ARTICLES
https://doi.org/10.1038/s41556-022-00947-3

nature cell biology

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

Xionghui Lin[1,10], Benjamin Swedlund[1,10], Mai-Linh N. Ton[2,3], Shila Ghazanfar[4], Carolina Guibentif[2,3,5], Catherine Paulissen[1], Elodie Baudelet[1], Elise Plaindoux[6], Younes Achouri[7], Emilie Calonne[8], Christine Dubois[1], William Mansfield[3], Stéphane Zaffran[6], John C. Marioni[4], Francois Fuks[8], Berthold Göttgens[2,3], Fabienne Lescroart[6,11] ✉ and Cédric Blanpain[1,9,11] ✉

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 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 transcription-factor 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.

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, endocardial cells, fibroblasts and smooth muscle cells[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, single-cell RNA sequencing (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 programmes during CP specification and differentiation in vitro and in vivo. Using RNA sequencing (RNA-seq), chromatin immunoprecipitation and sequencing (ChIP–seq) and assay for transposase-accessible chromatin using sequencing (ATAC-seq),

[1]Laboratory of Stem Cells and Cancer, Université Libre de Bruxelles, Brussels, Belgium. [2]Department of Haematology, Jeffrey Cheah Biomedical Centre, University of Cambridge, Cambridge, UK. [3]Wellcome and Medical Research Council Cambridge Stem Cell Institute, University of Cambridge, Cambridge, UK. [4]Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, UK. [5]Sahlgrenska Center for Cancer Research, Department of Microbiology and Immunology, University of Gothenburg, Gothenburg, Sweden. [6]Aix Marseille Université, INSERM, MMG U1251, Marseille, France. [7]Université Catholique de Louvain, Institut de Duve, Brussels, Belgium. [8]Laboratory of Cancer Epigenetics, Faculty of Medicine, ULB-Cancer Research Center (U-CRC), Université Libre de Bruxelles (ULB), Brussels, Belgium. [9]WELBIO, Université Libre de Bruxelles, Brussels, Belgium. [10]These authors contributed equally: Xionghui Lin, Benjamin Swedlund. [11]These authors jointly supervised this work: Fabienne Lescroart, Cédric Blanpain. ✉e-mail: fabienne.lescroart@univ-amu.fr; cedric.blanpain@ulb.ac.be

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 h after doxycycline (dox)-induced Mesp1 expression on day 2.5 of embryoid body (EB) differentiation, slightly earlier than endogenous *Mesp1* expression (Fig. 1a). *Mesp1* RNA fluorescent in situ hybridization (FISH) showed that dox-mediated Mesp1 overexpression induces *Mesp1* expression in almost every cell (89 ± 2%) in the EB, whereas only 22 ± 4% of the cells express *Mesp1* in control conditions on day 3.5. The level of *Mesp1* expression per cell was on average 2.2 higher in dox compared with 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 h) induced by Mesp1, called 'early genes' (for example, *Snai1*); genes whose expression increased constantly between 0 and 24 h, called 'constant genes' (for example, *Pdgfra*) and genes that were only activated after 24 h, called 'late genes' (for example, *Hoxb1*). A similar temporal pattern of early, constant and late downregulated genes could be identified (Fig. 1b–d and 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.

