stem cells
(WikiPathway) Neural Crest Differentiation WP2064
(WikiPathway) Cardiac Progenitor Differentiation WP2406
(WikiPathway) Heart Development WP1591
(BioPlanet) Heart development
(WikiPathway) Pleural Mesothelioma WP5087
(WikiPathway) Gene Regulatory Network Modelling Somitogenesis WP2854
(WikiPathway) Epithelial To Mesenchymal Transition In Colorectal Cancer WP4239
(BioPlanet) TGF-beta regulation of extracellular matrix
Fisher Exact -log10 p
E
2.5h
5h
7.5h
24h
C G4 G4D N A
lefty1.S
fst
fzd8
ccndx
calb1
gjb2
map3k12
rgs2
odc1
upregulated targets
F
igf2
ephA4
odc1
upregulated targets
G
prdm1
id3
trim29
rgl2
tnfsf10l
usp43
pllk2
odc1
downregulated targets
H
osr1
osr2
sox17a
sox17b
hhex
gata6
odc1
endodermal targets
C- control AC
G4- GATA4
G4D- GATA4+Dkk1
N- Nodal5
A- Activin

Figure 1

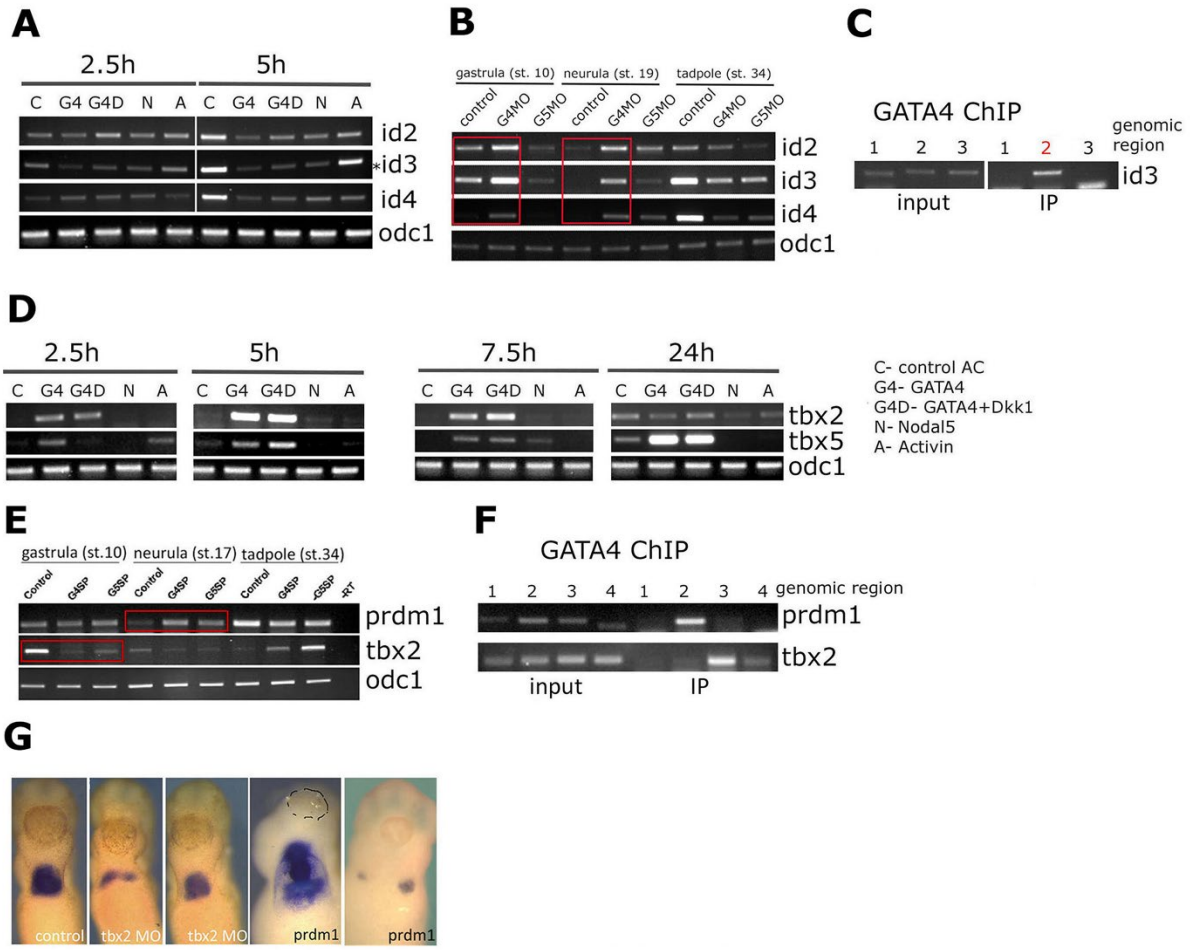
A
2.5h
5h
C G4 G4D N A C G4 G4D N A
id2
*id3
id4
odc1
B
gastrula (st. 10)
neurula (st. 19)
tadpole (st. 34)
control G4MO G5MO control G4MO G5MO control G4MO G5MO
id2
id3
id4
odc1
C
GATA4 ChIP
genomic region
1 2 3 1 2 3
id3
input
IP
D
2.5h
5h
7.5h
24h
C G4 G4D N A
tbx2
tbx5
odc1
C- control AC
G4- GATA4
G4D- GATA4+Dkk1
N- Nodal5
A- Activin
E
gastrula (st.10)
neurula (st.17)
tadpole (st.34)
Control G4SP G5SP Control G4SP G5SP Control G4SP G5SP RT
prdm1
tbx2
odc1
F
GATA4 ChIP
1 2 3 4 1 2 3 4 genomic region
prdm1
tbx2
input
IP
G
control
tbx2 MO
tbx2 MO
prdm1
prdm1

Figure 2

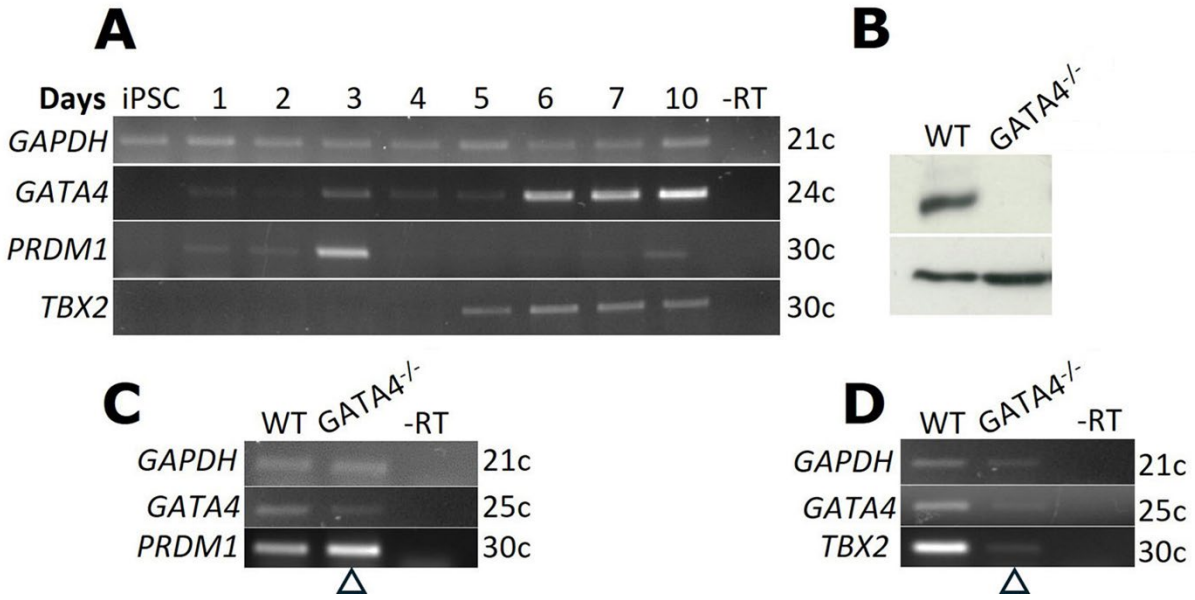
A
Days iPSC 1 2 3 4 5 6 7 10 -RT
GAPDH
GATA4
PRDM1
TBX2
21c
24c
30c
