Mesp1 controls the chromatin and enhancer landscapes essential for spatiotemporal patterning of early cardiovascular progenitors

Xionghui Lin[#1], **Benjamin Swedlund**[#1], **Mai-Linh N. Ton**[2,3], **Shila Ghazanfar**[5], **Carolina Guibentif**[2,3,4], **Catherine Paulissen**[1], **Elodie Baudelet**[1], **Elise Plaindoux**[8], **Younes Achouri**[6], **Emilie Calonne**[7], **Christine Dubois**[1], **William Mansfield**[3], **Stéphane Zaffran**[8], **John C. Marioni**[5], **Francois Fuks**[7], **Berthold Göttgens**[2,3], **Fabienne Lescroart**[8,11], **Cédric Blanpain**[1,9,11]

[1]Laboratory of Stem Cells and Cancer, Université Libre de Bruxelles, Brussels B-1070, Belgium

[2]Department of Haematology, Jeffrey Cheah Biomedical Centre, University of Cambridge, Cambridge CB2 0AW, UK

[3]Wellcome and Medical Research Council Cambridge Stem Cell Institute, University of Cambridge, Cambridge, UK

[4]Sahlgrenska Center for Cancer Research, Department of Microbiology and Immunology, University of Gothenburg, 413 90 Gothenburg, Sweden

[5]Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, UK

[6]Université Catholique de Louvain, Institut de Duve, Brussels, 1200, Belgium

[7]Laboratory of Cancer Epigenetics, Faculty of Medicine, ULB–Cancer Research Center (U-CRC), Université Libre de Bruxelles (ULB), Brussels, Belgium

[8]Aix Marseille Université, INSERM, MMG U1251, 13005 Marseille, France

[9]WELBIO, Université Libre de Bruxelles, Brussels B-1070, Belgium

[#] These authors contributed equally to this work.

Correspondence to: Fabienne Lescroart; Cédric Blanpain.

fabienne.lescroart@univ-amu.fr; cedric.blanpain@ulb.ac.be.
[11]These authors jointly supervised this work: Fabienne Lescroart, Cédric Blanpain

Abstract

The mammalian heart arises from various populations of Mesp1-expressing cardiovascular progenitors (CPs) that are specified during the early stages of gastrulation. Mesp1 is a transcription factor (TF) that acts as a master regulator of CP specification and differentiation. However, how Mesp1 regulates the chromatin landscape of nascent mesodermal cells to define the temporal and spatial patterning of the distinct populations of CPs remains unknown. Here, by combining ChIP-seq, RNA-seq and ATAC-seq during mouse pluripotent stem cell differentiation, we defined the dynamic remodelling of the chromatin landscape mediated by Mesp1. We identified different enhancers that are temporally regulated to erase the pluripotent state and specify the pools of CPs that mediate heart development. We identified Zic2 and Zic3 as essential cofactors that act with Mesp1 to regulate its TF activity at key mesodermal enhancers, thereby regulating the chromatin remodelling and gene expression associated with the specification of the different populations of CPs *in vivo*. Our study identifies the dynamics of the chromatin landscape and enhancer remodelling associated with temporal patterning of early mesodermal cells into the distinct populations of CPs that mediate heart development.

Introduction

During embryogenesis, tight temporal and spatial control of transcription is needed to correctly assign cell identities throughout the developing body. At the onset of gastrulation, epiblast cells shut down the gene regulatory network of the pluripotent state and start expressing genes specific to the different cell fates that are progressively specified. The transcription factors (TFs) that govern these cell fate transitions are relatively well known. However, how these lineage-specific master TFs control the temporal and spatial remodelling of the chromatin and enhancer landscape that mediate downstream gene expression to allocate the different cell fates at the correct place and time during embryonic development remain poorly understood.

The mammalian heart is the first organ to form during development, as it is essential for embryonic survival. It is a complex organ composed of four chambers and various cell lineages including cardiomyocytes (CMs), endocardial cells (ECs), fibroblasts and smooth muscle cells (SMCs)[1, 2]. The different regions of the heart (ventricles, atria and outflow tract) arise from the differentiation of distinct pools of cardiovascular progenitors (CPs) that are specified in a precise temporal and spatial pattern during gastrulation[3, 4]. Patterning defects during CP specification and differentiation lead to congenital heart disease, which represents the major cause of birth defects in humans[5–7].

Mesp1, a bHLH TF, is the earliest marker of CPs. Mesp1 is expressed transiently during mouse gastrulation in CPs that will contribute to all heart regions and cardiovascular cell types [8–12]. In mouse pluripotent stem cells (PSCs), Mesp1 directly promotes the expression of many TFs that compose the core gene regulatory network of cardiovascular development[13–16]. Lineage tracing experiments revealed that temporally distinct Mesp1+ CPs give rise to different heart regions and cardiovascular lineages[11, 17]. Moreover, scRNA-seq of Mesp1-expressing cells during mouse gastrulation *in vivo* demonstrated that Mesp1 CPs are temporally and spatially pre-patterned into progenitors of the different heart regions

and cardiovascular lineages during the early stages of mouse gastrulation[12, 18, 19]. The chromatin landscape and enhancer remodelling mediated by Mesp1 that promotes the specification of these different cardiovascular lineages and heart regions in a spatially and temporally regulated manner remains unknown.

Here, we define the precise dynamics of chromatin remodelling and enhancer logic by which Mesp1 temporally regulates gene expression programs during CP specification and differentiation *in vitro* and *in vivo*. Using RNA-seq, ChIP-seq and ATAC-seq, we found that Mesp1 induces remodelling of the chromatin and enhancer landscape to promote dynamic patterns of gene expression during development. We also identified Zic2 and Zic3 as regulators of Mesp1 TF activity at key mesodermal enhancers, promoting in a cooperative manner Mesp1 binding to the chromatin, chromatin remodelling and regulation of gene expression, allowing the specification and differentiation of CPs during the early stages of mouse gastrulation.

Results

Temporal regulation of gene expression mediated by Mesp1

To define the temporality of gene expression mediated by Mesp1, we performed RNA-seq during differentiation of mouse PSCs, 12 and 24 hours after doxycycline (dox)-induced Mesp1 expression at day 2.5 of embryoid body (EB) differentiation, slightly earlier than *Mesp1* endogenous expression (Fig. 1a). Mesp1 RNA-FISH showed that dox-mediated Mesp1 overexpression induces *Mesp1* expression in almost every cell (89 +/-2 %) within the EB whereas only 22 +/- 4% of the cells express *Mesp1* in control conditions at day (D) 3.5. Mesp1 expression per cell was on average 2.2 higher in dox compared to control conditions (Extended Data Fig. 1a-c). Our temporal analysis of gene expression showed that Mesp1 induces distinct patterns of expression that could be divided into different classes: genes that were rapidly and transiently (between 0 and 12 hours) induced by Mesp1 called "early genes" (e.g. *Snai1*), genes whose expression increased constantly between 0 and 24 hours called "constant genes", (e.g. *Pdgfra*) or genes activated only after 24 hours called "late genes" (e.g. *Hoxb1*). A Similar temporal pattern of early, constant and late downregulated genes could be identified (Fig. 1b-d, Extended data Fig 1d-f). Interestingly, the genes that were expressed at higher levels in embryonic day 6.75 (E6.75) *Mesp1*+ CPs *in vivo*[11] were mostly categorized as early upregulated *in vitro*, whereas the genes enriched at E7.25 were mostly classified as constant or late upregulated genes *in vitro* (Fig. 1e-h), showing that the temporality of Mesp1-regulated genes during PSC differentiation *in vitro* recapitulates the temporality of gene expression found during gastrulation *in vivo*.

Dynamics of chromatin and enhancer remodelling regulated by Mesp1

To investigate the temporal dynamics by which Mesp1 induces chromatin and enhancer remodelling associated with CP progenitor specification, we assessed the temporality of Mesp1 binding after its induction during PSC differentiation. To this end, we performed Mesp1 ChIP-seq 12 and 24 hours after overexpression of triple-HA tagged Mesp1 in PSCs induced at day 2.5 of embryonic bodies (EBs) differentiation (Fig. 1a and 2a). We defined a total of 2011 high-quality peaks, out of which 823 peaks were detected at 12 hours and 1792

peaks at 24 hours ($p < 10^{-10}$). The majority of these peaks were located either in intergenic regions or introns, with only 4% of peaks located at a transcriptional start site, demonstrating that Mesp1 acts as a lineage-specific TF through binding to distal regulatory elements (Extended Data Fig. 2a). We found that 536/1368 (39%) of upregulated genes presented a Mesp1 ChIP-seq peak within 500 kb of their transcription start site, versus194/1448 (13%) of downregulated genes, suggesting that Mesp1 acts mainly as a transcriptional activator (Fig. 2a-c).

Mesp1 ChIP-seq peaks could be classified into early (36/2011) (2%), constant (1451/2011) (72%), and late Mesp1 binding sites (525/2011) (26%). The kinetics of Mesp1 binding correlated with kinetics of gene expression, with early peaks being mostly associated with early upregulated genes, and late peaks with late upregulated genes (Fig. 2a-e).

To define how Mesp1 binding to its target sites affects the remodelling of the chromatin landscape, we performed ATAC-seq and histone 3 lysine 27 acetyl ($H_3K_{27}Ac$) ChIP-seq at 0, 12 and 24h following Mesp1 expression. Binding of Mesp1 to the regulatory regions of its target genes induced an opening of the chromatin, as defined by ATAC-seq, as well as an acetylation of histone 3 at lysine 27 ($H_3K_{27}Ac$ and monomethlyation of histone 3 at lysine 4 (H3K4me1) on neighbouring nucleosomes, two post-translational histone modifications associated with active enhancers (Fig. 2f-g)[20, 21]. Mesp1 promoted the opening of chromatin regions that were previously closed in the absence of Mesp1 expression at the majority (74%) of its binding sites (Fig. 2f-g), which we called *de novo* peaks. The remaining Mesp1-bound peaks were located in chromatin regions previously opened before Mesp1 binding, called primed peaks (21%) or that did not present detectable chromatin opening in any condition (5%). Positional analysis of nucleosomes in ATAC-seq data at Mesp1 binding sites confirmed that 68% of Mesp1-bound peaks were at least partially occluded by nucleosomes without Mesp1 induction, versus 13% after Mesp1 expression (Fig. 2h). Chromatin opening at these sites was associated with the temporality of Mesp1 binding, as only 20% of early Mesp1-bound peaks were classified as *de novo* peaks, versus 70% of constant and 87% of late peaks (Fig. 2i). The dynamics of chromatin remodelling and the temporality of H3K27Ac deposition around Mesp1-bound enhancers followed the kinetics of Mesp1 binding (Extended Data Fig. 2b-d).

We defined Mesp1-bound enhancers as DNA regions not located within a promoter that are bound by Mesp1, present opening of the chromatin, are flanked by histones marked with H3K27Ac and H3K4me1, and are associated with upregulation of gene expression following Mesp1 induction. To functionally validate the role of these putative enhancers, we deleted using CRISPR/Cas9n different Mesp1 binding regions (+/- 500 bp) presenting the hallmarks of enhancer activation after Mesp1 induction in regulatory regions of three different direct Mesp1 target genes that are important during cardiovascular development (*Hoxb1, Hand1*, and *Myocardin*). Deletion of a single Mesp1-bound putative enhancer strongly decreased the ability of Mesp1 to upregulate these target genes (Fig. 3b).

Different TFs are associated with temporally regulated Mesp1-bound enhancers

To define which TFs could cooperate with Mesp1 to control the different temporal patterns of gene expression induced by Mesp1, we performed motif discovery at Mesp1 ChIP-seq

peaks using Homer[22]. Most Mesp1 ChIP-seq peaks (88%) presented a very specific E-box motif: CAAATGG (Fig. 3c), equally present in early, constant and late ChIP-seq peaks, but significantly enriched in *de novo* peaks (90%) versus primed peaks (71%). This motif was strongly over-represented in comparison to other bHLH motifs in DNA footprints defined by ATAC-seq within Mesp1-bound enhancers (Extended Data Fig. 3a-c). The extra G next at the 3' end of the E-box motif is also found in the binding site of Ascl1, abHLH pioneer TF promoting reprograming of fibroblasts to neuronal cell fate[23, 24].

Mesp1-bound enhancers contained on average five bHLH binding motifs. To assess the importance of the most over-represented bHLH motif with an extra G within Mesp1 binding sites, we replaced one CAAATGG motif by GCTAGCG in the *Pdgfra* proximal enhancer, using CRISPR/Cas9-mediated homology directed repair (Fig. 3d). We showed that replacement of this single motif decreased by 3-fold the ability of Mesp1 to induce *Pdgfra* expression both at the transcriptional and protein level, demonstrating that this binding site is important for enhancer activity mediated by Mesp1 (Fig. 3e-g).

In addition to Mesp1 bHLH binding sites, we found a significant enrichment for binding motifs corresponding to other families of TFs, including Zic, Sox, Tcf/Lef, Oct/Sox/Nanog, Gat a and Fox motifs (Fig. 3c). Not all motifs were equally present in early, constant and late Mesp1 binding peaks, with a significant enrichment of a compound binding site of pluripotency TFs Oct4, Sox2 and Nanog in early peaks and an increased proportion of late peaks containing FOX and GATA motifs (Extended Data Fig. 3a).

Many TFs that have their binding motif enriched in Mesp1 ChIP-seq peaks are themselves upregulated by Mesp1 (e.g. *Gata4, Hand1/2, Meis1/2, Hoxb1, Foxc1/2*) and are known to play roles later during cardiovascular development (Fig. 3h)[25–30]. Moreover, a fraction of the enhancers that become active at 24 hours following Mesp1 expression were not directly bound by Mesp1 (23%, 899/3839), suggesting that additional TFs acting downstream of Mesp1 are activated and induce chromatin remodelling. GATA binding sites were the most enriched motifs in ATAC-seq peaks opened by Mesp1 but not directly bound by Mesp1. Analysis of Gata4 ChIP-seq performed in mesodermal cells derived from mouse PSCs[31] showed that Gata4 binds a fraction of the enhancers bound by Mesp1 (11%), the majority of which were predicted by motif discovery. In addition, Gata4 also bound to many (49%, 444/899) of the enhancers that are opened following Mesp1 induction but not directly bound by Mesp1, most of them (75%) being predicted from our bioinformatic analysis (Extended Data Fig. 3d-f).

We also found 2772 peaks that were associated with a closing of the chromatin and lower levels of flanking $H_3K_{27}Ac$ without presenting Mesp1 binding. These peaks were found in the vicinity of downregulated genes, including pluripotency TFs *Oct4* and *Sox2* and epithelial genes *Cdh1* and *Epcam*. Motif enrichment analysis on these repressed peaks revealed strong enrichment of a Oct4-Sox2-Tcf-Nanog compound motif. Sox2, Oct4 and Nanog ChIP-seq[32] showed that these core pluripotency TFs bound around 60% of enhancers that were indirectly repressed by Mesp1 (Extended Data Fig. 4a-c).

Chromatin landscape and transcriptional regulation at physiological level of Mesp1

To assess whether these findings are relevant in the absence of Mesp1 overexpression, we performed RNA-seq and ATAC-seq on Flk1/Pdgfra double positive and double negative cells at day 4 of PSC differentiation (Figure 4a), which are enriched for endogenous Mesp1 expression during mouse and human PSC differentiation *in vitro* and mouse gastrulation *in vivo*[11, 33, 34] First, we found that 32% of the genes upregulated by Mesp1 were significantly enriched in Flk1$^+$/Pdgfra$^+$ cells. On the other hand, 60% of the genes downregulated by Mesp1 were depleted in Flk1$^+$/Pdgfra$^+$ cells (Figure 4b-d). ATAC-seq on FACS-isolated Flk1$^+$/Pdgfra$^+$ and Flk1$^-$/Pdgfra$^-$ cells showed that 44% of Mesp1-bound sites were opened in Flk1$^+$/Pdgfra$^+$ cells, 50% of which (435/2011) were significantly enriched in Flk1$^+$/Pdgfra$^+$ cells (Figure 4e-g). Motif discovery on all peaks enriched in Flk1$^+$/Pdgfra$^+$ cells revealed strong enrichment of GATA, T-box and Mesp1 motifs (Figure 4h), showing that the same genes, enhancers and their associated TFs were enriched in Mesp1-expressing cells in the absence of Mesp1 overexpression.

Cell context dependency of Mesp1 binding and gene regulation

To define the importance of the cellular context for Mesp1 binding and transcriptional regulation, we induced Mesp1 overexpression in PSCs cultured in 2i medium, a condition that promotes the naïve pluripotent state[35], and performed Mesp1 ChIP-seq, RNA-seq, and ATAC-seq. We found that only 6/1368 (0.4%) of Mesp1 target genes were significantly upregulated 24 hours after Mesp1 overexpression in 2i. In 2i conditions, Mesp1 significantly bound to 13% of its binding sites identified during PSC differentiation, although low level of Mesp1 binding occurred at many of its binding sites. The proportion of primed versus de novo binding sites in 2i was similar than during PSC differentiation. However, there was a significant depletion of late Mesp1 binding sites in 2i conditions. There was a strong correlation between the strength of Mesp1 binding and chromatin opening at these sites. Surprisingly, even when Mesp1 binding and chromatin remodeling are unaffected by 2i conditions, Mesp1 overexpression did not result in gene upregulation (Extended Data Fig. 5a-g). Altogether, these data reveal the importance of the cellular context in Mesp1 binding, chromatin remodeling and transcription regulation mediated by Mesp1.

Zic2 and Zic3 are enriched at Mesp1 bound enhancers

To uncover the TFs that cooperate with Mesp1 to regulate its binding, chromatin remodelling and regulation of target gene expression, we performed TF motif discovery at Mesp1 bound enhancers. The most enriched non-bHLH TF motif in Mesp1 ChIP-seq peaks was the ZIC TF motif (Fig. 3c). As ZIC3 loss of function mutations in mouse and human are associated with gastrulation and congenital heart defects [36–41], we hypothesized that Zic TFs could regulate Mesp1 transcriptional activity and the expression of its downstream direct target genes. Zic2, the closest paralog of Zic3, is also broadly expressed during gastrulation[42] and was recently shown to promote CP specification and differentiation in human PSCs *in vitro*[43].

qRT-PCR and western blot showed that Zic2 and Zic3 are expressed during PSC differentiation *in vitro*, at the same time as *Mesp1* (Fig. 5a, Extended Data Fig. 6a). Single-molecule RNA fluorescent *in situ* hybridization (smRNA-FISH) in gastrulating mouse

embryos showed that *Zic2* and *Zic3* were both co-expressed with *Mesp1* in the prospective cardiogenic mesoderm at E6.75 (mid-streak stage) and E7.25 (early bud stage). At E7.25, *Zic2* was expressed at higher levels in the anterior side of the embryo, whereas *Zic3* was more abundant in the primitive streak (Fig. 5b).

To validate our bioinformatic prediction that Mesp1, Zic2 and Zic3 are binding to Mesp1-bound enhancers, we performed ChIP-seq of endogenous Zic2 and Zic3 proteins in the absence of Mesp1 expression (at day 2.5 of differentiation), at the onset of endogenous Mesp1 expression and 24 hours following Mesp1 overexpression (Fig. 5c). The high quality and specificity of the peaks were attested by strong enrichment of Zic motifs in these peaks (Extended Data Fig. 6b). Zic2 and Zic3 bound a significant fraction of Mesp1-bound enhancers (19% for Zic2 and 29% for Zic3), in good accordance with the bioinformatic prediction (Figure 5d-e). Out of these, 95% of Mesp1 peaks bound by Zic2 were also bound by Zic3 (Extended Data Fig. 6c). Among the Mesp1 and Zic3 co-bound peaks, 26% (153/577) were already bound by Zic3 before Mesp1 expression (day 2.5) and were also present at day 3.5, irrespective of Mesp1 overexpression (group 1). In group 1, 127/153 (83%) of Zic3/Mesp1 co-binding sites have their chromatin opened before Mesp1 binding (primed peaks) whereas the other 17% had their chromatin opened following Mesp1 overexpression (Fig. 5d-e, Extended Data Fig. 6d). On day 3.5, at the beginning of endogenous Mesp1 expression, 14% (79/577) of the Mesp1 and Zic3 co-bound peaks were bound by Zic3 without being associated with chromatin opening. Following Mesp1 overexpression, these chromatin regions were strongly opened (group 2). Finally, 345/577 (60%) of Mesp1/Zic3 co-bound peaks were bound by Zic3 and presented chromatin opening only following Mesp1 overexpression (group 3) (Extended Data Fig. 6d). Altogether, these data show the temporality and dynamic nature of the distinct groups of Mesp1 and Zic3 co-binding sites and their impact on chromatin remodeling.