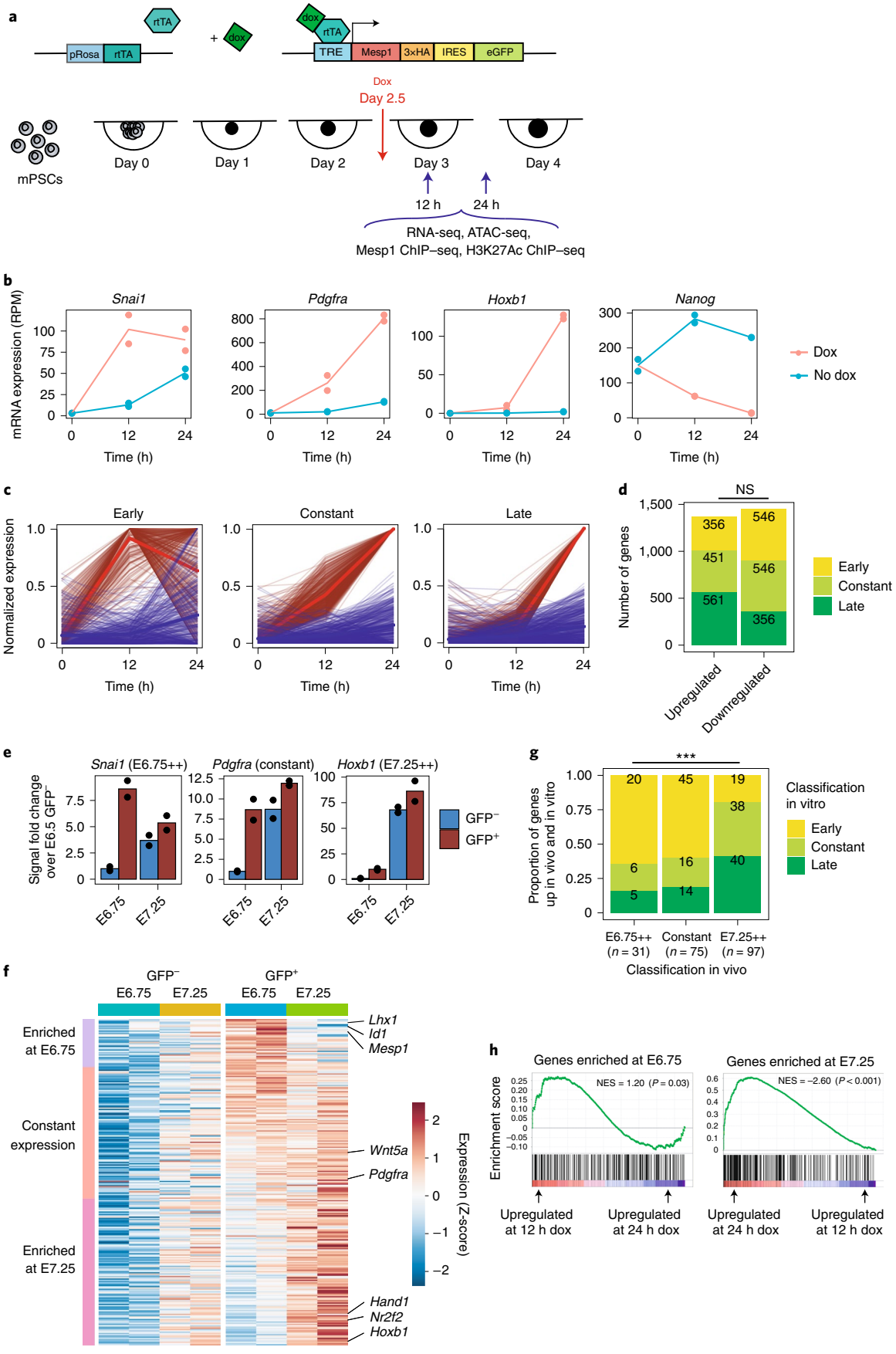
Dynamics of chromatin and enhancer remodelling regulated by Mesp1. To investigate the temporal dynamics by which Mesp1 induces chromatin and enhancer remodelling associated with CP specification, we assessed the temporality of Mesp1 binding after its induction during PSC differentiation. To this end, we performed Mesp1 ChIP–seq 12 and 24 h after overexpression of triple-haemagglutinin (HA) tagged Mesp1 in PSCs induced at day 2.5 of EB differentiation (Figs. 1a and 2a). We defined a total of 2,011 high-quality peaks, of which 823 were detected at 12 h and 1,792 at 24 h ($P < 10^{-10}$). 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 39% (536/1,368) of upregulated genes presented a Mesp1 ChIP–seq peak within 500 kb of their transcription start site, versus 13% (194/1,448) of downregulated genes, suggesting that Mesp1 acts mainly as a transcriptional activator (Fig. 2a–c).

Mesp1 ChIP–seq peaks could be classified into early (36/2,011, 2%), constant (1,451/2,011, 72%) and late (525/2,011, 26%) Mesp1 binding sites. The kinetics of Mesp1 binding correlated with the kinetics of gene expression, with early peaks predominantly 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 (H3K27Ac) ChIP–seq at 0, 12 and 24 h 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 (H3K27Ac) 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 compared with 13% after Mesp1 expression (Fig. 2h). Chromatin opening at these sites was associated with the temporality of Mesp1 binding, as only 20% of the early Mesp1-bound peaks were classified as de novo peaks versus 70% of the constant and 87% of the 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 used clustered regularly interspaced short palindromic repeats (CRISPR)–CRISPR associated protein 9 nickase (Cas9n) to delete different Mesp1 binding regions (±500 base pairs, bp) presenting the hallmarks of enhancer activation after Mesp1 induction in regulatory regions of three different