30c
B
WT GATA4-/-
C
WT GATA4-/- -RT
GAPDH
GATA4
PRDM1
21c
25c
30c
D
WT GATA4-/- -RT
GAPDH
GATA4
TBX2
21c
25c
30c

Figure 3

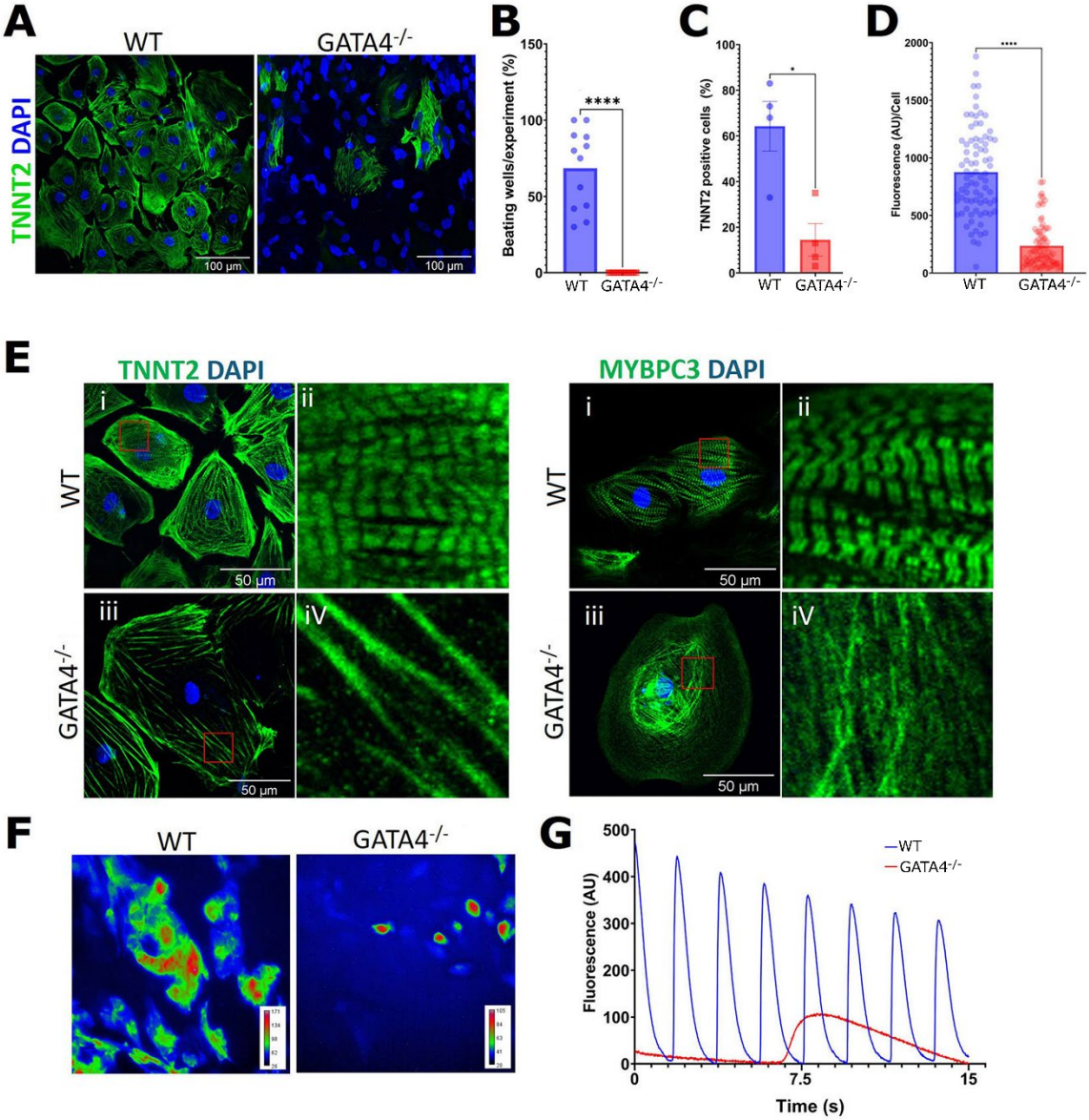
A
WT
GATA4-/-
TNNT2 DAPI
100 µm
100 µm
B
Beating wells/experiment (%)
****
WT
GATA4-/-
C
TNNT2 positive cells (%)
*
WT
GATA4-/-
D
Fluorescence (AU)/Cell
****
WT
GATA4-/-
E
TNNT2 DAPI
MYBPC3 DAPI
WT
GATA4-/-
i
ii
iii
iV
50 µm
F
WT
GATA4-/-
G
Fluorescence (AU)
Time (s)
WT
GATA4-/-

Figure 4

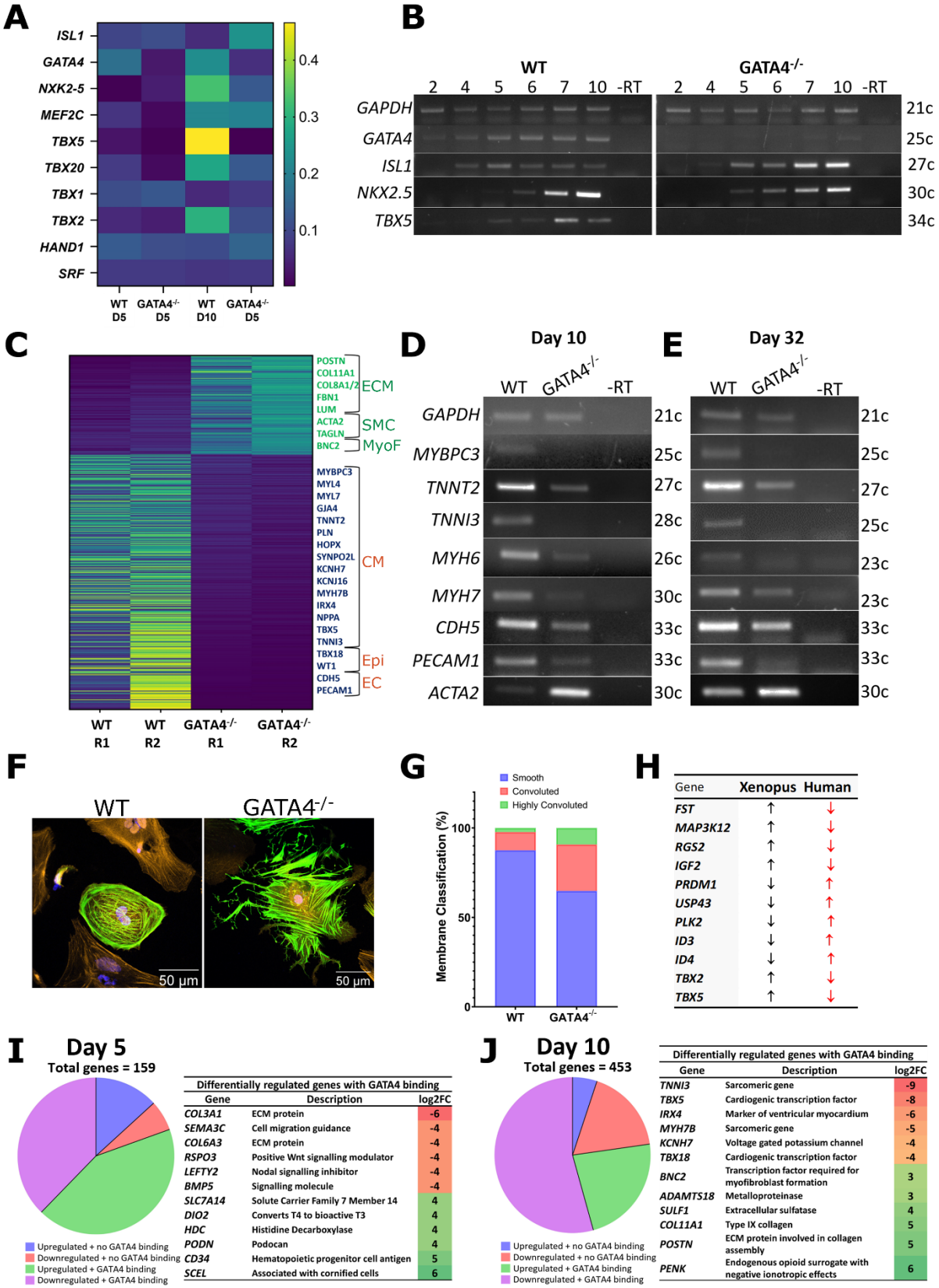
A
ISL1
GATA4
NKX2-5
MEF2C
TBX5
TBX20
TBX1
TBX2
HAND1
SRF
0.4
0.3
0.2
0.1
WT D5
GATA4-/- D5
WT D10
GATA4-/- D5
B
WT
GATA4-/-
2 4 5 6 7 10 -RT
GAPDH
GATA4
ISL1
NKX2.5
TBX5
21c
25c
27c
30c
34c
C
POSTN
COL11A1
COL8A1/2
FBN1
LUM
ACTA2
TAGLN
BNC2
ECM
SMC
MyoF
MYBPC3
MYL4