The binding of Zic2, Zic3 and Mesp1 to the same enhancers suggests that Mesp1 and Zic2/3 may physically interact. To assess this possibility, we performed co-immunoprecipitation using antibodies against endogenous Zic2 and Zic3 proteins, followed by western blot revealing the presence of Mesp1-3HA protein. We found that both Zic2 and Zic3 physically interact at the protein level with Mesp1 (Fig. 5f).

Zic3 regulates Mesp1 transcriptional activity at a subset of mesodermal enhancers

To assess whether Zic2 and Zic3 regulate Mesp1 function, we deleted *Zic2* or *Zic3* using CRISPR/Cas9n in PSCs allowing dox-induced Mesp1 overexpression in the context of Zic2 or Zic3 KO. Although Zic2 and Zic3 have been proposed to regulate pluripotency[44, 45], we did not find changes in the expression of core pluripotency TFs when culturing *Zic2* and *Zic3* KO PSCs in Lif+2i medium (Extended Data Fig. 7a). We then assessed whether *Zic2* or *Zic3* deletion affects CP specification and differentiation during PSC differentiation *in vitro*. Mesp1 CPs are characterized by co-expression of Flk1 and PDGFRa *in vitro* and *in vivo*[11, 13, 33, 34]. Upon Mesp1 overexpression, the number of Flk1[+]/PDGFRa[+] CPs was slightly decreased in *Zic2* and *Zic3* KO PSCs at day 4 and in *Zic3* KO PSCs day 5 (Fig. 6a). These data suggest a role for Zic3 in CP specification. We then assessed the impact of *Zic2* and *Zic3* KO on cardiac differentiation. Following Mesp1 overexpression, the number

of CMs was slightly decreased *Zic3* KO cell lines, but no significant difference was found in *Zic2* KO cells (Fig. 6b, Extended Data Fig. 7b).

To unravel the mechanisms by which Zic TFs regulate Mesp1 transcriptional activity, we performed RNA-seq on *Zic2* and *Zic3* KO cells 24 hours after Mesp1 overexpression. Single *Zic2* KO cells only presented 11 significantly differentially expressed genes following Mesp1 expression compared to WT cells. In contrast, 15% (87/588) of direct upregulated Mesp1 target genes were significantly downregulated in *Zic3* KO cells following Mesp1 overexpression, including several genes known to be important for specification of the different heart lineages (*e.g. Aldh1a2*, *Hoxb1*, *Aplnr*, Tenm4)[46–51] (Fig. 6c-e). Moreover, several genes controlling pluripotency and other cell lineages (*e.g. Nodal, Nanog, Sox2, Foxa2*) were upregulated in *Zic3* KO cells (Extended Data Fig. 7c). These data show that Zic3 alone regulates the expression of some of Mesp1 target genes.

Mesp1 and Zic3 co-regulate gene expression during mouse gastrulation

To assess the *in vivo* relevance of the cooperation between Mesp1 and Zic3 in regulating chromatin remodelling, enhancer activity and gene expression during CP specification, we investigated the expression of selected Mesp1 and Zic3 direct target genes *in vivo* in *Mesp1* and *Zic3* KO embryos. Among these genes, we selected a pan-marker of CPs, *Aplnr*[49–51] as well as *Alhd1a2* and *Hoxb1*, two known markers of the posterior SHF, a subpopulation of CPs that specifically contributes to the outflow region, atria and venous pole of the heart[3, 47, 48]. *Aldh1a2* and *Hoxb1* correspond to late Mesp1 target genes and *Aplnr* is an early upregulated Mesp1 target gene. smRNA-FISH of these three genes in WT embryos at E7.25 demonstrated that they co-localize with *Mesp1* within the mesodermal cells leaving the primitive streak and migrating towards the anterior pole of the embryo (Fig. 6f-h). In *Mesp1*-null embryos, the expression of these genes was strongly downregulated (Fig. 6i-k), showing that they are *bona fide* Mesp1 target genes *in vivo*.

To assess whether Zic3 also controls the expression of these genes *in vivo*, we examined their expression on gastrulating *Zic3*-null embryos. To this end, we generated *Zic3* KO mouse lines with CRISPR/Cas9 by injecting Cas9 protein and guidRNA directly into zygotes (Extended Data Fig. 8a). We obtained several founders with either big deletion or insertion, thereby disrupting the *Zic3* gene within its first exon, upstream of the DNA-binding domain coding sequence. The phenotypes of these novel *Zic3* KO alleles were identical to the previously reported *Zic3* null mice, with defects similar to the human phenotype associated with ZIC3 mutations including heterotaxy and exencephaly in a fraction of the *Zic3* mutant mice[39–41, 52]. Closer phenotypic analysis showed that *Zic3* KO embryos present persistent truncus arteriosus (1/32), situs inversus (2/32), hypoplasia of the right ventricle (5/32). Moreover, *Zic3* KO embryos showed a thinner compact myocardial layer at E14.5, in good accordance with the decrease in CM differentiation found in *Zic3* KO PSCs. On the other hand, the endocardial layer was unaffected in *Zic3 KO* hearts (Extended Data Fig. 8b-f). smRNA-FISH of *Hoxb1, Aldh1a2* and *Aplnr* in Zic3 KO embryos showed that these genes were downregulated during gastrulation in *Zic3* KO embryos (4/5 for *Hoxb1*, 5/8 for *Aldh1a2* and 9/11 for *Aplnr*) (Fig. 6l-n). These data demonstrate that Mesp1 and Zic3 co-regulate the expression of key cardiac genes *in vivo* during mouse gastrulation.

Zic3 and Zic2 co-regulate Mesp1 functions at mesodermal enhancers

As Zic2 binds a large fraction of Mesp1/Zic3 bound enhancers, we assessed whether *Zic2* could compensate for the loss of *Zic3* during Mesp1-induced CP specification and differentiation from PSCs *in vitro*. To this end, we generated *Zic2/3* double KO (dKO) cell lines in Mesp1-inducible PSCs. Western blot showed the absence of Zic2 and Zic3 protein expression in this KO cell line (Extended Data Fig. 6a). These cells were able to grow in 2i+Lif medium and sustain the expression of core pluripotency TFs (Extended Data Fig. 9a). The specification of F1k1$^+$/PDGFRa$^+$ CPs and their terminal differentiation into CMs was completely suppressed in *Zic2/3*dKO PSCs (Fig. 7a-b and Extended Data Fig. 9b).

To define which Mesp1 target genes and enhancers are co-regulated by Mesp1 and Zic2/3, we performed RNA-seq on *Zic2/3*dKO cells 24 hours after Mesp1 induction. Strikingly, 53% (310/588) of all up-regulated Mesp1 direct target genes and 49% (675/1368) of direct and indirect Mesp1 up-regulated genes were strongly downregulated following Mesp1 overexpression in *Zic2/3*dKO cells, including many genes that are essential for the specification and differentiation of cardiovascular lineages (Fig. 7c-f). Among the genes whose enhancers are co-bound by Mesp1 and Zic3, 199/319 (62%) presented a decrease in their expression following *Zic2/3*dKO, further demonstrating the functional importance of Zic2/3 in transcriptional regulation mediated by Mesp1.

We then assessed whether the major transcriptional defects after Mesp1 induction in *Zic2/3*dKO cells are caused by loss of Mesp1 binding to its enhancers (Fig. 7g-h). Mesp1 ChIP-seq in *Zic2/3*dKO cells showed that Mesp1 binding was decreased at 83% Mesp1 binding sites, enriched for the late Mesp1-binding peaks (Fig. 7g, Extended Data Fig. 9c, d).

Chromatin profiling of *Zic2/3*dKO PSCs following Mesp1 overexpression by ATAC-seq revealed a major defect of Mesp1-induced chromatin remodelling, with the signal of 56% (2130/3839) of the enhancers remodelled following Mesp1 overexpression and 48% of Mesp1-bound enhancers being strongly decreased (Fig. 7g-h, Extended Data Fig. 9e). Many of these enhancers were still bound by Mesp1 in the absence of Zic2/3, but their chromatin regions were not anymore opened by Mesp1 binding (group 3), showing that Zic2/3 are important for regulating chromatin opening mediated by Mesp1. Motif discovery on the downregulated 3558 ATAC-seq peaks in *Zic2/3*dKO cells revealed a strong enrichment of the Mesp1 motif, as well as Gata and Zic motifs (Extended Data Fig 9f).

To assess the cell-autonomous role of Zic2 and Zic3 in the regulation of Mesp1 mesoderm specification *in vivo*, we performed CRISPR/Cas9n deletion of *Zic2* and *Zic3* in Tomato-expressing PSCs, then injected these cells in WT embryos of the same genetic background, and analysed the chimeric embryos when the heart is formed, at E9.5 (Fig. 7i). Out of two litters of E9.5 chimeric embryos, we analysed 3 embryos with a significant percentage of chimerism. The contribution of WT tdTomato+ cells in similar chimeric experiments to the different embryonic lineages showed no difference in their differentiation potential compared to WT cells[53, 54]. In chimeric embryos with Zic2/3 double KO (tomato positive) that presented at least 5% of Tomato+ cells, we found a decreased contribution of tdTomato+/Zic2/3 double KO cells to the heart. While the chimerism of TdTomato+ cells in the embryo are 5.2+/- 0.3%, only 1.4+/- 0.6% of tdTomato+ cells

are found in cardiomyocytes, showing a significant decrease of Zic2/3 double KO cells in cardiomyocytes. The contribution to the endocardium was more variable, precluding to draw strong conclusions due to the low number of chimeric mice analysed (Fig 7i-k). These data further support a role of Zic2/3 during specification of cardiac progenitors and cardiomyocyte differentiation. Altogether, these data indicate that Zic2 and Zic3 act redundantly to regulate Mesp1 binding and Mesp1-induced chromatin remodelling, which are crucial for cardiac progenitor specification and differentiation.

Discussion

During mouse gastrulation, cells transit from a pluripotent state to a lineage committed state within 24 hours. The dramatic changes in gene expression accompanying these cell fate transitions are mediated by rapid and precise coordinated action of multiple TFs. Mesp1 acts as a master regulator of the specification and differentiation of cardiovascular lineages during embryonic development. By using Mesp1 ChIP-seq, $H_3K_{27}Ac$ and H_3K_4me1 ChIP-seq, ATAC-seq and RNA-seq, we demonstrated that Mesp1 induces rapid opening of initially closed chromatin regions within the regulatory elements of key cardiovascular genes, a defined characteristic of pioneer TFs [24, 55–57], which enables the recruitment other TFs and chromatin remodelling factors, leading to activation of gene expression[58]. Beside few exceptions [59–63], the co-factors that are required to regulate TF pioneer activity remain unknown.

Using bioinformatic predictions and functional validation by loss of function experiments *in vitro* and *in vivo*, we identified Zic3 and Zic2 as essential transcriptional cofactors that regulate Mesp1 TF activity and the expression of 50% of its direct cardiovascular target genes. Although CRISPR/Cas9 deletion of *ZIC2* in human PSCs leads to defective CM differentiation, no heart anomalies are found in *Zic2* deficient mice except a flatter outflow tract [43]. *ZIC3* deficiency causes X-linked heterotaxy, a syndrome in which organs present defects in their lateral positioning in both mouse and humans[36, 37] *Zic3* mutants present variable heart defects (about 50%) ranging from septal defects, conotruncal anomalies, and other outflow tract defects[38, 41]. In addition to the heterotaxy syndrome, *Zic3* null embryos present other gastrulation defects, ranging from failure to gastrulate to excess of mesoderm formation or axis duplication [40]. The reason for early gastrulation and LR asymmetry defects in *Zic3* KO remains unclear, as deletion of *Zic3* prior to gastrulation but not in CPs or in the node, cause heart malformations[64], suggesting that *Zic3*-associated defects are secondary to early anomalies occurring in the early stage of gastrulation. Our Mesp1, Zic2 and Zic3 ChIP-seq data demonstrate that Mesp1 and Zic TFs bind a common set of enhancers, many of which are in the regulatory regions of key genes of CP specification and differentiation.

Although the transcriptional activity of Mesp1 was reduced at key genes in *Zic3* KO PSCs, it was rarely abolished, suggesting that other Zic TFs could partially compensate for the absence of *Zic3*. Consistent with a genetic compensation between Zic2 and Zic3 during mesoderm formation, deletion of both *Zic2* and *Zic3* in Mesp1-inducible PSCs completely abolished the specification and cardiac differentiation of CPs following Mesp1 overexpression. RNA-seq, Mesp1 ChIP-seq and ATAC-seq of these double KO cell lines

demonstrate that Mesp1 overexpression cannot anymore activate many of its key enhancers that regulate the expression of many essential cardiovascular genes in *Zic2/3* double mutants. These data demonstrate that Zic3 and Zic2 act redundantly and are essential for cardiac mesoderm formation by regulating the TF activity of Mesp1. Consistent with an early role of Zic2/3 in regulating Mesp1 mesoderm specification, our chimeric embryo experiments between WT and Zic2/3 double KO show that cardiomyocytes were decreased in *Zic2/3* KO cells.

Mechanistically, Zic2 and Zic3 physically interact with Mesp1. Zic2 and Zic3 bind a significant number of Mesp1 bound enhancers prior to Mesp1 expression. In the absence of Zic2/3, Mesp1 present a decrease in its binding affinity at 87% of its binding sites, showing that Zic2/3 are required for Mesp1 binding to a subset of its target genes. In addition, whereas Mesp1 can bind some of its binding sites in the absence of Zic2/3, Mesp1 cannot anymore open the chromatin regions and upregulate gene expression, showing that Zic2/3 are important to regulate the pioneer activity of the transcriptional complex induced by Mesp1. Zic2/3 are also recruited by Mesp1 expression at 60% of Zic/Mesp1 co-bound enhancers.

Finally, at a minor proportion of Mesp1 binding sites, Zic2/3 do not control Mesp1 binding and chromatin opening but promote transcriptional activation, illustrating the different mechanisms by which Zic2/3 control the transcriptional activity of Mesp1. Our data uncover a novel role of Zic2/3 in regulating Mesp1 functions by dynamically controlling Mesp1 binding to its direct target genes, its ability to open the chromatin at key mesodermal enhancers and the timing of Mesp1 transcriptional regulation. Further studies will be important to define whether other TFs or chromatin regulators cooperate with Mesp1 to control cardiac mesoderm specification and differentiation and to assess whether Zic TFs also control the transcriptional activity of other master lineage-specific TFs during development.

Methods

Tetracycline-inducible PSC lines

A Mesp1-3HA stable transgenic PSC line was generated as described previously[67].

CRISPR/Cas9 Knockout PSCs

CRISPR/Cas9 knockout PSCs for *Zic2, Zic3*, and *Zic2/Zic3* together were generated following the protocol previously described [67, 68]. Briefly, two pairs guide RNAs for each gene that target two regions of the gene separated from 364bp to 2381bp were designed using a CRISPR/Cas9 online tool (Benchling [Biology Software](2019-2020), retrieved from https://benchling.com; Supplementary Table 1) and the vectors pX330-U6-Chimeric_BB-CBh-hSpCas9 (PX330) and pSpCas9n (BB)-2AGFP (PX461) were obtained from Addgene (#42230 and #48140). The guide RNAs were cloned into a modified short version of PX330 in which the Cas9 cassette was removed. Four plasmids containing the guide RNAs for Single KO or eight Plasmids for double KO were co-transfected together with PX461 (Cas9n) into Mesp1-3HA PSCs using Lipofectamine 2000 (Thermofisher,

11661089), in order to generate a big deletion either upstream or encompassing the DNA-binding domain of the targeted gene(s). Transfected GFP-positive cells were FACS-sorted individually into 96-well plates 48 hours after transfection using a FACSAria (BD Biosciences). After 7 to 12 days, colonies were passaged and then screened by PCR using primers flanking the expected deletion site (Supplementary Table 2). Zic2 and Zic3 single and double KO clones were selected based on homozygous PCR profiles and confirmed by Sanger sequencing.

CRISPR/Cas9 enhancer knockout PSCs

The same strategy as above was used to generate enhancer KO cell lines. Eight gRNAs were inserted into the shortened version of PX330 in order to generate deletions between 200 and 480 base pairs, centred around a Mesp1 ChIP-seq peak and encompassing the central bHLH motifs of the peak (Supplementary Table 1). At least two independent clones per peak for five enhancers were generated, targeting enhancers for *Myocd*, *Pdgfra*, *Hoxb1* and *Hand1*.

Motif replacement in the Pdgfra proximal Mesp1-bound enhancer

Two guide RNAs overlapping with the targeted motif (CCATTTG) were separately cloned into PX330 containing wildtype Cas9 (Supplementary Table 1). Single strand DNA containing 40 bp homology arms on each side of the motif, which as modified into CGCTAGC, a NheI restriction site, was synthetized by Eurogentec. Mesp1-inducible PSCs were transfected with PX330 containing the gRNA and the ssDNA template, then FACS sorted into single cells based on GFP as above. Selection of clones was performed using PCR for amplification of the whole region, followed by NheI digestion. Homozygous and precise editing of the motif was validated by Sanger sequencing.

PSC Culture and Differentiation

Mesp1-3HA PSCs were cultured as previously described[12, 13]. After generating KO cell line by CRISPR/Cas9n, PSC cell lines were cultured on feeder free condition with LIF/2i medium (ES medium supplemented with 1 μM PD0325901 (Sigma, PZ0162) and 3 μM Chir99021 (Sigma, SML1046). Differentiation was performed in hanging drops as for Mesp1-3HA PSCs.

RT-qPCR during PSC differentiation

RT-qPCR were performed as described previously[12, 13]. qPCR primers were listed in Supplemental Table 3.

RNA sequencing

RNA extraction was performed by using Qiagen RNeasy Micro Kit according to the manufacturer's instructions. Before sequencing, quality of the RNA was evaluated using a Bioanalyzer 2100 (Agilent). Indexed cDNA libraries were obtained using Ovation Solo RNA-seq Systems (NuGen) following the manufacturer's recommendations. The multiplexed libraries were loaded onto flow cells and sequences were produced using a HiSeq PE Cluster Kit v4 and TruSeq SBS Kit v3-HS (250 cycles) on a HiSeq 1500 (Illumina).

ChIP Sequencing

ChIP was performed as described previously on Mesp1-3HA-flagged Dox-inducible PSC lines[12]. Briefly, EBs were collected at 12h and 24h after induction of dox (working concentration 1μg/ml), fixed directly with 1% formaldehyde for 7 min at RT, and quenched with 0.125M glycine for 5 min. CHIP was performed according to the manufacturer's instructions (CHIP-IT express kit) with antibodies (Supplementary Table 4). 2-10 ng of pulled-down DNA was used to construct the sequencing library by using Truseq ChIP Library Preparation kit (Illumina) or NEB Next Ultra II DNA library Prep Kit for Illumina (NEB) according to manufacturer's instructions, and subsequently sequenced on a HIscanSQ module (Illumina).