Fig. 1 | Dynamics of gene expression regulated by Mesp1. a, Experimental scheme allowing *Mesp1* overexpression and the timing of the RNA-seq, ChIP–seq and ATAC-seq analyses. The Mesp1-3HA mPSC line was used and Mesp1 over expression induced at day 2.5 by addition of doxycycline (dox). **b**, Expression pattern of genes that are regulated by Mesp1, representing four categories of target genes: early, constant, late upregulated and downregulated. RPM, reads per million. **c**, Individual gene expression kinetics in dox (red) and no dox (blue) conditions (represented by thin lines) as well as the average profile of all genes (thick lines). **d**, Number of deregulated genes called using DESeq2 with a cutoff of 1.5-fold change in expression and an adjusted P (P_{adj}) < 0.05 (ref. [65]). NS, non-significant difference between the number of up- and downregulated genes shown in the graph. **e**, Representative expression levels (from $n = 2$ 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++; left) or E7.25 (E7.25++; right), or equally expressed in both populations (constant; middle), 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**, Number of genes upregulated in vivo (at E6.75, E7.25 or both) and in vitro at the early, constant and late time points. ***$P < 0.00001$, significance of the difference in the number of genes shown was assessed using a χ^2 test; χ^2 statistic, 36.7904; $n = 2$ independent experiments. **h**, Gene-set enrichment analysis showing the distribution of genes upregulated in GFP$^+$ cells in vivo that are enriched at E6.75 (left) or E7.25 (right) within the ranking of all genes based on their expression following 12 h of dox treatment versus 24 h of treatment. The nominal P values represent the statistical significance of the enrichment score relative to a null distribution calculated through a permutation test; NES, normalized enrichment score. **b**–**h**, $n = 2$ biologically independent experiments.


a
rtTA
pRosa
rtTA
dox
dox
rtTA
TRE
Mesp1
3×HA
IRES
eGFP
Dox
Day 2.5
mPSCs
Day 0
Day 1
Day 2
Day 3
Day 4
12 h
24 h
RNA-seq, ATAC-seq,
Mesp1 ChIP–seq, H3K27Ac ChIP–seq
b
Snai1
Pdgfra
Hoxb1
Nanog
mRNA expression (RPM)
Time (h)
Dox
No dox
c
Early
Constant
Late
Normalized expression
Time (h)
d
NS
Number of genes
356
451
561
546
546
356
Upregulated
Downregulated
Early
Constant
Late
e
Snai1 (E6.75++)
Pdgfra (constant)
Hoxb1 (E7.25++)
Signal fold change over E6.5 GFP⁻
GFP⁻
GFP⁺
E6.75
E7.25
g
***
Proportion of genes up in vivo and in vitro
20
6
5
45
16
14
19
38
40
Classification in vitro
Early
Constant
Late
E6.75++ (n = 31)
Constant (n = 75)
E7.25++ (n = 97)
Classification in vivo
f
GFP⁻
GFP⁺
E6.75
E7.25
E6.75
E7.25
Enriched at E6.75
Constant expression
Enriched at E7.25
Lhx1
Id1
Mesp1
Wnt5a
Pdgfra
Hand1
Nr2f2
Hoxb1
Expression (Z-score)
h
Genes enriched at E6.75
Genes enriched at E7.25
Enrichment score
NES = 1.20 (P = 0.03)
NES = −2.60 (P < 0.001)
Upregulated at 12 h dox
Upregulated at 24 h dox
Upregulated at 24 h dox
Upregulated at 12 h dox