MYL7
GJA4
TNNT2
PLN
HOPX
SYNPO2L
KCNH7
KCNJ16
MYH7B
IRX4
NPPA
TBX5
TNNI3
CM
TBX18
WT1
Epi
CDH5
PECAM1
EC
WT R1
WT R2
GATA4-/- R1
GATA4-/- R2
D
Day 10
WT
GATA4-/-
-RT
GAPDH
MYBPC3
TNNT2
TNNI3
MYH6
MYH7
CDH5
PECAM1
ACTA2
21c
25c
27c
28c
26c
30c
33c
33c
30c
E
Day 32
WT
GATA4-/-
-RT
21c
25c
27c
25c
23c
23c
33c
33c
30c
F
WT
GATA4-/-
50 µm
50 µm
G
Smooth
Convoluted
Highly Convoluted
Membrane Classification (%)
100
50
0
WT
GATA4-/-
H
Gene Xenopus Human
FST ↑ ↓
MAP3K12 ↑ ↓
RGS2 ↑ ↓
IGF2 ↑ ↓
PRDM1 ↓ ↑
USP43 ↓ ↑
PLK2 ↓ ↑
ID3 ↓ ↑
ID4 ↓ ↑
TBX2 ↑ ↓
TBX5 ↑ ↓
I
Day 5
Total genes = 159
Upregulated + no GATA4 binding
Downregulated + no GATA4 binding
Upregulated + GATA4 binding
Downregulated + GATA4 binding
Differentially regulated genes with GATA4 binding
Gene Description log2FC