ATAC Sequencing

ATAC-seq was perform following standard protocol[69]. Briefly, EBs were collected at the same time points and conditions than the RNA-seq datasets, rinsed with dPBS (ThermoFisher, 14190144) and dissociated with accutase (Sigma, A6964). 125,000 cells or FACS sorted cells were used to perform the ATAC-seq, and ATAC-seq library were performed according to manufacturer's instructions (Nextera DNA sample Preparation Kit, Illumina), and size selection from 200bp to 800bp was performed by Ampure XP beads (Beckman) before NGS sequencing.

Co-immunoprecipitation

Co-immunoprecipitation was performed according to a previously published protocol[70]. 3μg rabbit Igg control antibody, rabbit anti-Zic3 or rabbit anti-Zic2 (Supplementary Table 4) was added into same amount of protein (around 1mg in 300μl lysis) and rotated at 4 °C overnight, followed by addition of Dynabeads™ Protein G (ThermoFisher Scientific, 10003D) at 4°C for 4 hours. Washes were performed with NETN buffer (20mM Tris (pH 8), 1mM EDTA,900mM NaCl, 0.5% CA-630).15μl eluted samples were used to perform the western blot by mouse anti-HA antibody (Supplementary Table 4).

Western blot

Western blots were performed as previously described[67]. Antibodies were listed in Supplementary Table 4.

Flow Cytometry

Flow cytometry was performed as previously described[13]. Antibodies used are reported in Supplementary Table 4.

Immunofluorescence Analysis

Immunofluorescence on EBs were performed as previously described[13]. Antibodies used are reported in Supplementary Table 4. Images were acquired on a Zeiss Axio Imager with a Zeiss Axiocam MRn camera and using the Axiovison Rel. 4.6 Software.

On E14.5 hearts, standard histological procedures were used[71]. Antibodies used are reported in Supplementary Table 4. Images were acquired on a AxioZoom.V16 microscope (Zeiss)

and the thickness of the cardiomyocyte layer was measured with the Fiji software. Phenotypic analysis of *Zic3* mutant and WT hearts at E14.5 was performed with a careful stage matching between controls and KO embryos (based on the shape of the limbs).

Mouse lines

Mesp1–Cre mice were previously obtained from Y. Saga[9]. Mice colonies were maintained in certified animal facilities in accordance with European guidelines (7h to 19h light cycle, 20-25°C and 55% +/-15 humidity). The experiments were approved by the local ethical committee (CEBEA) under protocols #591N (CB) and Apafis #13031 (FL) and a national agreement (#B1301308) (FL). Novel *Zic3* KO mouse lines were generated by injecting two crRNAs targeting sequences flanking the DNA-binding domain of Zic3, along with tracrRNA and recombinant Cas9 protein, into zygotes (IDT). Offspring were screened for genomic aberrations within the *Zic3* gene. Two founders were selected to conduct further experiments; one containing a large deletion and the other a large insertion within the Zic3 gene, both leading to disruptions of the open reading frame (ORF) upstream of its DNA-binding domain encoding sequence.

Phenotypic analysis of *Zic3* KO embryos

Embryos collected at E14.5 were fixed in 4% paraformaldehyde and washed in PBS. Images of the heart and embryo were acquired on a AxioZoom V16 microscope (Zeiss). The areas of the RV and LV were measured with the Fiji software and a ratio was then calculated.

smRNA-FISH experiments

smRNA-FISH was performed according to the protocol of the RNAscope Multiplex Fluorescent v2 Assay (ACD-bio. 323110). In brief, embryos were fixed for 26-30h in 4%paraformaldehyde at 4°C and then dehydrated in methanol. For smRNA-FISH on sections at E6.5-E7.5, the decidua was embedded in paraffin after fixation, dehydrated in methanol and incubated for 16h in butanol at 4°C. For smRNA-FISH on sections of E9.5 chimera, the embryos were embedded in histogel (Epredia HG-4000-012) prior to paraffin embedding. Tissue sections were cut at 7 to 10 μm. Whole-mount smRNA-FISH was performed as previously described [72]. Probes are reported in Supplementary Table 5. To develop probes, we have used Opal dyes from Akoya Bioscience (Opal-520, Opal-570 and Opal-650, from 1:100 to 1:200). Embryos or sections were imaged using an AxioZoom.V16 microscope (Zeiss) or an LSM800 confocal microscope (Zeiss). When analysing litters from *Mesp1-Cre* or *Zic3* KO lines, embryos were genotyped by PCR after imaging.

Image analyses after smRNA-FISH experiments

To analyse the expression of *Mesp1* expression in EBs, images were analyses with the Fiji software (v2.1). We used guidelines from ACD-bio on how to quantify RNAscope Fluorescent Assay Results. First, the *Average Background Intensity (ABI)* was calculated based on $\frac{\sum integrate\ intensity\ of\ selected\ background\ regions}{\sum area\ of\ selected\ background\ regions}$. To measure signal intensity (SI), cells where *Mesp1* expression was detected were selected (ROI). SI = Total intensity of ROI — ABI × Total Area. To quantify the percentage of cells expressing *Mesp1*, we also selected pixels with SI>ABI. We used the DAPI channel to measure the surface area (SA) of the EBs.

Similarly, in another channel, we estimated the SA of regions expressing *Mesp1* transcripts. We used the ratio between the SA of *Mesp1* + pixels over the SA of the nuclei staining (DAPI+) to estimate the percentage of cells within the EBs that express *Mesp1*.

To estimate the percentage of chimerism in *Zic2/3*dKO cells in WT embryos, images from RNAscope experiments were analysed. We estimated the *ABI* for each channel as described above. We then selected pixels with SI>ABI. All positive pixels were selected, and their SA measured. We used the DAPI channel to measure the SA of the embryo in a selected section. Similarly, in another channel, we estimated the SA of regions expressing *tdTomato*. We used the ratio between the SA of *tdTomato+* pixels over the SA of the nuclei staining (DAPI+) to estimate the percentage of chimerism in the embryo. To estimate the percentage of chimerism in cardiomyocytes, we selected the region with *Tnnt2*+ cells. To estimate the percentage of chimerism in non-cardiomyocytes heart cells, we selected the cardiac region and subtracted the area with *Tnnt2* expression.

Chimeric embryo generation with Zic2/3 KO TdTomato+ cells

TdTomato-expressing mouse PSCs were generated as previously described[53], and maintained in Lif/2i conditions[35]. Zic3 knockout was performed using exactly the same strategy with same targeting guide RNAs used for knocking out *Zic3* in Mesp1-3HA inducible PSCs (Supplementary Table 1). Then, Zic2 knockout was performed on this Zic3 KO TdTomato-expressing PSC line. Three Zic2/Zic3 KO clones were selected, and their genotype was confirmed using PCR and sequencing. These PSCs were injected into WT blastocysts in 6 rounds, chimeras were harvested at E7.5, E9.5 and E13.5. No tdTomato+ cells were found at E13.5. tdTomato+ cells were analysed at E9.5 in three embryos, which were sectioned for RNAscope experiments.

Bioinformatic analysis of bulk RNA-seq

Raw sequenced reads were trimmed using Trimmomatic (v0.39) for quality of bases and eliminating sequencing adapters[73]. Cleaned FastQ files were aligned to the mouse genome using STAR (v2.7.3)[74]. Read counts were generated for all refSeq genes using HTSeq-count (v0.11.3)[75]. Counts were normalized for all samples as reads per million of mapped reads for each gene. Only genes that had an expression > 1 read per million for both duplicates of one or more conditions were kept for further analysis.

Differentially expressed genes were defined with DESeq2 (v1.34) using a cut-off adjusted p-value of 0.05 and a minimum of 1.5 fold-change expression between both conditions [65]. Dox and no dox conditions were compared separately at 12 hours and 24 hours, then these two lists of differentially expressed genes were pooled together to define their kinetics of expression. Genes with a standard deviation higher than 50% of the mean expression were further eliminated. Genes were first classified as up- or downregulated by using a threshold of 1.5 fold-change $\left(\frac{\text{mean dox}}{\text{mean no dox}} > 1.5 \text{or} < \frac{1}{1.5} \text{respectively} \right)$. To classify upregulated genes into early, constant and late activated, we calculated the slope of increased expression from 0 to 12h (a) and from 12 to 24h (b) in dox and no dox conditions. The resulting slopes $a = a_{dox} - a_{no\ dox}$ and $b = b_{dox} - b_{no\ dox}$ were calculated to take into account endogenous patterns of gene expression. Genes were defined as early if $a/b > 3$, late if $b/a > 3$, and

constant when a/b was intermediate. If the two slopes went in opposite direction, the slope with highest absolute value was preferred to the other, and the gene was defined either as early (a>0 and b<0) or late (a<0 and b>0). Basic plots were constructed using the R package ggplot2[76]. When comparing Mesp1-inducible WT cells to KO cells for Zic2, Zic3 or both, we used DESeq2 to call differentially expressed genes between cell types and added a cut-off fold-change value of 2 [65]. For comparing Flk1$^-$/Pdgfra$^-$ with Flk1$^+$/Pdgfra$^+$ cells, DESeq2 was used in addition to a 1.5-fold change threshold to call differentially expressed genes. The same was used for RNA-seq samples collected in 2i, taking into account only genes expressed at least 1 read per million.

Analysis of microarray data in Mesp1-expressing cells *in vivo*

Genes up-regulated in GFP+ (Mesp1-positive) expressing cells *in vivo* were defined by microarray as genes presenting at least 2-fold change stronger signal in E6.75 or E7.25 GFP+ versus GFP-. We then classified these genes as early, constant and late depending on the ratio of expression between E6.75 and E7.25 GFP+ cells. Genes with at least 1.5-fold higher expression at E6.75 were classified as early, whereas genes with at least 1.5-fold higher expression at E7.25 as late. All intermediate genes were classified as constant.

Analysis of Mesp1 ChIP-seq and H$_3$K$_{27}$Ac ChIP-seq

Raw sequenced reads were trimmed using Trimmomatic for quality of bases and eliminating sequencing adapters [73]. Single-end (replicate 1) or paired-end (replicate 2) sequencing reads were uniquely aligned to the mouse genome using Bowtie2 (v2.4.2)[77]. Picard tools were used to remove PCR duplicates (Broad Institute), samtools (v1.2.10) to remove mitochondrial reads and low-quality alignments[78]. Peaks were called using MACS2 (v2.4.2) with a threshold of p=10^{-10} and using the input sequencing as background[79]. Mesp1 ChIP-seq peaks from 12 and 24 hours were compared between both replicates, and only the intersection of both were kept for further analysis. Then, peaks from both time points were merged in a single file for annotation, motif analysis and read counting. Reads in all peaks were counted using HTseq-count[75]. Mesp1 ChIP-seq peaks were classified as early binding if average nreads(12h)/nreads(24h) > 2, late if this ratio < 0.5, and constant for all intermediate values. To annotate peaks to genes, we used GREAT (v4.0.4)[80] with a maximum distance of 500 kb to the TSS. Ngsplot (v2.63) was used for data visualization[81].

Analysis of H$_3$K$_4$me1, Mesp1 ChIP-seq in 2i and in Zic2/3 double KO cells

FastQ files were cleaned and trimmed using Trimmomatic[73]. Paired-end reads were aligned using Bowtie2[77]. Picard tools were used to remove PCR duplicates (Broad Institute), samtools to remove mitochondrial reads and low-quality alignments[78]. Peaks were called using MACS2[79] with a threshold p-value of 10^{-10}. Where duplicates available (all the Mesp1 ChIP-seq samples), the intersection peaks of both replicates were kept for further analysis. For Mesp1 ChIP-seq samples, peaks were then compared to those obtained in Mesp1 ChIp-seq during differentiation using BEDtools[82].

Analysis of ATAC-seq

FastQ files were cleaned and trimmed using Trimmomatic[73]. Paired-end reads were aligned using Bowtie2 [77]. Picard tools were used to remove PCR duplicates (Broad Institute), samtools to remove mitochondrial reads and low-quality alignments[78]. Peaks were called using MACS2[79] with a threshold p-value of 10^{-10}. Peaks from all experiments from the WT cell line were merged, and reads were counted for each condition in all merged peaks using HTSeq-count[75]. Peaks that were up-or downregulated in dox versus no dox were defined with DESeq2, using an adjusted p-value of 0.05 as threshold[65]. Peaks were annotated with the same parameters as above using GREAT[80]. To define the peaks that were upregulated in dox conditions without being bound by Mesp1, we excluded all Mesp1 ChIP-seq peaks with a p-value cut-off of 0.1 from the file of upregulated peaks.

To define *de novo* versus primed Mesp1-bound enhancers, we performed a second round of peak calling in our ATAC-seq samples at 24 hours dox and 0 hours using a q-value cut-off of 0.05. *De novo* peaks were defined as Mesp1 ChIP-seq peaks not presenting a peak at 0 hours but presenting one at 24 hours dox. Primed peaks encompassed peaks that presented an ATAC-seq peak at 0 hours and undefined Mesp1 peaks, which did not present a peak neither at 0 nor 24 hours dox. Nucleosome eviction after Mesp1 induction was also assessed using HHMRATAC (v1.2.10)[66], a tool for assessing nucleosome positioning in ATAC-seq data. We quantified the overlap between Mesp1 ChIP-seq peaks and nucleosome-bound DNA, as predicted by HHMRATAC (v1.2.10).

For the analysis of ATAC-seq performed in 2i and Zic2/3 dKO cell lines, reads in peaks were counted as above, then up-and downregulated peaks were defined using DESeq2 with a threshold of $p_{adj}<0.05$ and minimum fold-change = 2. For Flk1^{-}/Pdgfra^{-} and Flk1^{+}/Pdgfra^{+} samples, DESeq2 was used with a cut-off at p=0.05 and a minimum fold-change average read enrichment of 2.

Footprinting and bHLH motif quantification

ATAC-seq footprints were defined within Mesp1 ChIP-seq peaks using HINT-ATAC (v0.13.1)[83] on merged ATAC-seq data at 24h dox. The number of occurrence of each possible bHLH motif was quantified within these footprints and represented as a barplot.

Motif discovery and analysis

Different sets of peaks were analysed for known and de novo motif enrichment of TF motifs using Homer (v4.11.1)[22]. Algorithm parameters were set to find motifs of 6, 8, 10 and 12 nucleotides in a region of 500 bp around the peak centre. Homer was also used to define which peaks within a set of peaks contained a particular motif.

Analysis of Zic2 and Zic3 ChIP-seq

Paired-end sequencing reads were treated as for the ATAC-seq for trimming, filtering and alignment. Peaks were called with a cutoff at q=0.05, since these ChIP-seq were generated using endogenous-tagging antibodies. For Zic3 ChIP-seq, replicates were performed and only the peaks present in the intersection of both replicates of the same condition were kept for further analysis. Motif enrichment was performed as above.

Statistics and reproducibility

For pluripotent stem cell work, results shown and statistical tests come from at least three biologically independent experiments performed in at least one cell line for WT cells and two cell lines for CRISPR/Cas9 KO cell lines. Analysis and statistical testing of next-generation sequencing results were performed in 2 biological independent experiments unless stated in the legends. At least three independent embryos of matching stages were analysed for statistical testing for each of the *in vivo* experiments shown and quantified here. Statistical tests were chosen and performed accordingly to data structure (categorical versus quantitative, normal versus non-normal distribution).

For all statistical analyses, data were obtained from a minimum of three independent experiments unless specified in the legends. Details of replicate numbers, quantification and statistics for each experiment are specified in the figure legends.

Extended Data

a

b

c

d

e

f
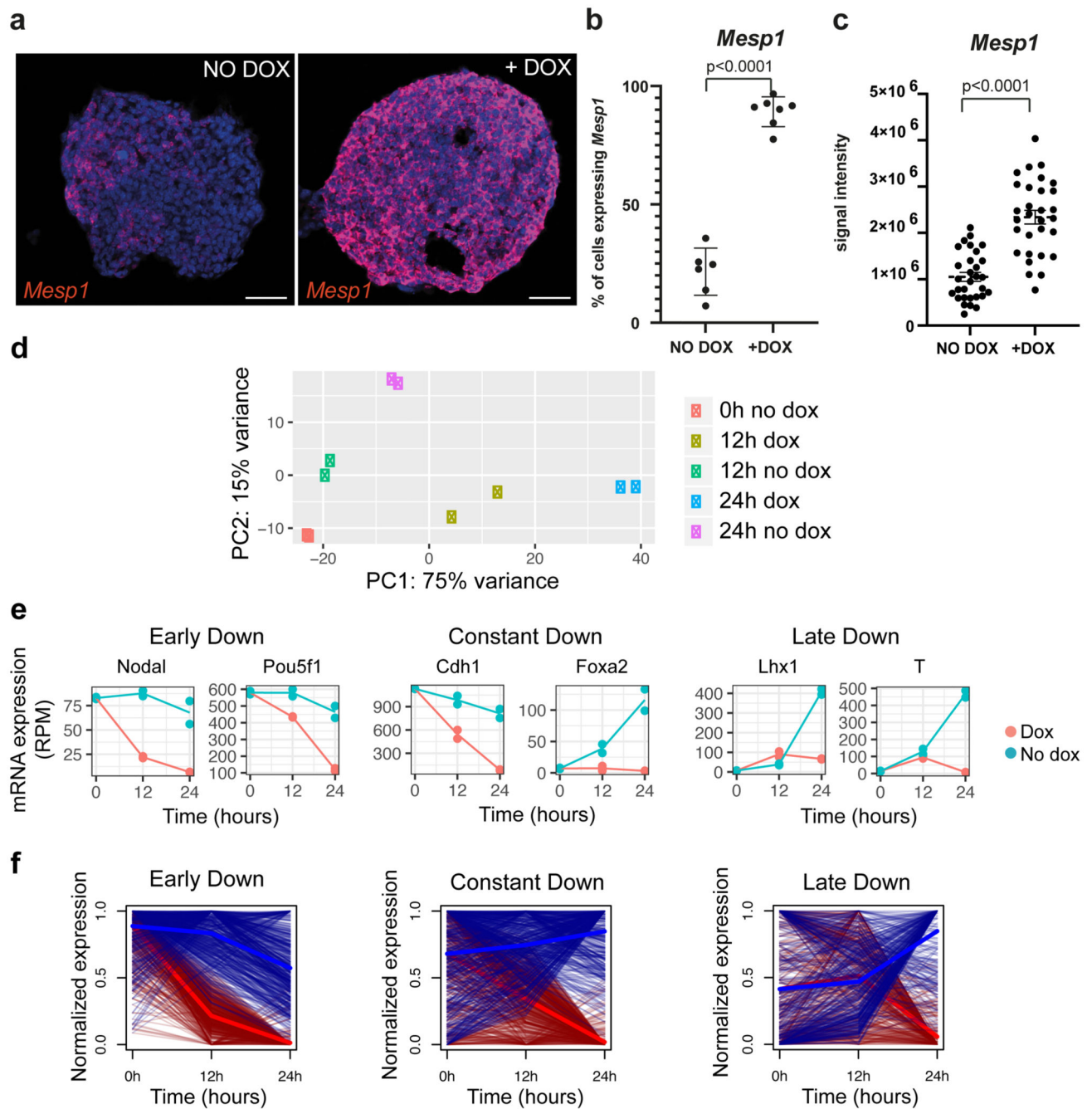
Extended Data Fig. 1. Temporal regulation of gene expression mediated by homogeneous Mesp1 induction within embryoid bodies.