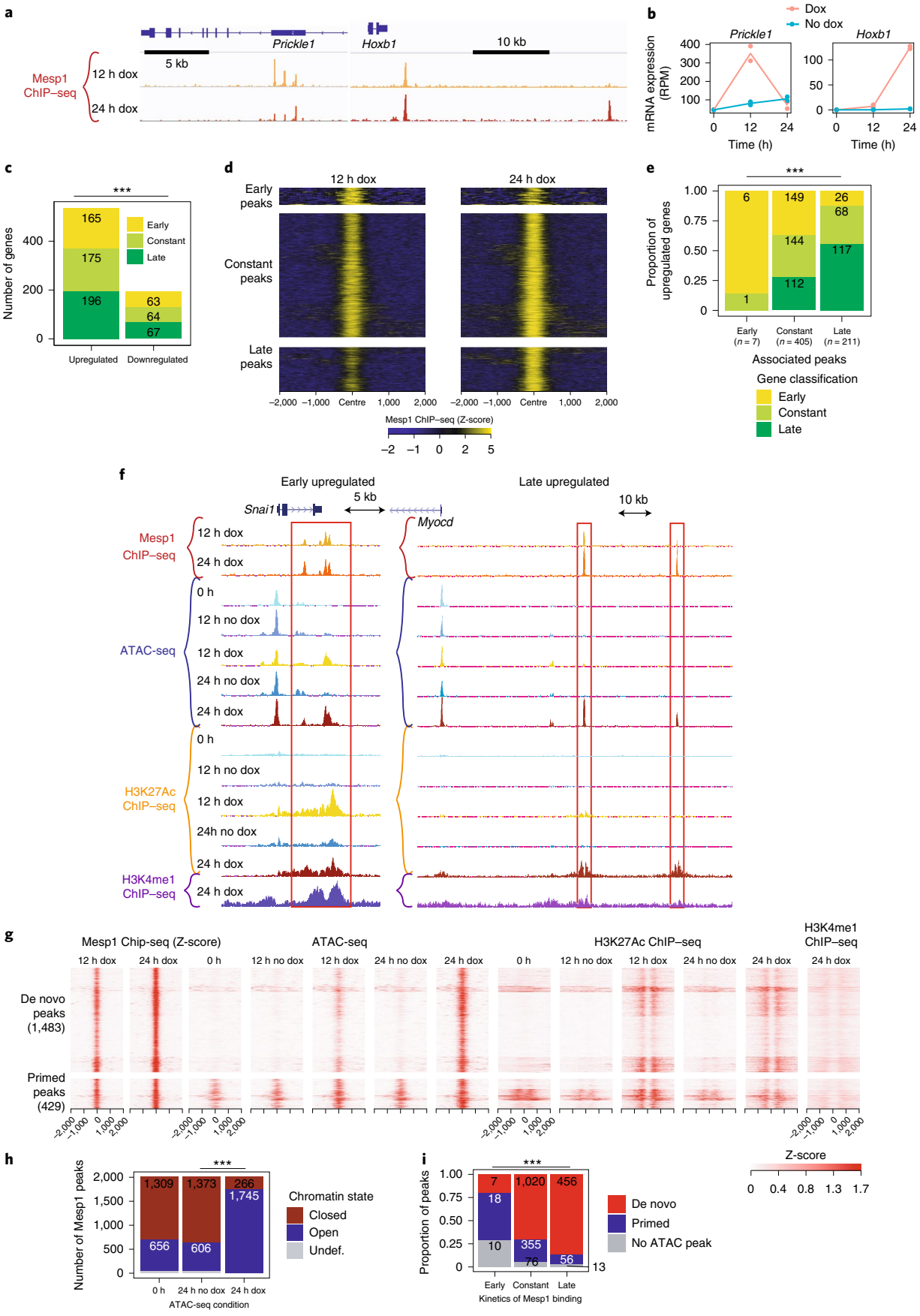

a
Prickle1
Hoxb1
5 kb
10 kb
Mesp1 ChIP–seq
12 h dox
24 h dox
b
Dox
No dox
Prickle1
Hoxb1
mRNA expression (RPM)
Time (h)
c
***
Number of genes
Early
Constant
Late
165
175
196
63
64
67
Upregulated
Downregulated
d
12 h dox
24 h dox
Early peaks
Constant peaks
Late peaks
Centre
Mesp1 ChIP–seq (Z-score)
e
***
Proportion of upregulated genes
6
1
149
144
112
26
68
117
Early (n = 7)
Constant (n = 405)
Late (n = 211)
Associated peaks
Gene classification
Early
Constant
Late
f
Early upregulated
Late upregulated
Snai1
Myocd
5 kb
10 kb
Mesp1 ChIP–seq
ATAC-seq
H3K27Ac ChIP–seq
H3K4me1 ChIP–seq
0 h
12 h no dox
24h no dox
g
Mesp1 Chip-seq (Z-score)
ATAC-seq
H3K27Ac ChIP–seq
H3K4me1 ChIP–seq
De novo peaks (1,483)
Primed peaks (429)
h
***
Number of Mesp1 peaks
1,309
1,373
266
1,745
656
606
Chromatin state
Closed
Open
Undef.
ATAC-seq condition
i
***
Proportion of peaks
7
18
10
1,020
355
76
456
56
13
De novo
Primed
No ATAC peak
Kinetics of Mesp1 binding
Z-score