COL3A1 ECM protein -6
SEMA3C Cell migration guidance -4
COL6A3 ECM protein -4
RSPO3 Positive Wnt signalling modulator -4
LEFTY2 Nodal signalling inhibitor -4
BMP5 Signalling molecule -4
SLC7A14 Solute Carrier Family 7 Member 14 4
DIO2 Converts T4 to bioactive T3 4
HDC Histidine Decarboxylase 4
PODN Podocan 4
CD34 Hematopoietic progenitor cell antigen 5
SCEL Associated with cornified cells 6
J
Day 10
Total genes = 453
Upregulated + no GATA4 binding
Downregulated + no GATA4 binding
Upregulated + GATA4 binding
Downregulated + GATA4 binding
Differentially regulated genes with GATA4 binding
Gene Description log2FC
TNNI3 Sarcomeric gene -9
TBX5 Cardiogenic transcription factor -8
IRX4 Marker of ventricular myocardium -6
MYH7B Sarcomeric gene -5
KCNH7 Voltage gated potassium channel -4
TBX18 Cardiogenic transcription factor -4
BNC2 Transcription factor required for myofibroblast formation 3
ADAMTS18 Metalloproteinase 3
SULF1 Extracellular sulfatase 4
COL11A1 Type IX collagen 5
POSTN ECM protein involved in collagen assembly 5
PENK Endogenous opioid surrogate with negative ionotropic effects 6

Figure 5

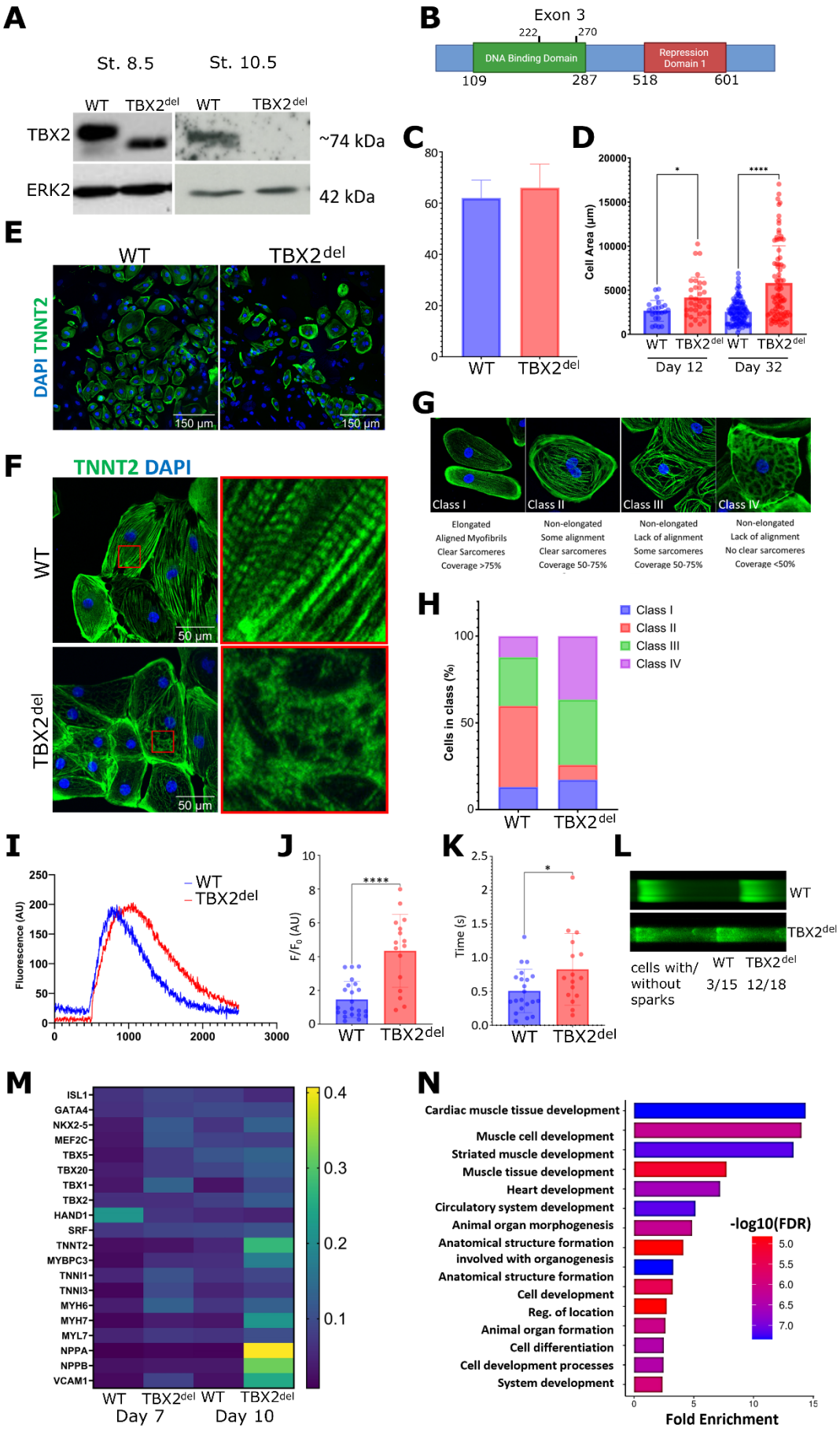
A
St. 8.5
St. 10.5
WT
TBX2del
WT
TBX2del
TBX2
ERK2
~74 kDa
42 kDa
B
Exon 3
222
270
DNA Binding Domain
Repression Domain 1
109
287
518
601
C
WT
TBX2del
D
Cell Area (µm)
WT
TBX2del
WT
TBX2del
Day 12
Day 32
E
WT
TBX2del
DAPI TNNT2
150 µm
150 µm
F
TNNT2 DAPI
WT
TBX2del
50 µm
50 µm
G
Class I
Class II
Class III
Class IV
Elongated
Aligned Myofibrils
Clear Sarcomeres