a, RNA-FISH on sections of EBs from Mesp1 Dox-inducible PSC lines in control conditions (NO DOX) or 24h after doxycycline induction (+DOX) showing *Mesp1* expression in red. (representative image of 6 independent embryonic bodies). **b,** Percentage of cells that are positive for *Mesp1* using RNA in situ hybridization, in control (NO DOX) or upon Mesp1 overexpression (+DOX) (n=6 for NO DOX and n=7 for +DOX independent

embryoid bodies). Error bars indicate SEM. Statistical analyses were performed by two-tail unpaired student t tests. p= $2.53^{\times}10^{-6}$ **c,** Level of *Mesp1* expression in control (NO DOX) or doxycycline condition (+DOX) as measured by the signal intensity from the smRNA-FISH. Error bars indicate SEM. Statistical analyses were performed by two-tail unpaired student t tests. n= 30 representative Mesp1+ cells per condition. p= $6.87^{\times}10^{-10}$ **d**, Principal component analysis of RNA-seq samples performed at day 2.5 (0h), day 3 (12h following Mesp1 overexpression) and day 3.5 (24h following Mesp1 overexpression) in control (no dox) and Mesp1 overexpression (dox) conditions during PSC differentiation. Note the excellent concordance between biological duplicates. **e,** Representative examples of genes that undergo early, constant or late downregulation mediated by Mesp1. Examples were chosen to represent the diversity of kinetics we could find in genes repressed by Mesp1. **f,** Plots representing all genes classified as early, constant or late downregulated, as well as their respective average profile (thick lines).
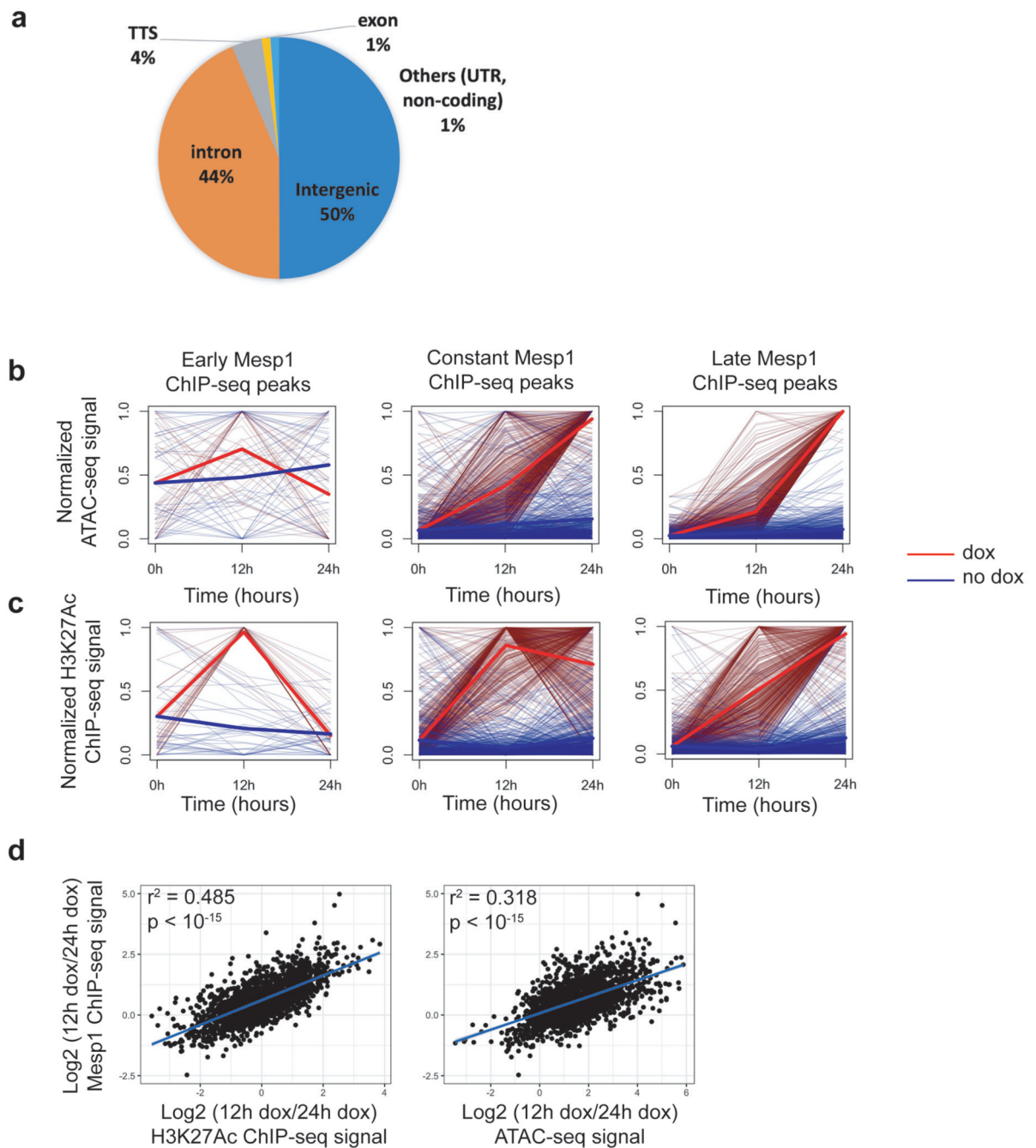
**Extended Data Fig. 2. Temporal analysis of Mesp1 transcription factor activity.**
a, Pie chart representing the distribution of the position of Mesp1 ChIP-seq peaks relative
to protein-coding genes (data shown represent two biologically independent replicates. **b-c**,
Graph representing the dynamics of ATAC-seq (**b**) and $H_3K_{27}Ac$ ChIP-seq (**c**) signal within
early, constant and late Mesp1 ChIP-seq peaks. Average profiles are shown by the think red
(dox) and blue (no dox) lines. **d**, Dot plot illustrating the correlation between temporality
of Mesp1 binding and surrounding $H_3K_{27}Ac$ deposition (left), or Mesp1 binding and ATAC-
seq opening of the chromatin (right). For each individual Mesp1 ChIP-seq peak, the average

signal was measured for all three types of experiments at 12 hours and 24 hours dox, in order to extract a measure of fold-change between 12 and 24 hours. The log2 value of this fold-change for each peak was compared between each type of experiment. Data shown represent two biologically independent replicates; p-values were calculated through a two-tailored t-test.
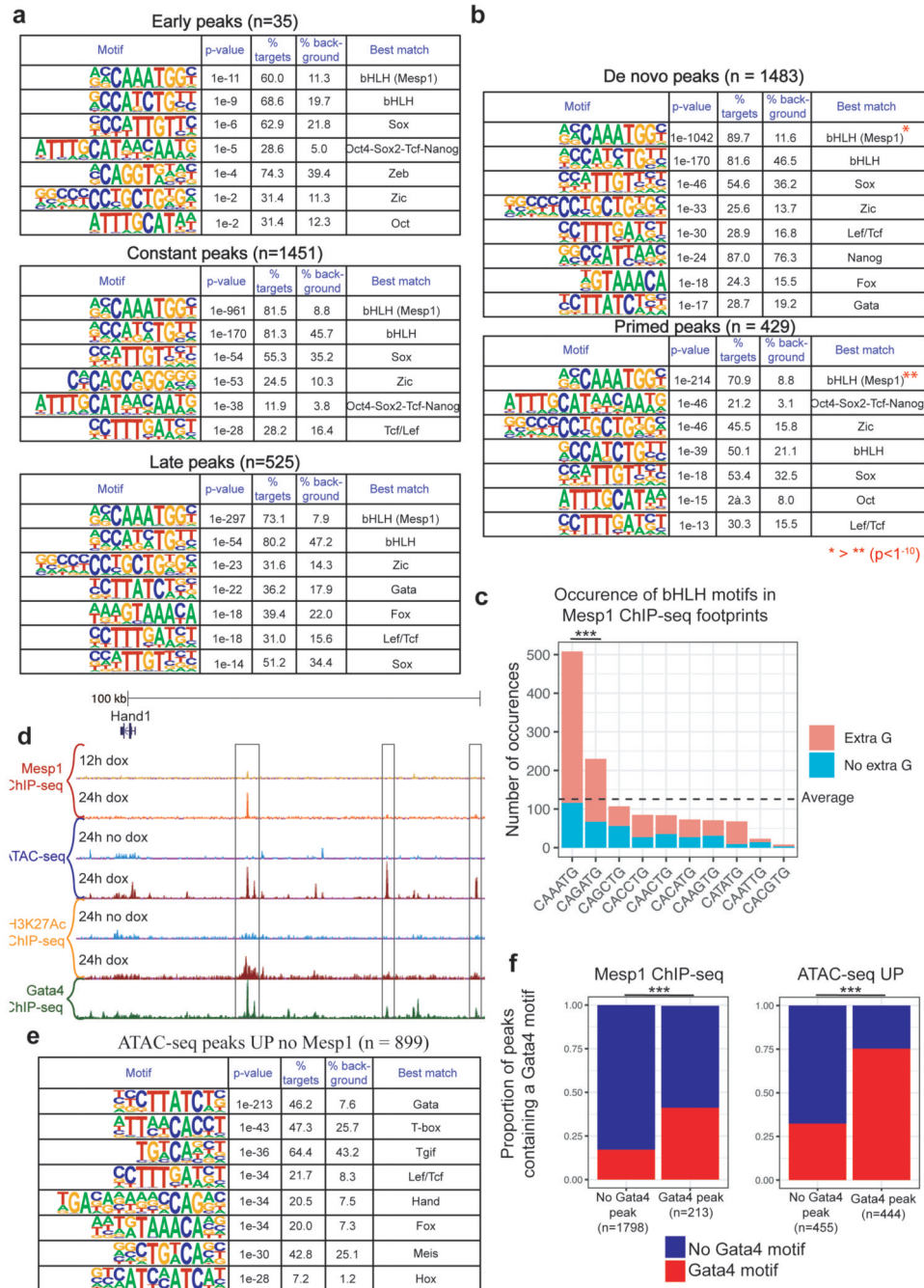

Extended Data Fig. 3. Motif discovery of enhancers activated by Mesp1.

a, Motif discovery performed separately in Mesp1 ChIP-seq peaks classified as early, constant or late peaks. p-values are calculated through a binomial test. **b**, Same pipeline as panel a but separating peaks as *de novo* versus primed Mesp1 peaks. Stars represent the enrichment of the CAAATGG motif in pioneer peaks in comparison to non-pioneer peaks through a two-tailored Z-test. Data shown represent two biologically independent replicates. **c**, Quantification of the occurrence of all forms of bHLH motifs in Mesp1 ChIP-seq peaks, showing the prevalence of CAAATG motif, most often with an extra G. *** The top 2 bHLH motifs were significantly overrepresented (p<0.00001) through a two-tailored Z-test, with z = 17.5 for CAAATG and z = 5.9 for CAGATG. **d**, Representative genomic locus where ATAC-seq and $H_3K_{27}Ac$ ChIP-seq peaks are upregulated 24h following Mesp1 expression but which are not directly bound by Mesp1 (red boxes), suggesting that their regulation is mediated by other TFs, whose expression is induced directly or indirectly by Mesp1. Gata4 ChIP-seq data[31] (green) shows a strong overlap between these *de novo* opened peaks not bound but induced by Mesp1. **e**, Motif discovery searching for known TF binding sites within ATAC-seq peaks that get opened by Mesp1 but are not directly bound by Mesp1. p-values are calculated through a binomial test. **f**, Quantification of peaks containing a Gata4 motif, both in Mesp1 ChIP-seq peaks and ATAC-seq peaks UP without Mesp1 binding, separated as peaks bound by Gata4 or not bound by Gata4 using a published ChIP-seq dataset[31]. *** Gata4 motifs are statistically significantly enriched in Gata4-binding peaks in Mesp1 ChIP-seq (z=8.34, p<0.00001) and in ATAC-seq peaks UP (z=13.6, p<0.00001). Data shown represent two biologically independent replicates; n=1 for previously published Gata4 ChIP-seq[31]. These values were calculated through a two-tailored Z-test.
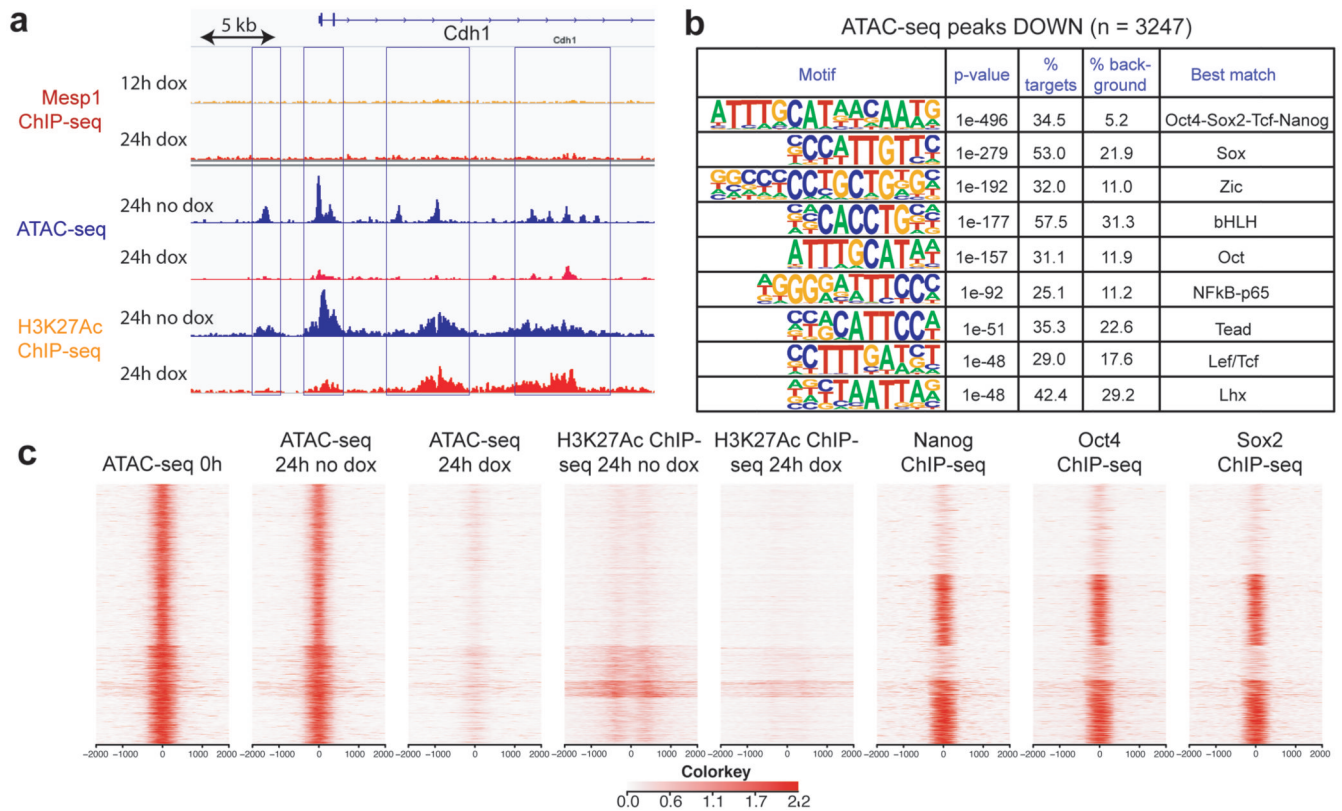
**Extended Data Fig. 4. Repression of the core pluripotency network by Mesp1.**
a, Representative genomic locus (*Cdh1*) where ATAC-seq and H3K27Ac ChIP peaks found in control (no dox) conditions are absent or smaller in dox conditions, without presenting any Mesp1 binding (blue boxes), suggesting indirect repression of chromatin opening by other factors. **b**, Motif discovery within peaks that are closed in dox conditions, including enrichment of a compound OCT-SOX-TCF-NANOG motif. p-values are calculated through a binomial test. **c**, Heatmap showing signal of ATAC-seq, $H_3K_{27}Ac$ ChIP-seq and published Nanog, Oct4 and Sox2 ChIP-seq data within peaks that were closed after dox-induced Mesp1 overexpression[32].
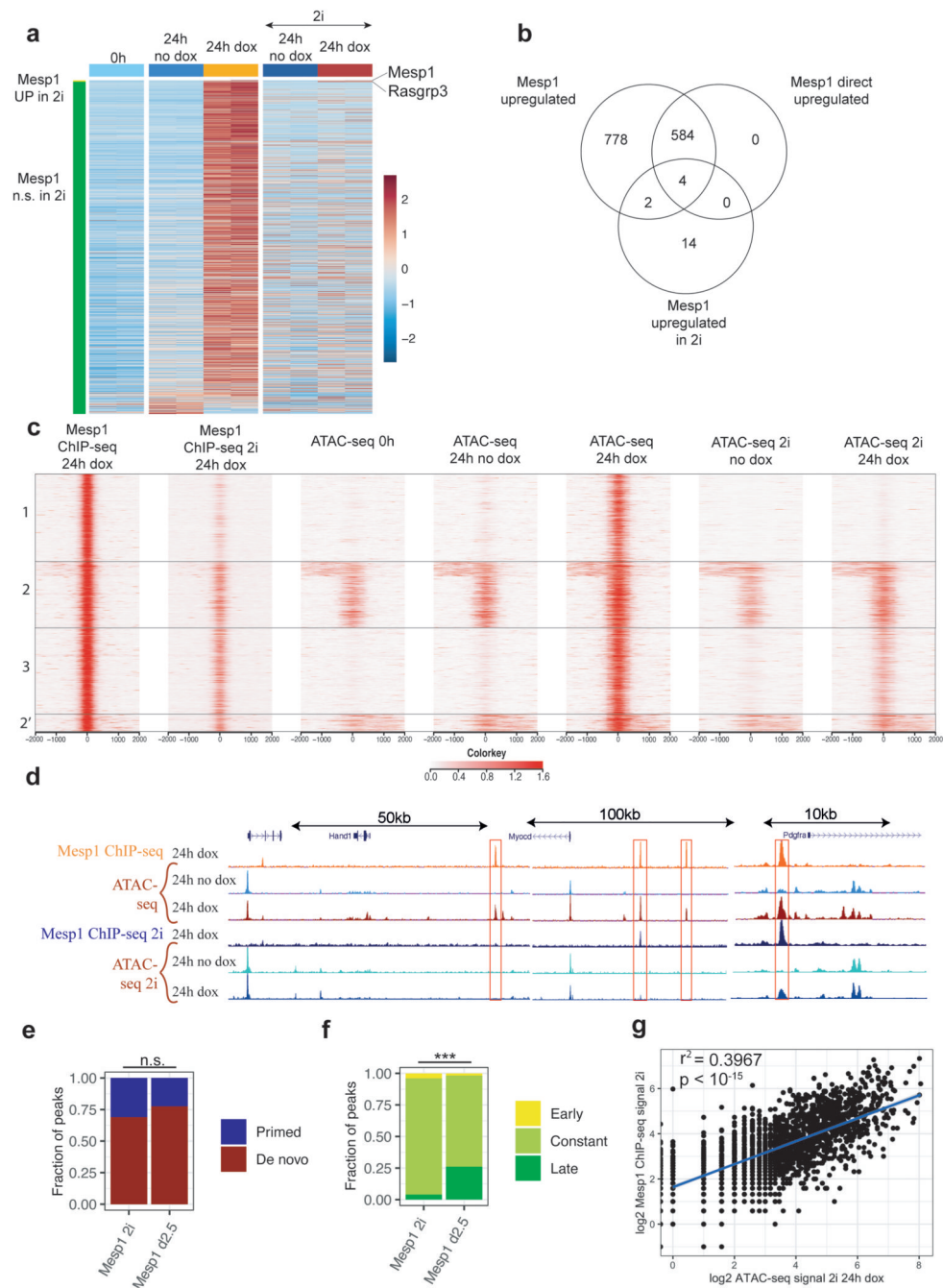
Extended Data Fig. 5. Context-dependency of Mesp1 activator potential.
a, Heatmap of the expression values of Mesp1 upregulated genes in undifferentiated PSCs (2i conditions), with or without Mesp1 induction (dox). RNA-seq samples in 2i were performed twice. **b**, Overlap between genes directly and indirectly activated by Mesp1 during PSC differentiation or in 2i, illustrating the paucity of Mesp1-mediated gene activation in pluripotency. **c**, Heatmap illustrating Mesp1 binding affinity to its enhancers in 2i by Mesp1 ChIP-seq and the subsequent lack of chromatin opening by ATAC-seq in 2i conditions. Each row represents a Mesp1 binding site detected during differentiation at