Fig. 2 | Temporal dynamic of chromatin remodelling regulated by Mesp1. a, Mesp1 ChIP–seq 12h (yellow) and 24 h (red) after dox addition at 2.5 days of EB differentiation. Genome tracks of *Prickle1* and *Hoxb1* are shown (top). **b**, Expression dynamics of the corresponding early (*Fgf15*) and late (*Hoxb1*) Mesp1 direct upregulated genes. RPM, reads per million. **a**,**b**, $n=2$ biologically independent experiments. **c**, 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 using a two-tailed z-test, where $z=15.6501$. **d**, Heatmap showing normalized signals at the 2,011 Mesp1 ChIP–seq peaks following 12 (left) and 24 h (right) of dox treatment. The peaks were separated into early, constant and late categories based on the ratio of normalized reads in these peaks at 12 and 24 h. **e**, Number (indicated in the bars) and proportion of early, constant and late Mesp1-regulated genes in the vicinity of early, constant and late Mesp1 ChIP–seq peaks. **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 ($n=2$ biologically independent experiments), H3K27Ac ($n=1$ experiment) and H3K4me1 ($n=1$ experiment) with genome tracks of *Snai1* and *Myocd* (top). **g**, Heatmap representing the signal for ATAC-seq, H3K27Ac, H3K4me1 and Mesp1 ChIP–seq data within all Mesp1 binding sites separated into de novo and primed ATAC-seq peaks. **h**, Quantification of the overlap between Mesp1 ChIP–seq peaks and predicted nucleosome positioning in different ATAC-seq samples, as predicted by HMMRATAC[66]. The number of peaks predicted to overlap with nucleosome (closed), devoid of nucleosomes (open) or undefined (undef.) are indicated. **i**, 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. **e**,**h**,**i**, Significance was assessed using a χ^2 test; χ^2 statistic, 66.48 (**e**), 1,328.6 (**h**) and 124.3 (**i**). **h**,**i**, $n=2{,}011$ total peaks. $^{***}P<0.00001$.

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—that is, CAAATGG (Fig. 3c)—that was equally present in early, constant and late ChIP–seq peaks but notably enriched in de novo peaks (90%) compared with 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, a bHLH pioneer TF promoting reprogramming of fibroblasts to neuronal cell fate[23,24].

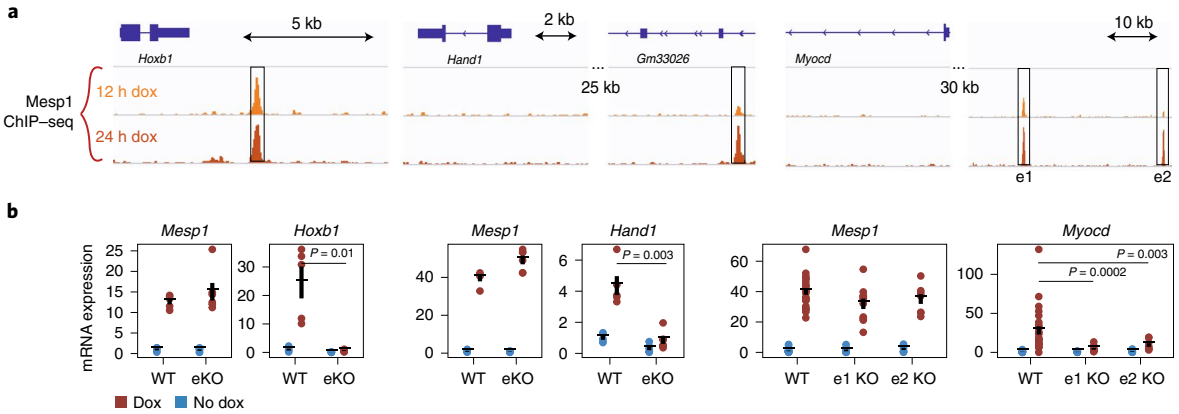
Mesp1-bound enhancers contained an average of five bHLH binding motifs. To assess the importance of the most over-represented bHLH motif with an extra G within the Mesp1 binding sites, we replaced one CAAATGG motif with GCTAGCG in the *Pdgfra* proximal enhancer using CRISPR–Cas9n-mediated homology directed repair (Fig. 3d). The replacement of this single motif led to a threefold decrease in the ability of Mesp1 to induce *Pdgfra* expression, both at the transcriptional and protein levels, 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, Gata 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 the pluripotency TFs Oct4, Sox2 and Nanog in the 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 (for example, *Gata4*, *Hand1* and *-2*, *Meis1* and *-2*, *Hoxb1*, and *Foxc1* and *-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 h following Mesp1 expression were not directly bound by Mesp1 (899/3,839, 23%), 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. Furthermore, Gata4 also bound to many (444/899, 49%) 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).