Coverage >75%
Non-elongated
Some alignment
Clear sarcomeres
Coverage 50-75%
Non-elongated
Lack of alignment
Some sarcomeres
Coverage 50-75%
Non-elongated
Lack of alignment
No clear sarcomeres
Coverage <50%
H
Cells in class (%)
Class I
Class II
Class III
Class IV
WT
TBX2del
I
Fluorescence (AU)
WT
TBX2del
J
F/F0 (AU)
WT
TBX2del
K
Time (s)
WT
TBX2del
L
WT
TBX2del
cells with/ without sparks
WT
TBX2del
3/15
12/18
M
ISL1
GATA4
NKX2-5
MEF2C
TBX5
TBX20
TBX1
TBX2
HAND1
SRF
TNNT2
MYBPC3
TNNI1
TNNI3
MYH6
MYH7
MYL7
NPPA
NPPB
VCAM1
WT
TBX2del
WT
TBX2del
Day 7
Day 10
N
Cardiac muscle tissue development
Muscle cell development
Striated muscle development
Muscle tissue development
Heart development
Circulatory system development
Animal organ morphogenesis
Anatomical structure formation involved with organogenesis
Anatomical structure formation
Cell development
Reg. of location
Animal organ formation
Cell differentiation
Cell development processes
System development
-log10(FDR)
Fold Enrichment

Figure6

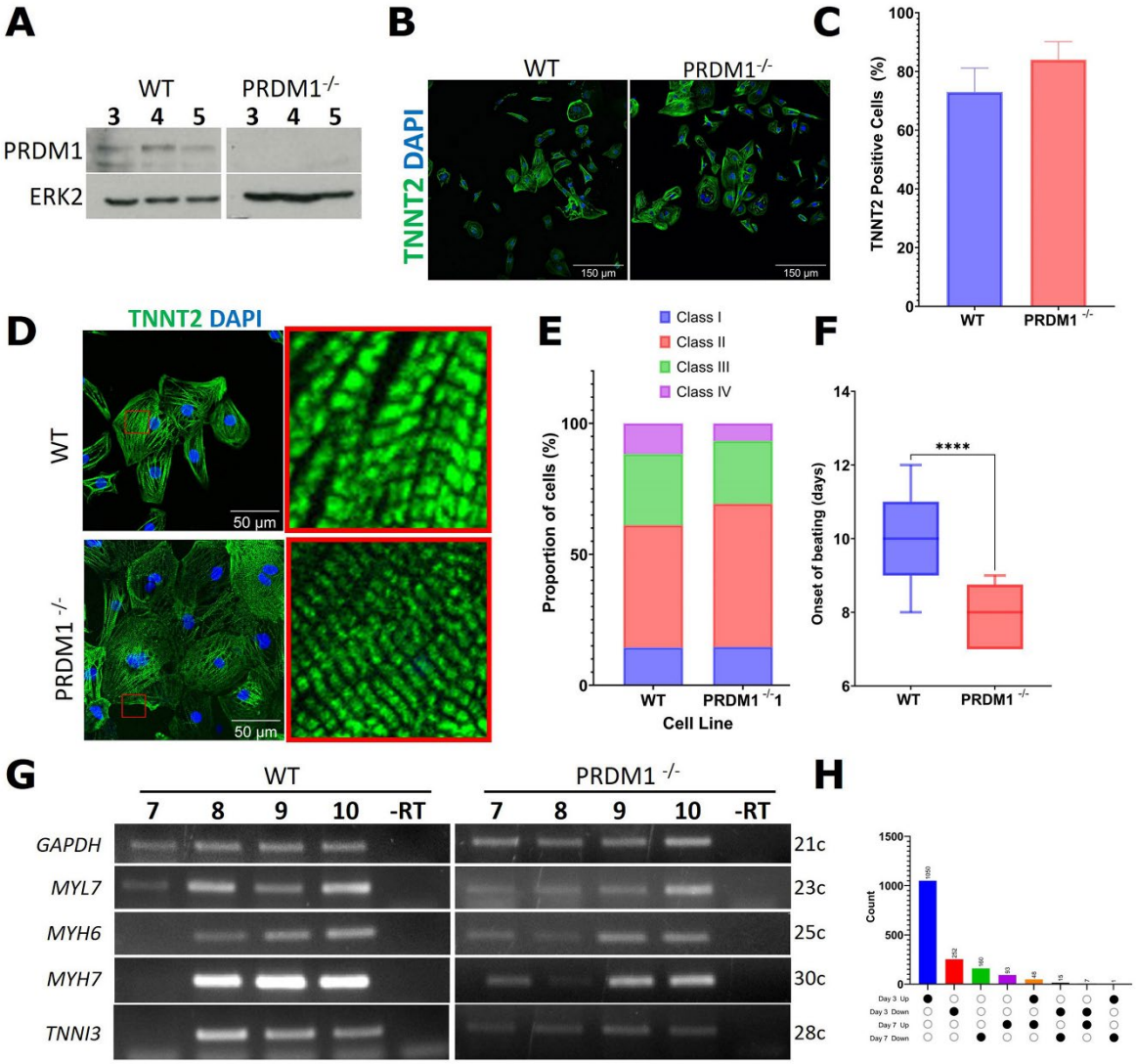
A
WT
PRDM1-/-
3 4 5 3 4 5
PRDM1
ERK2
B
WT
PRDM1-/-
TNNT2 DAPI
150 µm
150 µm
C
TNNT2 Positive Cells (%)
100
80
60
40
20
0
WT
PRDM1 -/-
D
TNNT2 DAPI
WT
PRDM1 -/-
50 µm
50 µm
E
Class I
Class II
Class III
Class IV
Proportion of cells (%)
100
50
0
WT
PRDM1 -/-1
Cell Line
F
Onset of beating (days)
14
12
10
8
6
****
WT
PRDM1 -/-
G
WT
PRDM1 -/-
7 8 9 10 -RT 7 8 9 10 -RT
GAPDH
MYL7
MYH6
MYH7
TNNI3
21c
23c
25c
30c
28c
H
Count
1500
1000
500
0
Day 3 Up
Day 3 Down
Day 7 Up
Day 7 Down

Figure 7