24h dox. Peaks were ordered by unsupervised k-means clustering. 1, *de novo* peaks where Mesp1 binding and subsequent chromatin opening is lost in 2i; 2 and 2', primed peaks with conserved Mesp1 binding and chromatin opening in 2i; 3, *de novo* ATAC-seq peaks where Mesp1 binding and chromatin opening is conserved in 2i. All samples collected in 2i were performed twice. **d**, Representative examples of Mesp1 binding loci in 2i conditions. **e-f**, Classification of Mesp1 ChIP-seq peaks found in 2i into previously detailed chromatin opening (e) or kinetic (f) groups. *** Late peaks were significantly depleted in 2i conditions (z=-7.86, p< 0.00001). Data shown represent two biologically independent replicates. These values were calculated through a two-tailored Z-test. **g**, Quantification of the correlation between Mesp1 binding strength measured by Mesp1 ChIP-seq and chromatin opening in ATAC-seq in 2i (24h dox), demonstrating a linear correlation between these two variables. Data shown represent two biologically independent replicates.
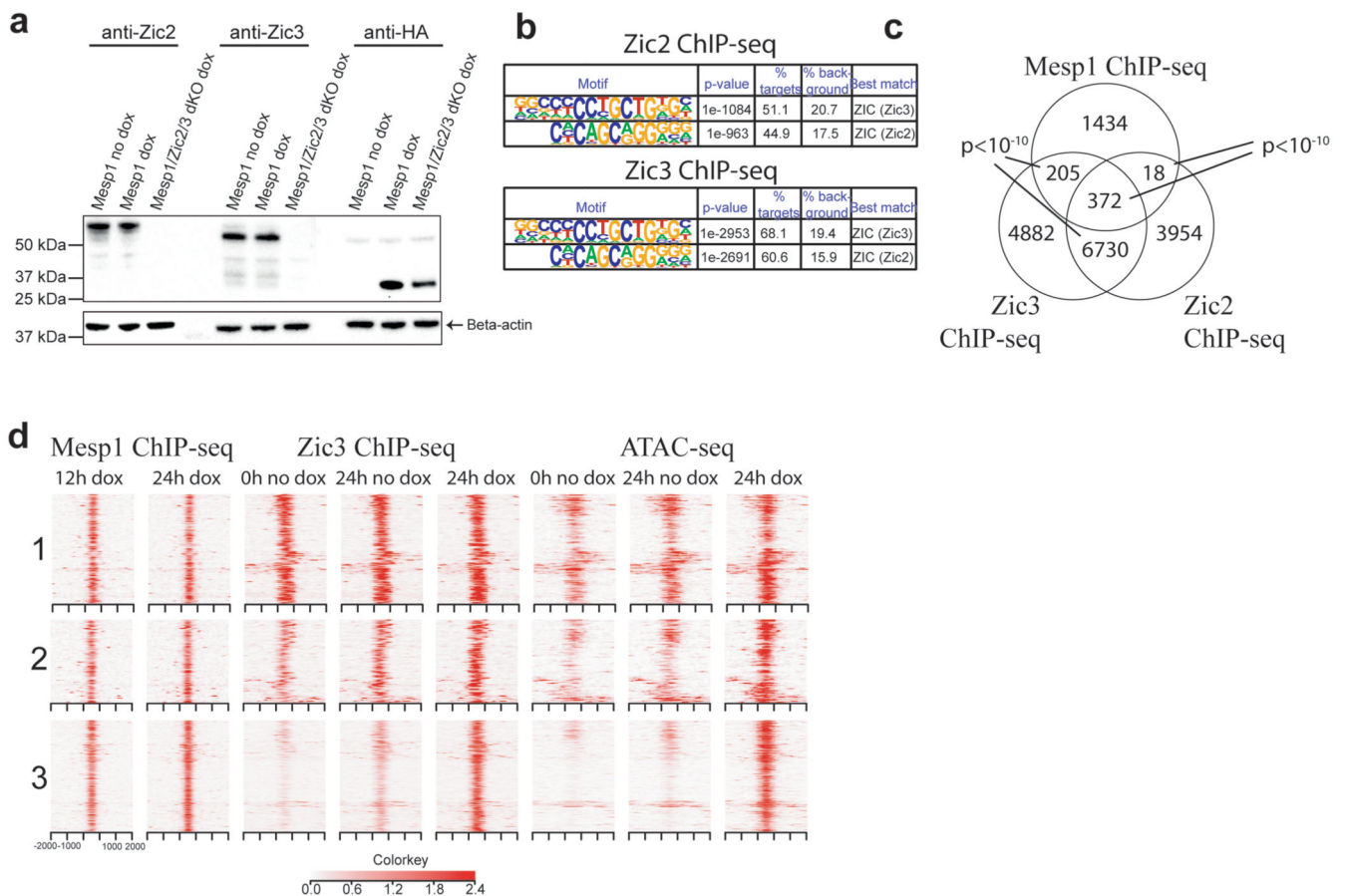
Extended Data Fig. 6. Zic2 and Zic3 cooperate with Mesp1 and potentially with other mesoderm-inducing TFs.

a, Western blot illustrating the expression of Zic2 and Zic3 with and without Mesp1 induction, at day 3.5 (24 hours) of PSC differentiation, as well as the lack of Zic2 and Zic3 protein expression in Zic2/3 double KO cell lines. (Data shown represent 2 independent experiments) **b**, Illustration of the two most enriched motifs in all Zic2 and Zic3 ChIP-seq detected peaks. p-values are calculated through a binomial test. **c**, Venn diagram illustrating

the number of overlapping peaks between Mesp1, Zic2 and Zic3 ChIP-seq datasets. P-value was calculated using a hypergeometric test, using bedtools fisher. **d**, Illustration of the temporality of Zic3 binding within primed and *de novo* Mesp1-bound peaks, with associated ATAC-seq signal. 1, peaks already bound by Zic3 at day 2.5 (0 hours); 2, peaks bound by Zic3 at 24h no dox; 3, peaks bound by Zic3 at 24h dox.
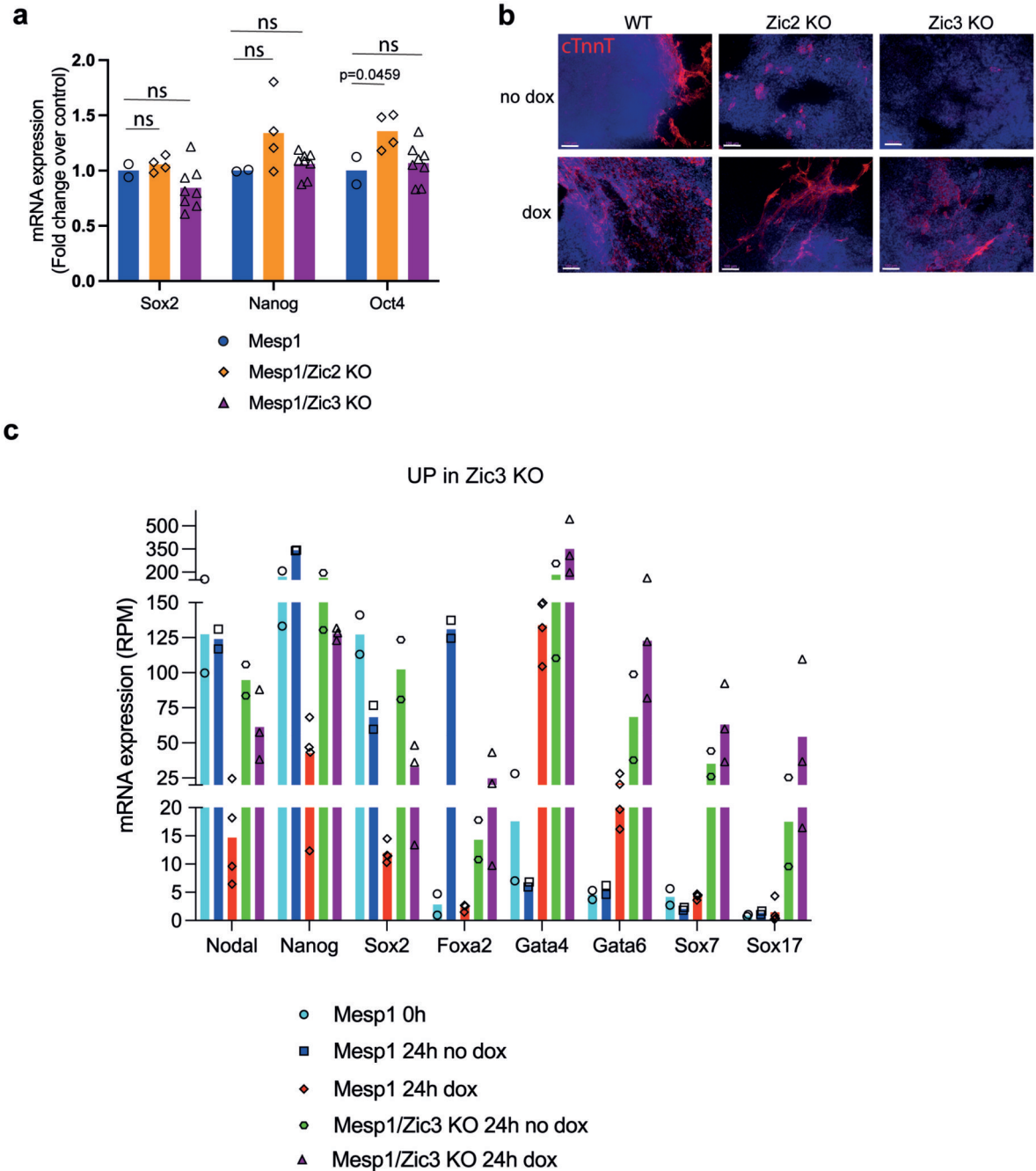
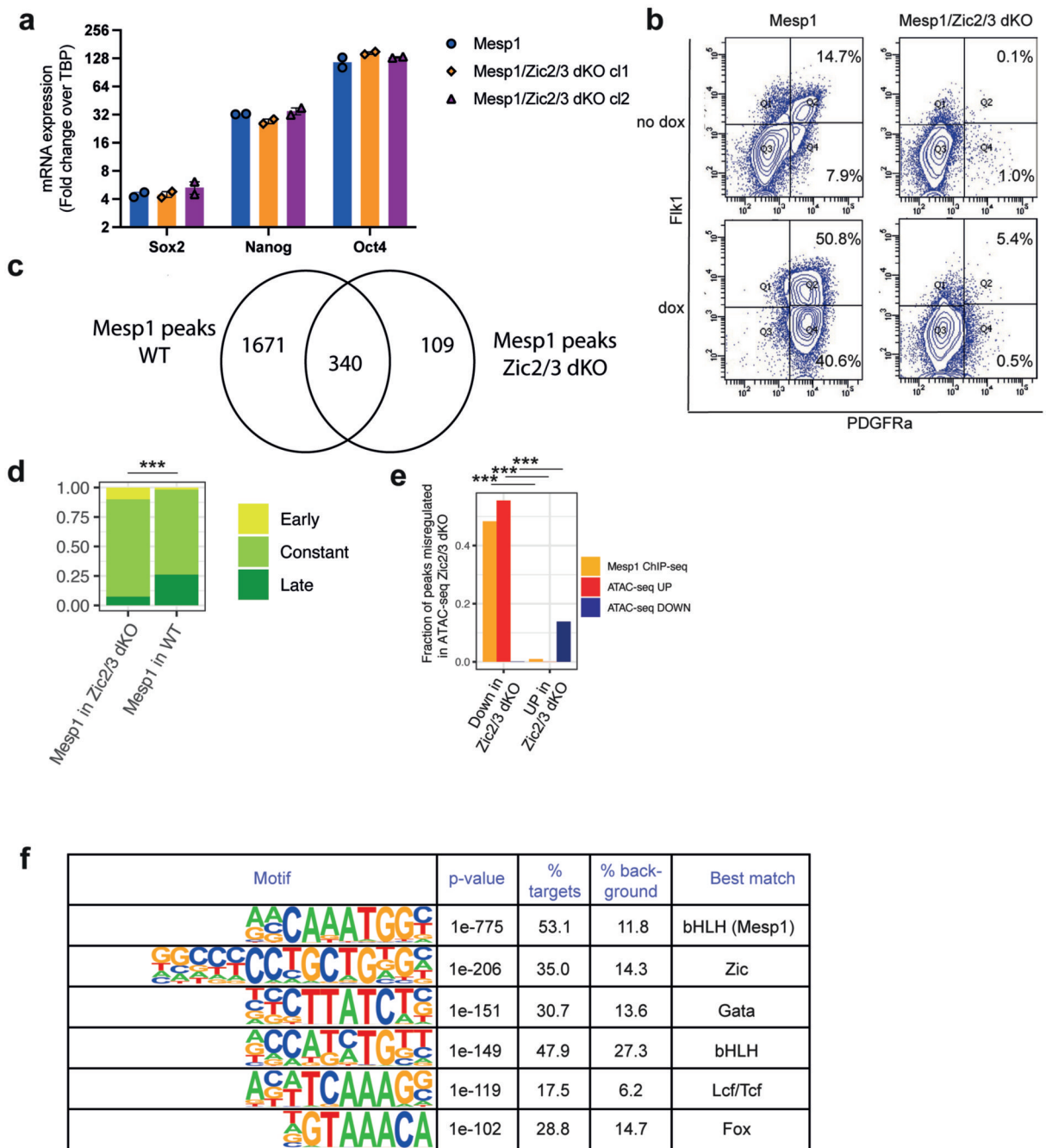**Extended Data Fig. 7. Zic2 and Zic3 regulate Mesp1-induced CP specification and differentiation.**

a, mRNA expression of the core pluripotency associated TFs in Mesp1-inducible WT, *Zic2* and *Zic3* KO PSCs in Lif/2i pluripotency conditions, as measured by RT-qPCR. (n=4 biologically independent replicates covering two independent KO clones with each assessed by two independent experiments. Error bars indicate mean +/-SEM, statical analysis was performed by 2-way ANOVA. **b**, Representative immunofluorescence for Troponin T in Mesp1-inducible WT, Zic2 and Zic3 KO cell lines at day 10 of differentiation, illustrating the ability of Mesp1 overexpression to overcome cardiac differentiation defects in Zic2 and Zic3 KO cell lines. (Data shown represent 6 independent experiments. Scale bars=100 μm. **c**, Illustrative examples of genes that are significantly upregulated in Zic3 KO cells in comparison to WT cells, including known important factors of pluripotency and endoderm differentiation.
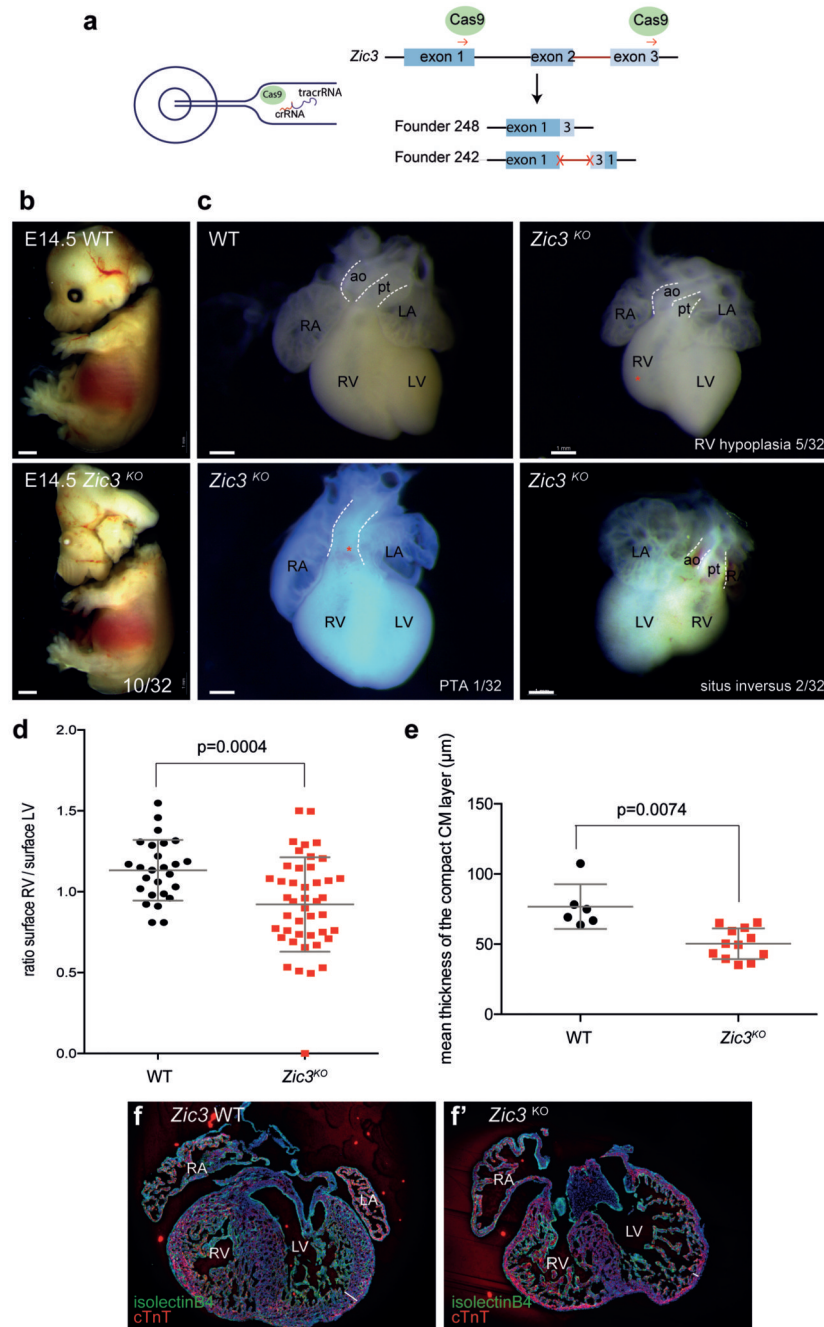
Extended Data Fig. 8. Heart defects observed in the newly generated *Zic3* KO line.
a, Zygote injection strategy used to generate *Zic3* KO mice. **b**, Pictures of Wild type
(WT) and homozygous null (*Zic3*^{KO}) E14.5 embryos, showing severe neural tube closure
defects and exencephaly found in a subset of *Zic3 KO* embryos (n=10/32). Scale bars=
1mm. **c**, Range of cardiac morphological abnormalities found in *Zic3* KO embryos at
E14.5. We observe outflow tract defects with persistent truncus arteriosus (PTA bottom
left panel – n=1/32), hypoplasia of the right ventricle (RV) (upper right panel – n=5/32)
and mutants with a *situs inversus* phenotype (bottom right panel – n=2/32). Scale bars=

1mm. **d**, ratio of the surface area of the right ventricle compared to the surface area of the left ventricle in wild type (black – n= 26) and *Zic3* KO (red – n=44). Error bars indicate mean +/-SEM. Unpaired, two-tailed t-test showed a p-value=0.0004. **e**, Mean thickness of the compact myocardial (CM) layer of the ventricles in wild type (black – n=6) and *Zic3 KO* embryos (red – n=12). Error bars indicate mean +/-SEM. Unpaired, twotailed t-test showed a p-value=0.0074. **f-f'**, Immunofluorescence on E14.5 wild type (**f**) and *Zic3* KO (**f'**) hearts using an anti-cardiac Troponin T (cTnT) antibody to label the cardiomyocytes and isolectinB4 to label the endocardium (representative pictures from 4 independent hearts of each genotype). No endocardial defect was observed in *Zic3 KO* embryos while the cTnT+ layer was thinner in f'. Scale bars= 200μm. RV, right ventricle; LV, left ventricle; RA, right atrium; LA, left atrium; pt, pulmonary trunk; ao, aorta.

Motif	p-value	% targets	% back-ground	Best match
AACAAATGGC	1e-775	53.1	11.8	bHLH (Mesp1)
GGCCCCCTGCTGGG	1e-206	35.0	14.3	Zic
TCCTTATCTG	1e-151	30.7	13.6	Gata
ACCATCTGTT	1e-149	47.9	27.3	bHLH
ACATCAAAGG	1e-119	17.5	6.2	Lcf/Tcf
TGTAAACA	1e-102	28.8	14.7	Fox

Extended Data Fig. 9. Zic2 and Zic3 redundantly regulate Mesp1 activity.
a, Expression of three core pluripotency genes in Mesp1 WT and two independent Zic2/3 KO PSC cell lines cultured in Lif/2i medium. Data from two independent experiments, **b**, FACS profiles of EBs at day 4 of differentiation from Mesp1 WT and *Zic2*/3dKO cell lines, illustrating the decrease in Flk1 and PDGFRa expression in *Zic2/3*dKO cell lines both in no dox and dox conditions. **c**, Table shows the distribution of genes that were downregulated in *Zic3* KO and *Zic2/3*dKO cells within the temporal categories of Mesp1 direct upregulated target genes. There was no particular enrichment for early, constant and

late genes within the Zic2/3-dependent fraction of Mesp1 target. **d**, Barplot illustrating the proportion of early, constant and late Mesp1 binding sites within Mesp1 ChIP-seq peaks conserved in *Zic2/3*dKO cell lines. *** for late genes, z=-5.845, p<0.00001. These values were calculated through a two-tailored Z-test. Data shown represent two biologically independent replicates. **e**, Representation of the proportion of Mesp1 ChIP-seq peaks as well as ATAC-seq peaks that are opened (UP) or closed (DOWN- upon Mesp1 induction in WT cells which are preferentially closed in Zic2/3 dKO cells after Mesp1 induction. n = 2 independent experiments for Mesp1 ChIP-seq and ATAC-seq in WT cells; n = 3 independent experiments for ATAC-seq in Zic2/3 dKO cell lines. *** all three comparisons were significant with p<0.00001 and respectively z = 34.9 (Mesp1 ChIP-seq), z = 54.2 (ATAC-seq UP) and z = -20.1 (ATAC-seq DOWN). These values were calculated through a two-tailored Z-test. **f**, Motif enrichment analysis of ATAC-seq peaks that were preferentially closed in *Zic2/3*dKO cells in comparison to WT cells. p-values are calculated through a binomial test.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Data Availability

NGS data (ChIP-seq, ATAC-seq and RNA-seq) generated during this study has been deposited in Gene Expression Omnibus (GEO) and is accessible through GEO Series accession number GSE165107. Previously published NGS data or microarray data that were re-analysed here are available under accession code GSE41361, GSE44288 and GSE59033. Source data for Figs. 3,5,6 and 7 and Extended Data Figs. 1,6,7,8 and 9 are provided with this paper. All other data supporting the findings of this study are available from the corresponding author on reasonable request.

References

1. Xin M, Olson EN, Bassel-Duby R. Mending broken hearts: cardiac development as a basis for adult heart regeneration and repair. Nat Rev Mol Cell Biol. 2013; 14: 529–541. [PubMed: 23839576]

2. Harvey RP. Patterning the vertebrate heart. Nat Rev Genet. 2002; 3: 544–556. [PubMed: 12094232]

3. Meilhac SM, Buckingham ME. The deployment of cell lineages that form the mammalian heart. Nat Rev Cardiol. 2018; 15: 705–724. [PubMed: 30266935]

4. Protze SI, Lee JH, Keller GM. Human Pluripotent Stem Cell-Derived Cardiovascular Cells: From Developmental Biology to Therapeutic Applications. Cell Stem Cell. 2019; 25: 311–327. [PubMed: 31491395]

5. Hoffman JI, Kaplan S. The incidence of congenital heart disease. J Am Coll Cardiol. 2002; 39: 1890–1900. [PubMed: 12084585]

6. Srivastava D. Making or breaking the heart: from lineage determination to morphogenesis. Cell. 2006; 126: 1037–1048. [PubMed: 16990131]

7. Bruneau BG. The developmental genetics of congenital heart disease. Nature. 2008; 451: 943–948. [PubMed: 18288184]

8. Saga Y, et al. MesP1: a novel basic helix-loop-helix protein expressed in the nascent mesodermal cells during mouse gastrulation. Development. 1996; 122: 2769–2778. [PubMed: 8787751]

9. Saga Y, et al. MesP1 is expressed in the heart precursor cells and required for the formation of a single heart tube. Development. 1999; 126: 3437–3447. [PubMed: 10393122]

10. Kitajima S, Takagi A, Inoue T, Saga Y. MesP1 and MesP2 are essential for the development of cardiac mesoderm. Development. 2000; 127: 3215–3226. [PubMed: 10887078]

11. Lescroart F, et al. Early lineage restriction in temporally distinct populations of Mesp1 progenitors during mammalian heart development. Nat Cell Biol. 2014; 16: 829–840. [PubMed: 25150979]

12. Lescroart F, et al. Defining the earliest step of cardiovascular lineage segregation by single-cell RNA-seq. Science. 2018; 359: 1177–1181. [PubMed: 29371425]

13. Bondue A, et al. Mesp1 acts as a master regulator of multipotent cardiovascular progenitor specification. Cell Stem Cell. 2008; 3: 69–84. [PubMed: 18593560]

14. David R, et al. MesP1 drives vertebrate cardiovascular differentiation through Dkk-1-mediated blockade of Wnt-signalling. Nat Cell Biol. 2008; 10: 338–345. [PubMed: 18297060]

15. Lindsley RC, et al. Mesp1 coordinately regulates cardiovascular fate restriction and epithelial-mesenchymal transition in differentiating ESCs. Cell Stem Cell. 2008; 3: 55–68. [PubMed: 18593559]

16. Chan SS, et al. Mesp1 patterns mesoderm into cardiac, hematopoietic, or skeletal myogenic progenitors in a context-dependent manner. Cell Stem Cell. 2013; 12: 587–601. [PubMed: 23642367]

17. Devine WP, Wythe JD, George M, Koshiba-Takeuchi K, Bruneau BG. Early patterning and specification of cardiac progenitors in gastrulating mesoderm. Elife. 2014; 3

18. Zhang Q, et al. Unveiling Complexity and Multipotentiality of Early Heart Fields. Circ Res. 2021; 129: 474–487. [PubMed: 34162224]

19. Ivanovitch K, et al. Ventricular, atrial and outflow tract heart progenitors arise from spatially and molecularly distinct regions of the primitive streak. Plos One. 2021; 19 (5) e3001200

20. Creyghton MP, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A. 2010; 107: 21931–21936. [PubMed: 21106759]

21. Shlyueva D, Stampfel G, Stark A. Transcriptional enhancers: from properties to genome-wide predictions. Nat Rev Genet. 2014; 15: 272–286. [PubMed: 24614317]

22. Heinz S, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010; 38: 576–589. [PubMed: 20513432]

23. Wapinski OL, et al. Hierarchical mechanisms for direct reprogramming of fibroblasts to neurons. Cell. 2013; 155: 621–635. [PubMed: 24243019]

24. Soufi A, et al. Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. Cell. 2015; 161: 555–568. [PubMed: 25892221]

25. Paige SL, et al. A temporal chromatin signature in human embryonic stem cells identifies regulators of cardiac development. Cell. 2012; 151: 221–232. [PubMed: 22981225]

26. Wamstad JA, et al. Dynamic and coordinated epigenetic regulation of developmental transitions in the cardiac lineage. Cell. 2012; 151: 206–220. [PubMed: 22981692]

27. Pikkarainen S, Tokola H, Kerkela R, Ruskoaho H. GATA transcription factors in the developing and adult heart. Cardiovasc Res. 2004; 63: 196–207. [PubMed: 15249177]

28. George RM, Firulli AB. Hand Factors in Cardiac Development. Anat Rec (Hoboken). 2019; 302: 101–107. [PubMed: 30288953]

29. Greulich F, Rudat C, Kispert A. Mechanisms of T-box gene function in the developing heart. Cardiovasc Res. 2011; 91: 212–222. [PubMed: 21498422]

30. Kume T. Novel insights into the differential functions of Notch ligands in vascular formation. J Angiogenes Res. 2009; 1: 8. [PubMed: 20016694]

31. Oda M, et al. DNA methylation restricts lineage-specific functions of transcription factor Gata4 during embryonic stem cell differentiation. PLoS Genet. 2013; 9 e1003574 [PubMed: 23825962]

32. Whyte WA, et al. Master transcription factors and mediator establish superenhancers at key cell identity genes. Cell. 2013; 153: 307–319. [PubMed: 23582322]

33. Bondue A, et al. Defining the earliest step of cardiovascular progenitor specification during embryonic stem cell differentiation. J Cell Biol. 2011; 192: 751–765. [PubMed: 21383076]

34. Kattman SJ, et al. Stage-specific optimization of activin/nodal and BMP signaling promotes cardiac differentiation of mouse and human pluripotent stem cell lines. Cell Stem Cell. 2011; 8: 228–240. [PubMed: 21295278]

35. Ying QL, et al. The ground state of embryonic stem cell self-renewal. Nature. 2008; 453: 519–523. [PubMed: 18497825]

36. Gebbia M, et al. X-linked situs abnormalities result from mutations in ZIC3. Nat Genet. 1997; 17: 305–308. [PubMed: 9354794]

37. Ware SM, et al. Identification and functional analysis of ZIC3 mutations in heterotaxy and related congenital heart defects. Am J Hum Genet. 2004; 74: 93–105. [PubMed: 14681828]

38. Cowan J, Tariq M, Ware SM. Genetic and functional analyses of ZIC3 variants in congenital heart disease. Hum Mutat. 2014; 35: 66–75. [PubMed: 24123890]

39. Purandare SM, et al. A complex syndrome of left-right axis, central nervous system and axial skeleton defects in Zic3 mutant mice. Development. 2002; 129: 2293–2302. [PubMed: 11959836]

40. Ware SM, Harutyunyan KG, Belmont JW. Zic3 is critical for early embryonic patterning during gastrulation. Dev Dyn. 2006; 235: 776–785. [PubMed: 16397896]

41. Zhu L, et al. Identification of a novel role of ZIC3 in regulating cardiac development. Hum Mol Genet. 2007; 16: 1649–1660. [PubMed: 17468179]

42. Elms P, et al. Overlapping and distinct expression domains of Zic2 and Zic3 during mouse gastrulation. Gene Expr Patterns. 2004; 4: 505–511. [PubMed: 15261827]

43. Xu J, et al. Genome-wide CRISPR screen identifies ZIC2 as an essential gene that controls the cell fate of early mesodermal precursors to human heart progenitors. Stem Cells. 2020; 38: 741–755. [PubMed: 32129551]

44. Lim LS, Hong FH, Kunarso G, Stanton LW. The pluripotency regulator Zic3 is a direct activator of the Nanog promoter in ESCs. Stem Cells. 2010; 28: 1961–1969. [PubMed: 20872845]

45. Lim LS, et al. Zic3 is required for maintenance of pluripotency in embryonic stem cells. Mol Biol Cell. 2007; 18: 1348–1358. [PubMed: 17267691]

46. Perl E, Waxman JS. Retinoic Acid Signaling and Heart Development. Subcell Biochem. 2020; 95: 119–149. [PubMed: 32297298]

47. Bertrand N, et al. Hox genes define distinct progenitor sub-domains within the second heart field. Dev Biol. 2011; 353: 266–274. [PubMed: 21385575]

48. Ryckebusch L, et al. Retinoic acid deficiency alters second heart field formation. Proc Natl Acad Sci U S A. 2008; 105: 2913–2918. [PubMed: 18287057]

49. Deshwar AR, Chng SC, Ho L, Reversade B, Scott IC. The Apelin receptor enhances Nodal/TGFbeta signaling to ensure proper cardiac development. Elife. 2016; 5

50. Scott IC, et al. The g protein-coupled receptor agtrl1b regulates early development of myocardial progenitors. Dev Cell. 2007; 12: 403–413. [PubMed: 17336906]

51. Zeng XX, Wilm TP, Sepich DS, Solnica-Krezel L. Apelin and its receptor control heart field formation during zebrafish gastrulation. Dev Cell. 2007; 12: 391–402. [PubMed: 17336905]

52. Klootwijk R, et al. A deletion encompassing Zic3 in bent tail, a mouse model for X-linked neural tube defects. Hum Mol Genet. 2000; 9: 1615–1622. [PubMed: 10861288]

53. Pijuan-Sala B, et al. A single-cell molecular map of mouse gastrulation and early organogenesis. Nature. 2019; 566: 490–495. [PubMed: 30787436]

54. Guibentif C, et al. Diverse Routes toward Early Somites in the Mouse Embryo. Dev Cell. 2021; 56: 141–153. e146 [PubMed: 33308481]

55. Zaret KS. Pioneer Transcription Factors Initiating Gene Network Changes. Annu Rev Genet. 2020; 54: 367–385. [PubMed: 32886547]

56. Zaret KS. Pioneering the chromatin landscape. Nat Genet. 2018; 50: 167–169. [PubMed: 29374252]

57. Zaret KS, Carroll JS. Pioneer transcription factors: establishing competence for gene expression. Genes Dev. 2011; 25: 2227–2241. [PubMed: 22056668]

58. Iwafuchi-Doi M, Zaret KS. Cell fate control by pioneer transcription factors. Development. 2016; 143: 1833–1837. [PubMed: 27246709]

59. Liu Z, Kraus WL. Catalytic-Independent Functions of PARP-1 Determine Sox2 Pioneer Activity at Intractable Genomic Loci. Mol Cell. 2017; 65: 589–603. e589 [PubMed: 28212747]

60. Swinstead EE, et al. Steroid Receptors Reprogram FoxA1 Occupancy through Dynamic Chromatin Transitions. Cell. 2016; 165: 593–605. [PubMed: 27062924]

61. Donaghey J, et al. Genetic determinants and epigenetic effects of pioneer-factor occupancy. Nat Genet. 2018; 50: 250–258. [PubMed: 29358654]

62. Mayran A, et al. Pioneer and nonpioneer factor cooperation drives lineage specific chromatin opening. Nat Commun. 2019; 10 3807 [PubMed: 31444346]

63. Cernilogar FM, et al. Pre-marked chromatin and transcription factor co-binding shape the pioneering activity of Foxa2. Nucleic Acids Res. 2019; 47: 9069–9086. [PubMed: 31350899]

64. Sutherland MJ, Wang S, Quinn ME, Haaning A, Ware SM. Zic3 is required in the migrating primitive streak for node morphogenesis and left-right patterning. Hum Mol Genet. 2013; 22: 1913–1923. [PubMed: 23303524]

65. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014; 15: 550. [PubMed: 25516281]

66. Tarbell ED, Liu T. HMMRATAC: a Hidden Markov ModeleR for ATAC-seq. Nucleic Acids Res. 2019; 47 e91 [PubMed: 31199868]

67. Chiapparo G, et al. Mesp1 controls the speed, polarity, and directionality of cardiovascular progenitor migration. J Cell Biol. 2016; 213: 463–477. [PubMed: 27185833]

68. Ran FA, et al. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell. 2013; 154: 1380–1389. [PubMed: 23992846]

69. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Curr Protoc Mol Biol. 2015; 109: 21 29 21–21 29 29.

70. Sambrook J, Russell DW. Identification of associated proteins by coimmunoprecipitation. CSH Protoc. 2006; 2006

71. Roux M, Laforest B, Capecchi M, Bertrand N, Zaffran S. Hoxb1 regulates proliferation and differentiation of second heart field progenitors in pharyngeal mesoderm and genetically interacts with Hoxa1 during cardiac outflow tract development. Dev Biol. 2015; 406: 247–258. [PubMed: 26284287]

72. de Soysa TY, et al. Single-cell analysis of cardiogenesis reveals basis for organ-level developmental defects. Nature. 2019; 572: 120–124. [PubMed: 31341279]

73. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014; 30: 2114–2120. [PubMed: 24695404]

74. Dobin A, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013; 29: 15–21. [PubMed: 23104886]

75. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics (Oxford, England). 2015; 31: 166–169.

76. Wickham, H. Ggplot2: Elegant Graphics for Data Analysis. 2nd. Springer Publishing Company, Incorporated; 2009.

77. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012; 9: 357–359. [PubMed: 22388286]

78. Li H, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; 25: 2078–2079. [PubMed: 19505943]

79. Zhang Y, et al. Model-based Analysis of ChIP-Seq (MACS). Genome Biology. 2008; 9 R137 [PubMed: 18798982]

80. McLean CY, et al. GREAT improves functional interpretation of cis-regulatory regions. Nature Biotechnology. 2010; 28: 495–501.

81. Shen L, Shao N, Liu X, Nestler E. ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. BMC Genomics. 2014; 15: 284. [PubMed: 24735413]

82. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010; 26: 841–842. [PubMed: 20110278]

83. Li Z, et al. Identification of transcription factor binding sites using ATAC-seq. Genome Biology. 2019; 20: 45. [PubMed: 30808370]
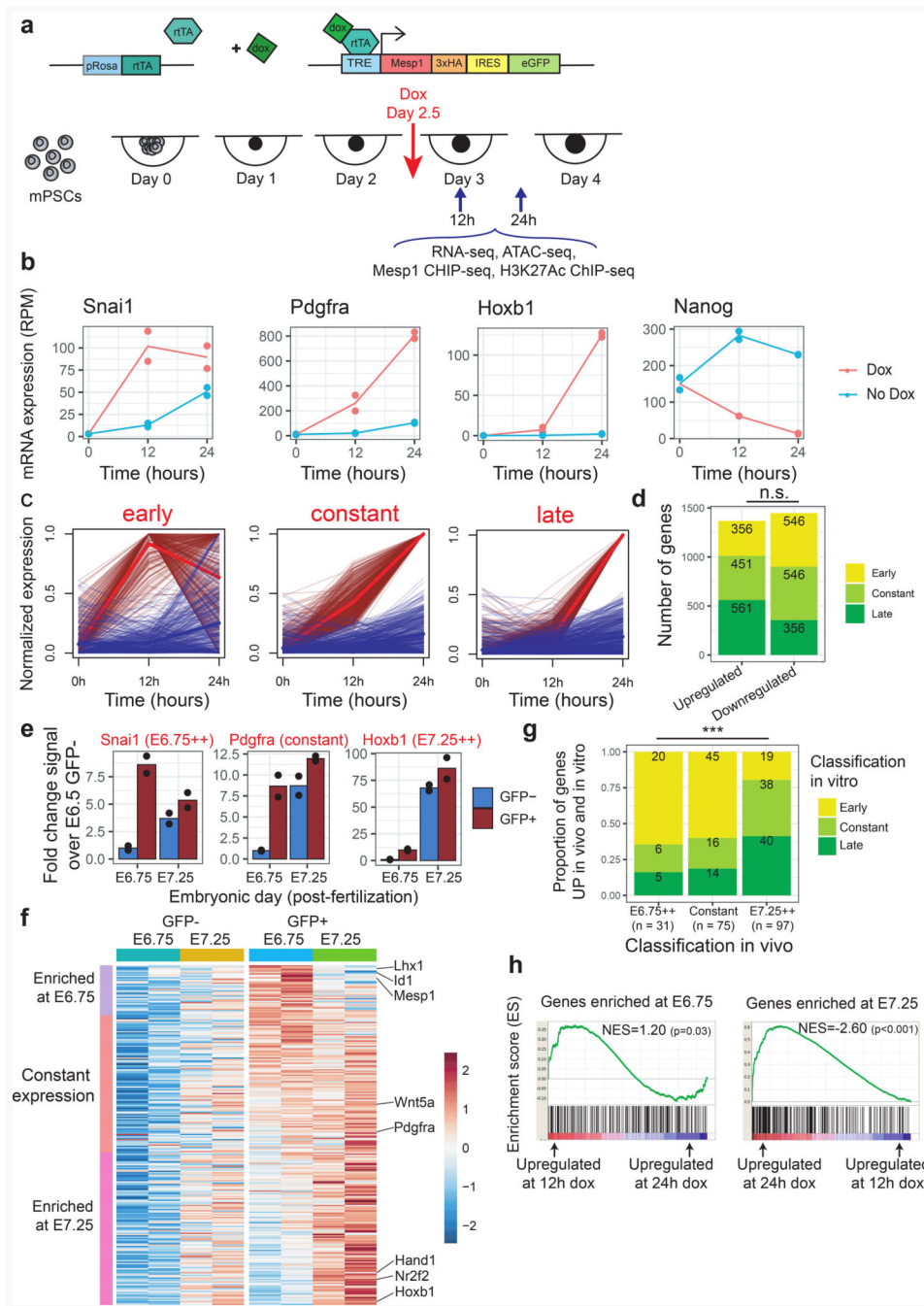
Figure 1. Dynamics of gene expression regulated by Mesp1

a, Experimental scheme allowing Mesp1 overexpression and the timing of RNA-seq, ChIP-seq and ATAC-seq analyses. **b**, Expression pattern of genes that are regulated by Mesp1, representing 4 categories of target genes: early, constant, late upregulated and downregulated. n=2 biologically independent experiments.]. **c**, Thin lines represent individual gene expression kinetics in dox (red) and no dox (blue) conditions, and thick lines represent the average profile of all genes. **d**, Barplotshowing the number of deregulated genes called using DESeq2 with a cut-off of 1.5-fold change expression and an adjusted

p-value < 0.05[65]. n = 2biologically independent experiments.n.s., non significant difference between the number of up- and downregulated genes shown in the graph. **e**, Representative expression from 2 independent experiments of upregulated genes with different patterns of expression in Mesp1-expressing (GFP[+]) cells *in vivo*, which are either enriched at E6.75 (E6.75++) or E7.25 (E7.25++), or equally expressed in both populations (constant), as measured by microarray[11]. **f**, Heatmap representing the *in vivo* patterns of expression of genes both upregulated by Mesp1 induction *in vitro* and in Mesp1-expressing cells *in vivo*, sorted according to the ratio of expression between E6.75 and E7.25 GFP+ cells. **g**, Barplotshowing the number of gene upregulated *in vivo (at* E6.75, E7.25 or both) and upregulated *in vitro* at the early, constant and late time points. n = 2independent experiments. *** Significance of the difference in the number of genes shown was assessed by a chi-square test; the chi-square statistic is 36.7904. The p-value is < 0.00001. **h**, GSEA showing the distribution of genes upregulated in GFP[+] cells *in vivo* that are enriched at E6.75 or enriched at E7.25, within the ranking of all genes based on their expression at 12h dox versus 24h dox. n = 2 biologically independent experiments. Nominal p-value represents the statistical significance of the enrichment score, relative to a null distribution calculated through a permutation test.
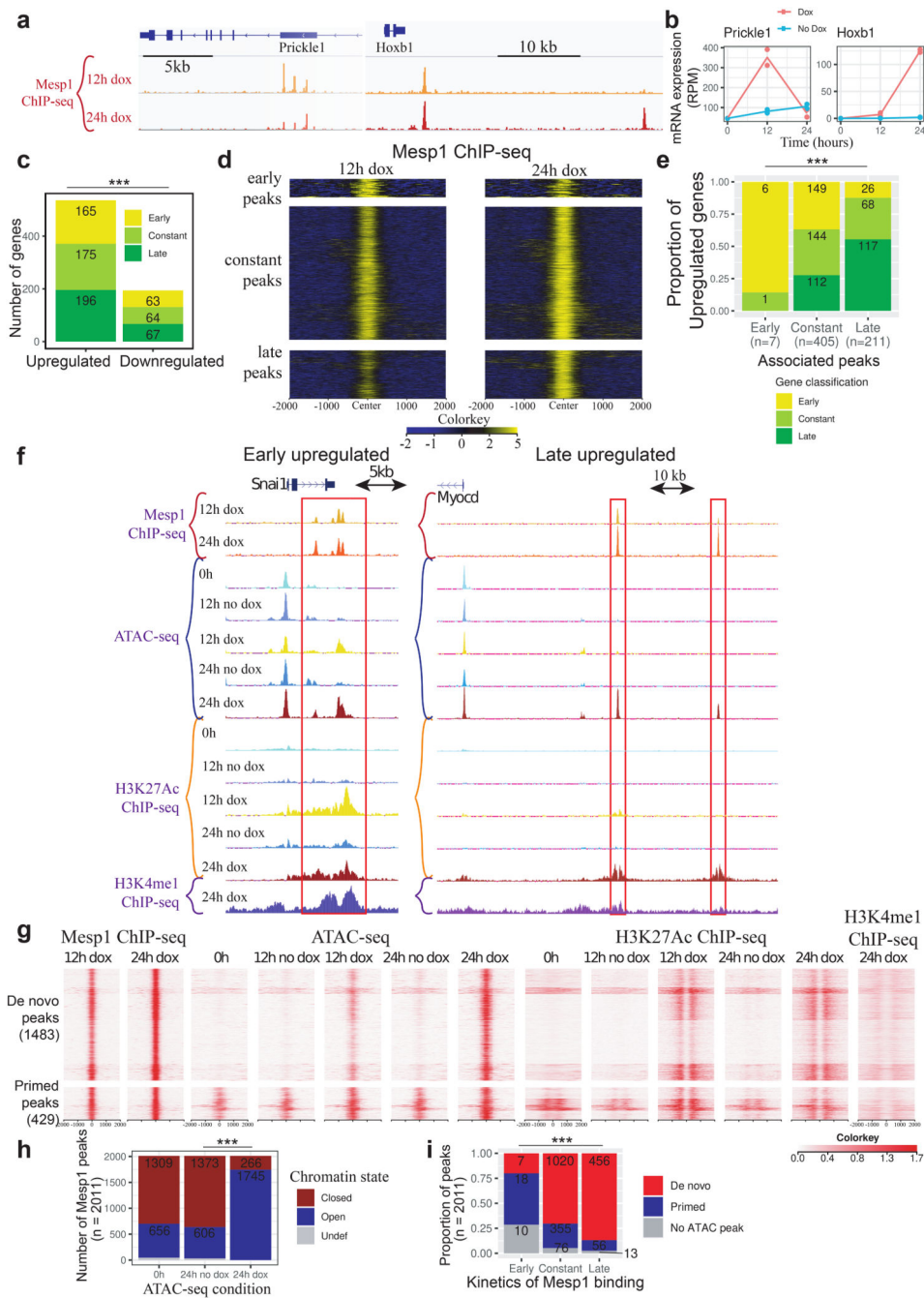
Figure 2. Temporal dynamic of chromatin remodelling regulated by Mesp1

a, Mesp1 ChIP-seq 12 and 24 hours after dox addition at 2.5 days of EB differentiation, n = 2 biologically independent experiments. **b**, Expression dynamics of the corresponding early (*Fgf15*) and late (*Hoxb1*) Mesp1 direct upregulated genes. n = 2 biologically independent experiments. **c**, Barplot illustrating the number of Mesp1 direct target upregulated and downregulated genes within early, constant and late regulated genes. *** The proportion of up- and downregulated genes among all genes were compared as categorical variables, by a two-tailored Z-test, where z=15.6501 and p <.00001.**d**, Heatmap showing normalized

signal at the 2011 Mesp1 ChIP-seq peaks 12 and 24 hours after dox. Peaks were separated into early, constant and late categories based on the ratio of normalized reads in these peaks at 12 and 24 hours. **e**, Number and proportion of early, constant and late Mesp1 regulated genes in the vicinity of early, constant and late Mesp1 ChIP-seq peaks. *** Significance was assessed by a chi-square test; the chi-square statistic is 66.48. The p-value is < 0.00001. **f**, Representative example of chromatin remodelling associated with Mesp1 binding, as measured by ATAC-seq(n = 2 biologically independent experiments) along with ChIP-seq for Mesp1 (n=2 independent experiments), $H_3K_{27}Ac$ (n=1 experiment) and H_3K_4me1 (n=1 experiment). **g**, Heatmap representing the signal for ATAC-seq,$H_3K_{27}Ac$, H_3K_4me1 and Mesp1 ChIP-seq data within all Mesp1 binding sites, separated into *de novo* and primed ATAC-seqpeaks. **h**, Quantification of overlap between Mesp1 ChIP-seq peaks and predicted nucleosome positioning in different ATAC-seq samples, as predicted by HMMR-ATAC[66]. The number of peaks predicted to overlap with nucleosome (closed), devoided of nucleosomes (open) or undefined (Undef) are indicated. *** The chi-square statistic is 1328.6. The p-value is < 0.00001. **i**, Barplotshowing the relationship between *de novo* opening of the chromatin and temporality of Mesp1 binding. No peak (grey) represents Mesp1 binding sites where the ATAC-seq signal was undetectable in all conditions. *** Chi-squared test was used to assess the significance of the number of peaks at different times and the *de novo* chromatin opening. The chi-square statistic is 124.3. The p-value is < 0.00001.
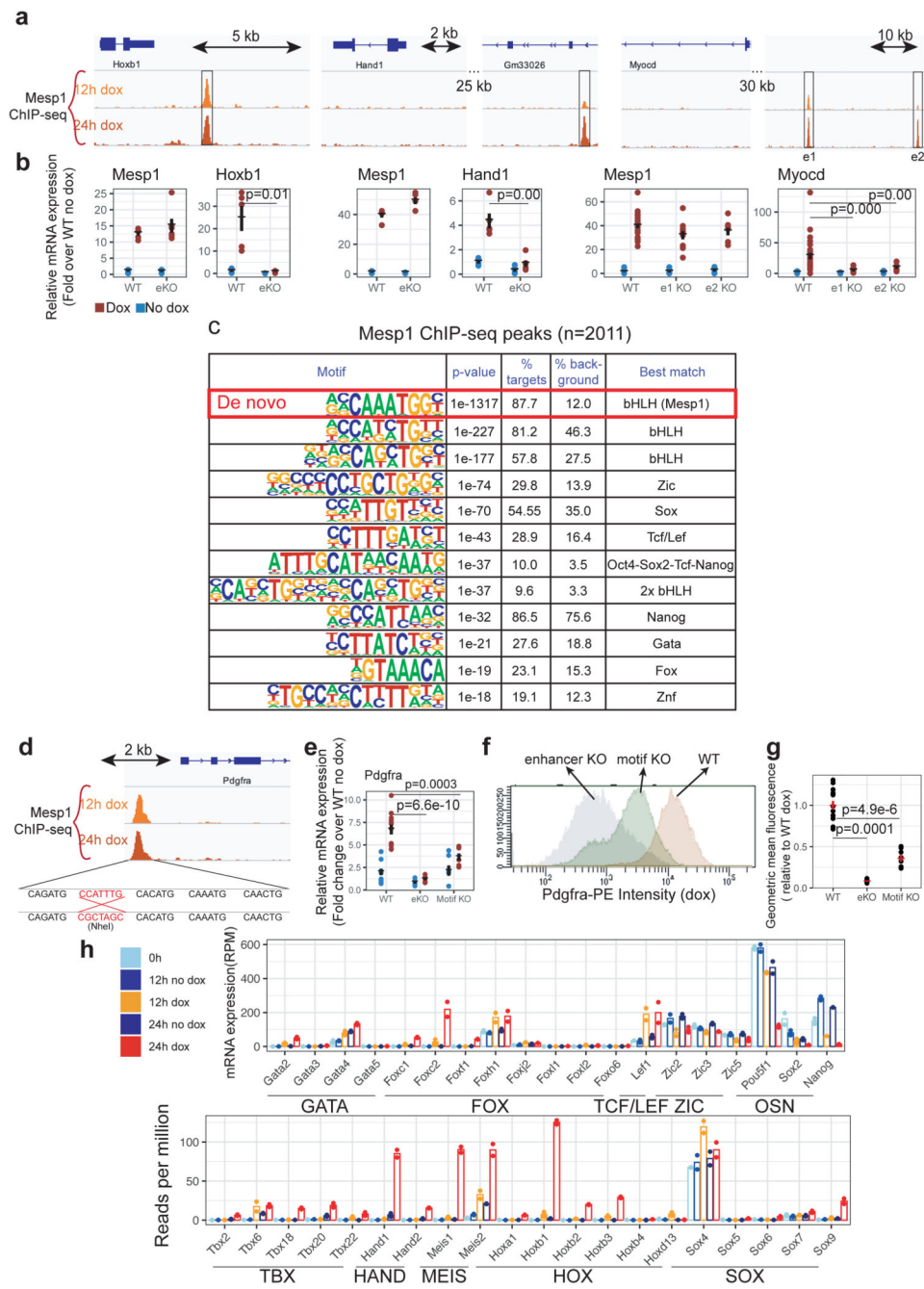
Figure 3. Characterization of Mesp1-bound enhancers and prediction of putative transcriptional cofactors

a, Genomic regions containing the four Mesp1 ChIP-seqbound enhancers that were deleted using CRISPR/Cas9. **b**, RT-qPCR showing the expression of Mesp1 target genes in WT and KO enhancer cell lines, with and without Mesp1 induction. Two-sided t-tests were used to compare gene expression in WT and KO cells. Horizontal line and error bar represent mean +/- standard error. n = 3 independent experiments. **c**, Motif enrichment analysis of all Mesp1 ChIP-seq peaks. Only motifs with p<10^{-10} and present in at least 10% of target

peaks are shown. The CAAATGG motif was found by *de novo* motif discovery, whereas the other motifs were known TF binding sites within the Homer database[22]. p-values are calculated through a binomial test. **d**, Locus containing the *Pdgfra* proximal enhancer. The CCATTTG motif of this enhancer was replaced by CGCTAGC. **e**, Expression of *Pdgfra* by RT-qPCR in cell lines with homozygous WT, KO or mutated (Motif KO) Mesp1-bound proximal *Pdgfra* enhancer. Horizontal line and error bar represent mean +/- standard error. n = 3 independent experiments for Hoxb1 and Hand1 enhancer KOs; n = 4 independent experiments for two KO and two WT cell lines for Myocd enhancer KO. RT-qPCR data in WT and KO cell lines were compared by a two-sided t-test. **f**, Histogram of PDGFRa protein expression quantified by FACS at day 4 of differentiation upon Dox addition. **g**, Quantification of PDGFRa expression after Mesp1 induction in WT, KO and mutated enhancer cell lines. Horizontal line and error bar represent mean +/- standard error. n = 5 independent experiments. Two-sided t-test was used to compare the values of geometric mean fluorescence. **h**, Expression values of TFs that have their motif enriched in Mesp1 ChIP-seq peaks or ATAC-seq peaks. n = 2 independent experiments.
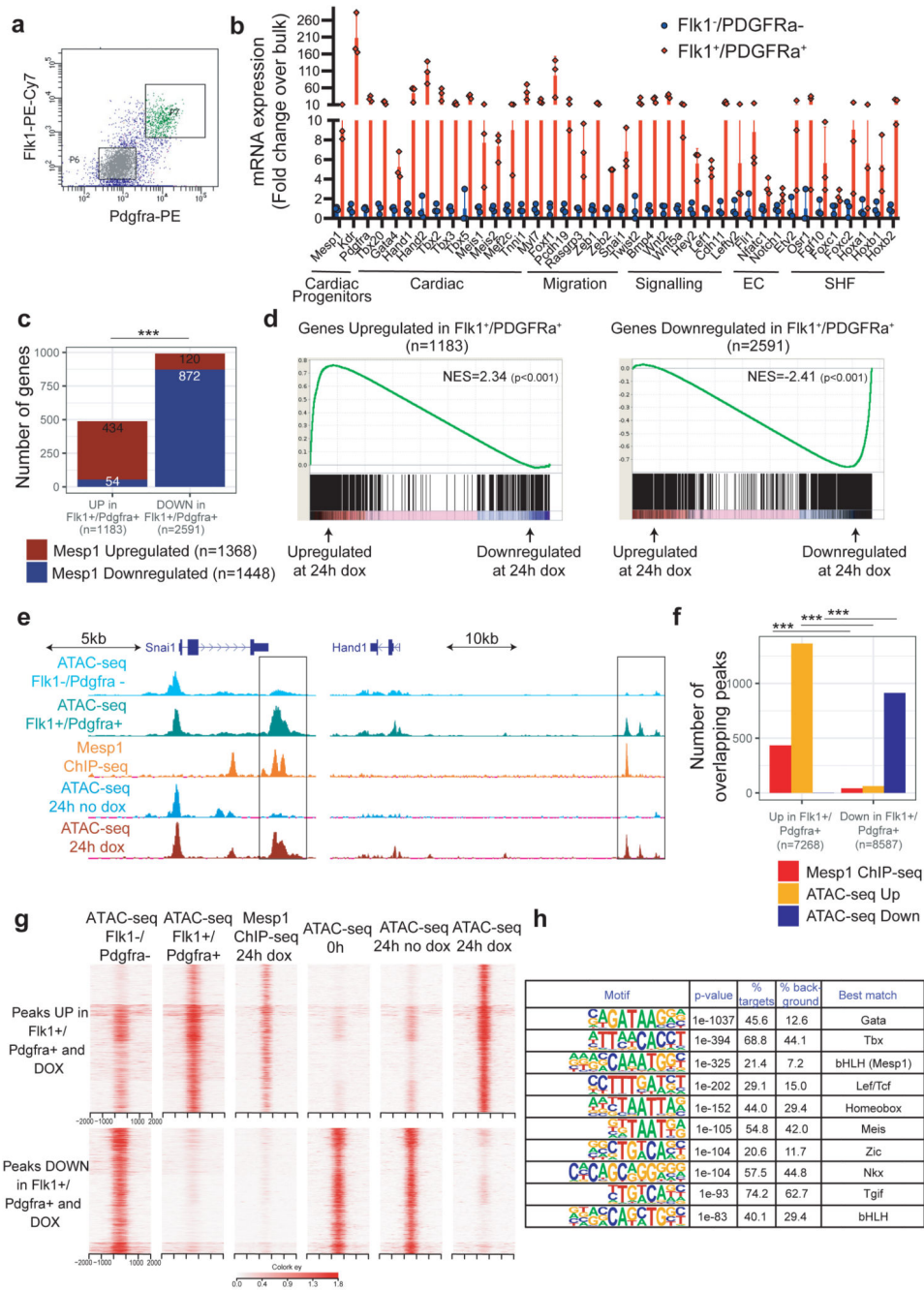
Figure 4. Validation of Mesp1 target genes and enhancer remodeling in the presence of endogenous Mesp1

a, FACS plot used to isolate Flk1$^+$/Pdgfra$^+$ (Mesp1 expressing cells) and Flk1$^-$/Pdgfra$^-$ cells at day 4 of PSC differentiation in the absence of Mesp1 overexpression. RNA-seq and ATAC-seq experimentswere performed in triplicates in these two populations. **b**, Examples of Mesp1 target genes with strong enrichment of their expression in Flk1$^+$/Pdgfra$^+$ versus Flk1$^-$/Pdgfra$^-$ cells (n=3 independent experiments). Error bars represent SEM. **c**, Quantification of the number of Mesp1 upregulated and downregulated genes

that are significantly enriched or depleted in Flk1$^+$/Pdgfra$^+$ versus Flk1$^-$/Pdgfra$^-$ cells. ***
Two-tailed Z-test statistic=13.007, p <0.00001. **d**, GSEA illustrating the enrichment of the
Mesp1 upregulated genes in Flk1$^+$/Pdgfra$^+$cells, and of Mesp1 downregulated genes in
Flk1$^-$/Pdgfra$^-$ cells. Nominal p-value represents the statistical significance of the enrichment
score, relative to a null distribution calculated through a permutation test. **e,** Representative
locus showing the opening of Mesp1-bound enhancers in Flk1$^+$/Pdgfra$^+$ and closed in Flk1$^-$/
Pdgfra$^-$ cells at day 4 of differentiation. **f**, Quantification of the number of Mesp1 ChIP-seq
peaks, as well as peaks that are selectively opened (UP) or closed (DOWN) after Mesp1
induction, within peaks enriched in Flk1$^+$/Pdgfra$^+$ of Flk1$^-$/Pdgfra$^-$ populations. *** A two-
tailored z-test was performed for all 3 comparisons, and all three were significantly different
between peaks UP and DOWN in Flk1$^+$/Pdgfra$^+$ cells, with p<0.00001 and respectively z =
20.2 (Mesp1 ChIP-seq), 39.6 (ATAC-seq UP), and -28.5 (ATAC-seq DOWN). **g**, Heatmaps
illustrating peaks that are both enriched or depleted in Flk1$^+$/Pdgfra$^+$ cells and upon
Mesp1 overexpression. **h**, Motif enrichment analysis on peaks enriched in Flk1$^+$/Pdgfra$^+$
cells, showing the strong enrichment of Mesp1 binding sites and binding sites of other
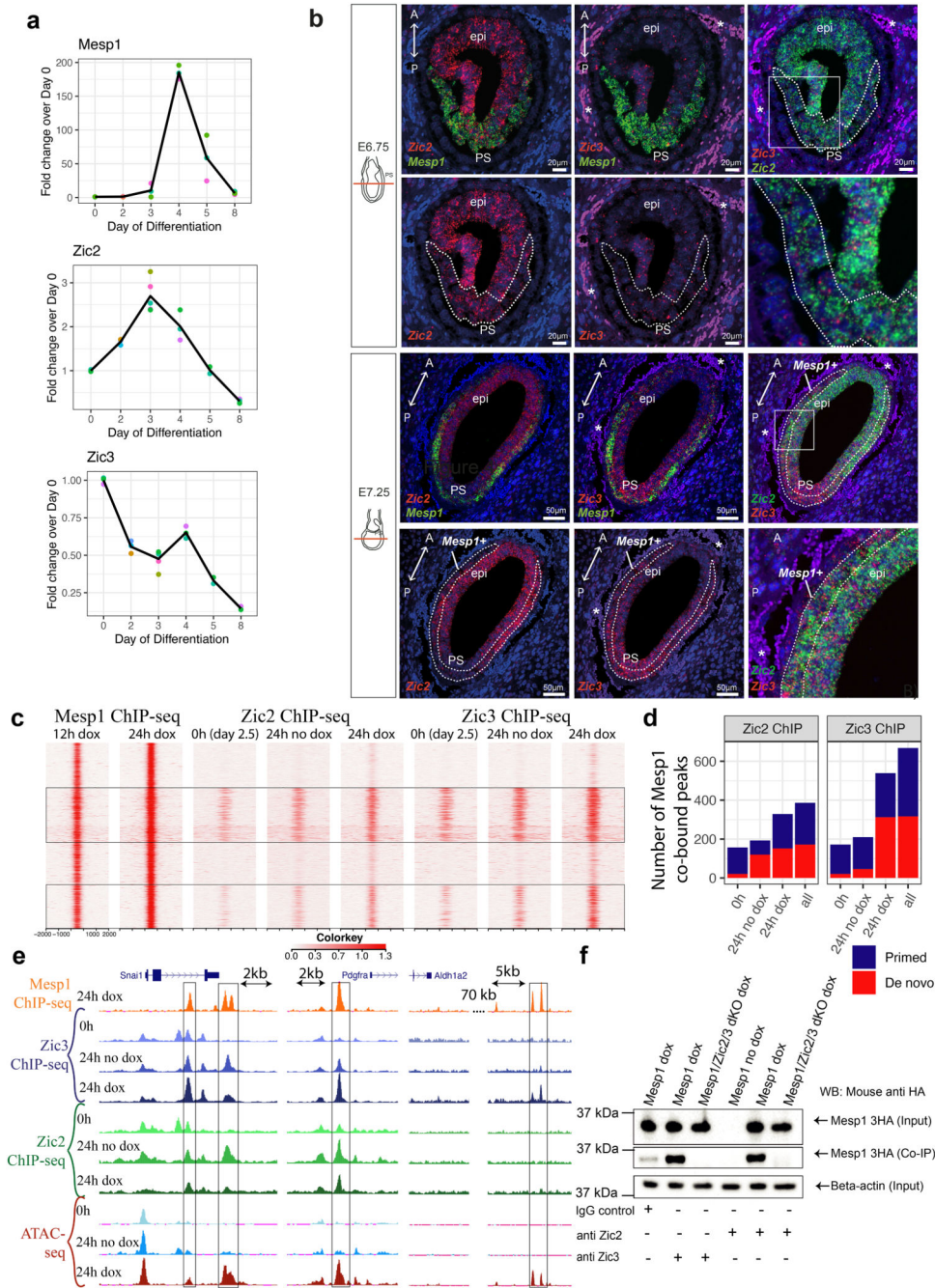cardiovascular TFs. p-values are calculated through a binomial test[22].

Figure 5. Zic2 and Zic3 bind to a fraction of Mesp1 bound enhancers

a, Expression profiles of *Mesp1*, *Zic2* and *Zic3* throughout differentiation of PSCs *in vitro* as measured by qRT-PCR. **b**, Single molecule RNA Fluorescent ISH(smRNA-FISH) of *Zic3* and *Zic2* together with *Mesp1* in transversal sections of gastrulating embryos *in vivo* at E6.75 and E7.25. epi, epiblast; PS, primitive streak; A, anterior; P, posterior. Dotted lines indicate *Mesp1* expression domains. * Indicates background fluorescence found in the decidua and that differs from signal found in the embryo. Data shown represent n = 4 stage-matched embryos from 3 independent litters. **c**, Heatmap representing

the signal of Mesp1, Zic2 and Zic3 ChIP-seq at different times within Mesp1-bound enhancers, demonstrating the co-occurrence of Zic2 and Zic3 as well as increasing signal as developmental time progresses. Mesp1 and Zic3 ChIP-seq analysis was performed in two biologically independent experiments; Zic2 ChIP-seq was performed once. **d**, Barplot representing the dynamic binding of Zic2 and Zic3 within Mesp1-bound enhancers, separated in primed versus *de novo* peaks based on ATAC-seq data presented in Fig. 2. **e**, Representative examples of enhancers co-occupied by Mesp1, Zic2 and Zic3, illustrating the diversity of dynamic binding of Zic2 and Zic3 within Mesp1 ChIP-seq peaks. **f**, Co-immunoprecipitation by using anti-Zic2, Zic3 or IgG control antibodies, followed by Western blot of anti-HA antibody revealing the presence of Mesp1-3HA protein. (data shown represent 3 independent experiments).
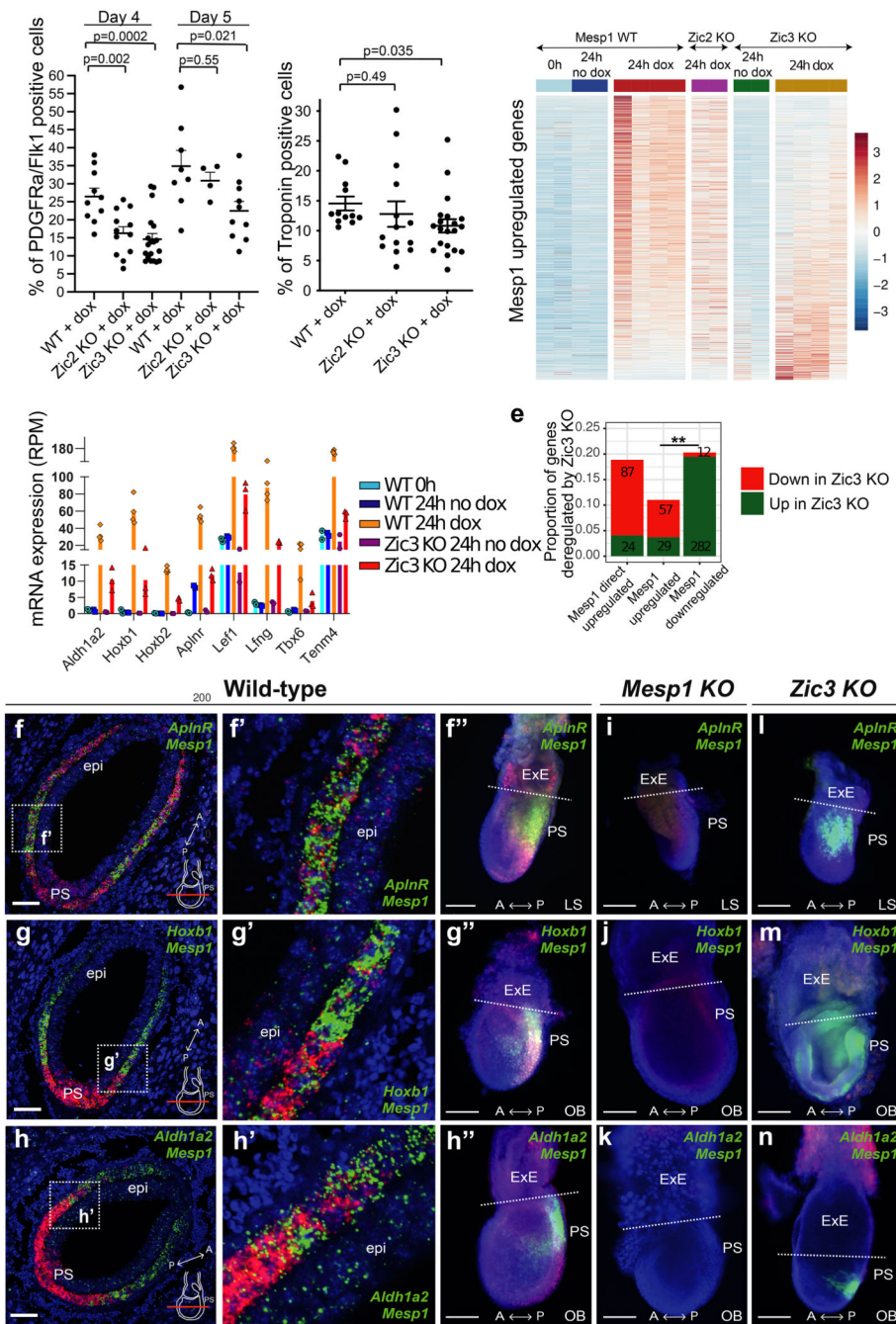
Figure 6. Mesp1 and Zic3 co-regulate gene expression during mouse gastrulation

a, Graph showing the number of Flk1/PDGFRa double positive cardiac progenitors upon Mesp1 overexpression (Dox) in WT, *Zic2*KO and *Zic3*KO cells at day 4 and 5 of PSC differentiation (n = 10 at day 4 and n = 4 at day 5 of biologically independent experiments.) (mean and SEM, unpaired two-sided t test p<0.0001). **b**, Proportion of Troponin-positive cardiomyocytes upon Mesp1 overexpression in WT, *Zic2* KO and *Zic3* KO cells at day 10 (n=12 for WT, n=14 for *Zic2 KO* and n=20 biologically independent experiments for *Zic3 KO*, mean with SEM, unpaired two-sided t test p<0.0001). **c**, Heatmap of the expression

of Mesp1 upregulated genes in WT, *Zic2 KO*, *Zic3* KO cells. Each column represents one sample. **d**, Direct Mesp1 upregulated genes downregulated in *Zic3* KO cell lines, as measured by RNA-seq. (n = 2 in no dox and n = 4 in dox conditions). **e**, Barplot showing the number and the proportion of genes regulated by Mesp1 up- or downregulated in Zic3 KO cell lines, using DESeq2[65] with cut-off values of 2-fold change and $p_{adj} < 0.05$. n = 4 biologically independent experiments. ** two-tailored Z-test with z=3.0679, p=0.00214. **f-n**, smRNA-FISH of *Mesp1* and *Aplnr, Hoxb1* or *Aldh1a2* in E7.5 embryos (representative of four embryos per genotype from different litters). The co-expression of *Mesp1* with *Aplnr* (**f-f''**), *Hoxb1* (**g-g''**) or *Aldh1a2* (**h-h''**) was assessed in WT embryos on sections (f, g, h (Scale bars= 50μm) - higher magnifications are found in f', g' and h') and in whole mount (f', g'', h'' - Scale bars= 200μm). **i-n**, Expression of *Aplnr, Hoxb1* or *Aldh1a2* in *Mesp1* KO and *Zic3* KO embryos at E7.5 at the late streak (LS) stage or no bud stage (OB). Decreased *Aplnr* expression was found in n=6/6 *Mesp1* KO embryos (i) and in 9/11 *Zic3* KO embryos (l). *Hoxb1* expression was decreased in n=4/4 *Mesp1* KO embryos (j) and 4/5 *Zic3* KO embryos (m). *Aldh1a2* expression was decreased in n=4/4 *Mesp1* KO embryos (k) and in n=5/8 *Zic3* KO embryos (n). A, anterior; P, posterior; epi, epiblast; EXE, extra-embryonic region; PS, primitive streak. Scale bars= 200μm.
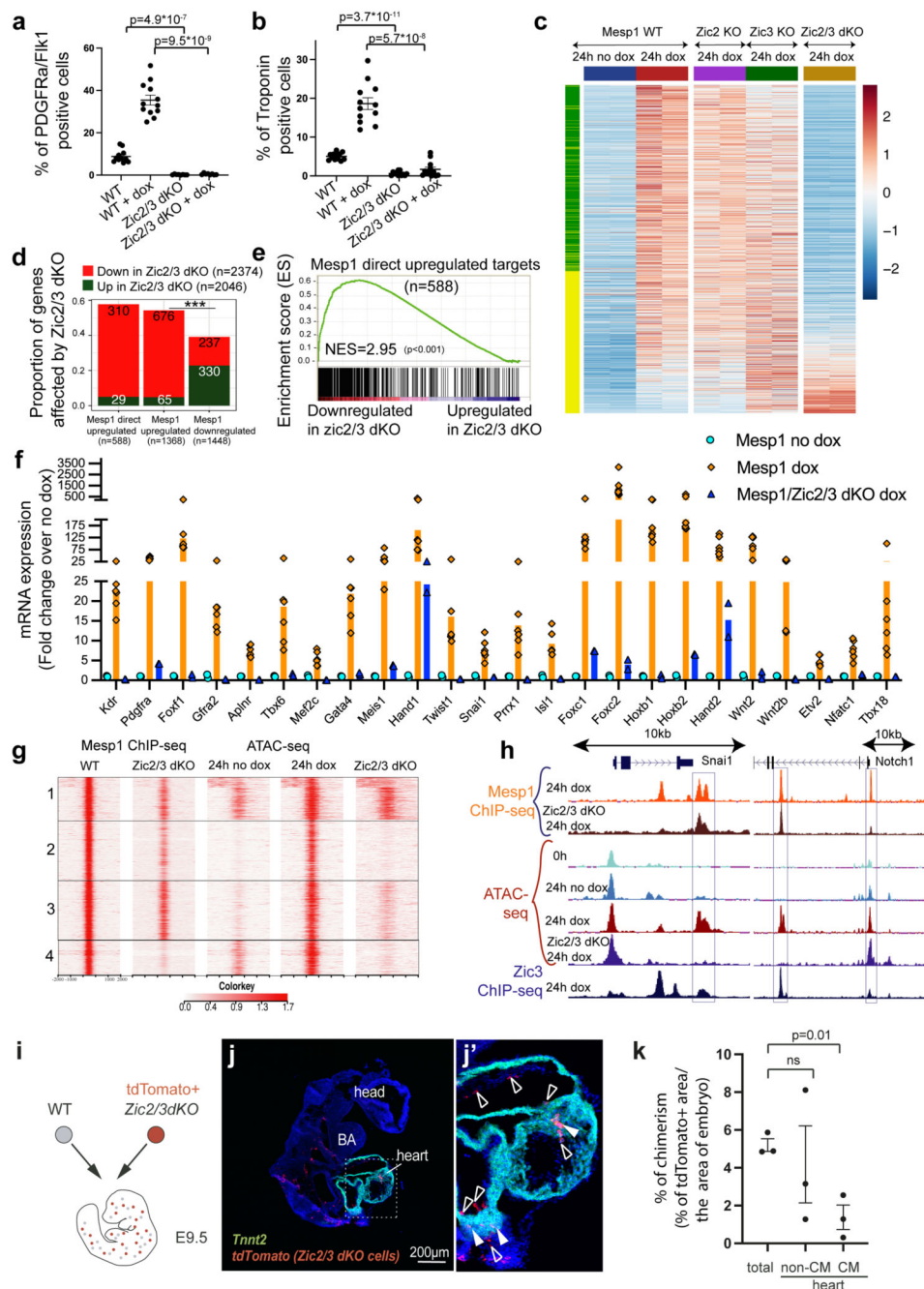
Figure 7. Zic3 and Zic2 are essential regulators of Mesp1 activity

a, Proportion of Flk1+/PDGFRa+ WT and *Zic2/Zic3* dKO cells at day 4 (n=12 WT, n=12 WT+dox, n=8 *Zic2/3*dKO and n=8 *Zic2/3*dKO+dox biologically independent experiments, mean with SEM), unpaired student two-tailed t test. **b**, Proportion of TNNT2 positive cardiomyocytes in WT and *Zic2/Zic3* dKO cells at day 10 (n= 12 biologically independent experiments, mean with SEM). Unpaired student t test. **c**, Heatmap of the expression of Mesp1 upregulated genes in WT and *Zic2/Zic3* dKO cells. Each column represents one sample. **d**, Barplot showing the number and proportion of Mesp1 regulated genes up- or

downregulated in *Zic2/Zic3* dKO cells. n = 2 biologically independent experiments. *** z=16.166; p < 0.00001. **e**, GSEA showing the expression of Mesp1 direct upregulated genes up or downregulated in *Zic2/3*dKO cells 24h after dox. **f**, Examples of genes downregulated in Zic2/3 dKO cells (n=2 for WT no dox and dox, n=5 for Zic2/3 dKO of biologically independent experiments. **g**, Heatmap of Mesp1 ChIP-seq and ATAC-seq peaks in WT (=2) and *Zic2/3*dKO (n = 3 biologically independent experiments) cells 24h after Mesp1 induction. A, primed peaks with maintained Mesp1 binding in Zic2/3 dKO cells; 2, de novo peaks with loss of Mesp1 binding; 3, primed peaks with loss of Mesp1 binding in dKO cells; 4, de novo peaks with sustained Mesp1 binding but no chromatin opening. **h**, Illustrative examples of Mesp1 Chip-seq and ATAC-seq in *Zic2/3*dKO cells, where chromatin opening upon Mesp1 induction is reduced. **i**, Schematic overview of the chimeric experiment. **j,** smRNA-FISH on section of a E9.5 chimeric embryo showing *tdTomato* and *Tnnt2* expression (representative picture of 3 independent embryos). BA, branchial arch. Scale bar= 200μm. j' higher magnification on the cardiac region. Empty arrowheads show *Tnnt2* negative *Zic2/3*dKO *tdTomato*+ cells corresponding to ECs. **k**, Percentage of Tomato+ chimerism in the whole embryo (total), in CM and in non-CM as measured by the ratio between the area of *tdTomato*+ pixels and the area of DAPI staining nuclei (n=10, 12 and 13 slides examined over 3 biologically independent embryos, mean with SEM). Two-tail paired student *t* tests.