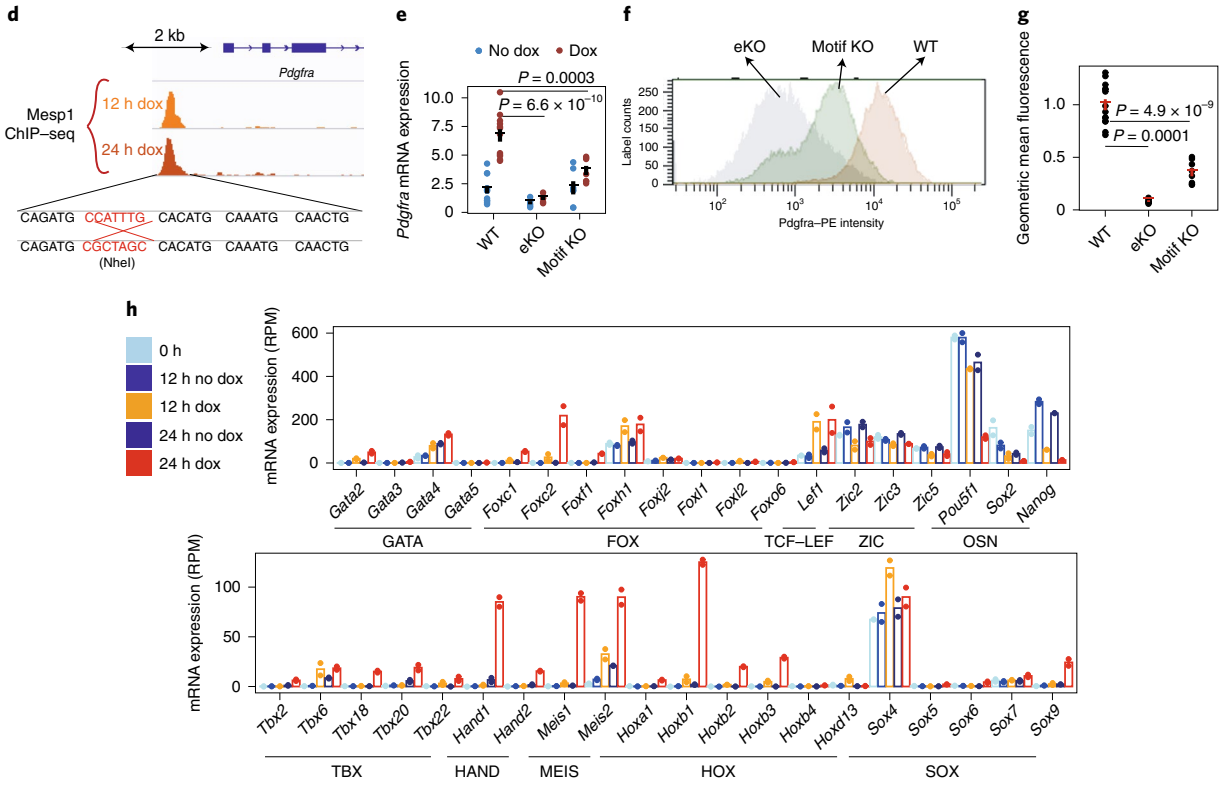
We also identified 2,772 peaks that were associated with closing of the chromatin and lower levels of flanking H3K27Ac without presenting Mesp1 binding. These peaks were found in the vicinity of downregulated genes, including the pluripotency TFs *Oct4* and *Sox2*, and the epithelial genes *Cdh1* and *Epcam*. Motif-enrichment analysis on these repressed peaks revealed strong enrichment of a Oct4–Sox2–Tcf–Nanog compound motif. Sox2, Oct4 and Nanog

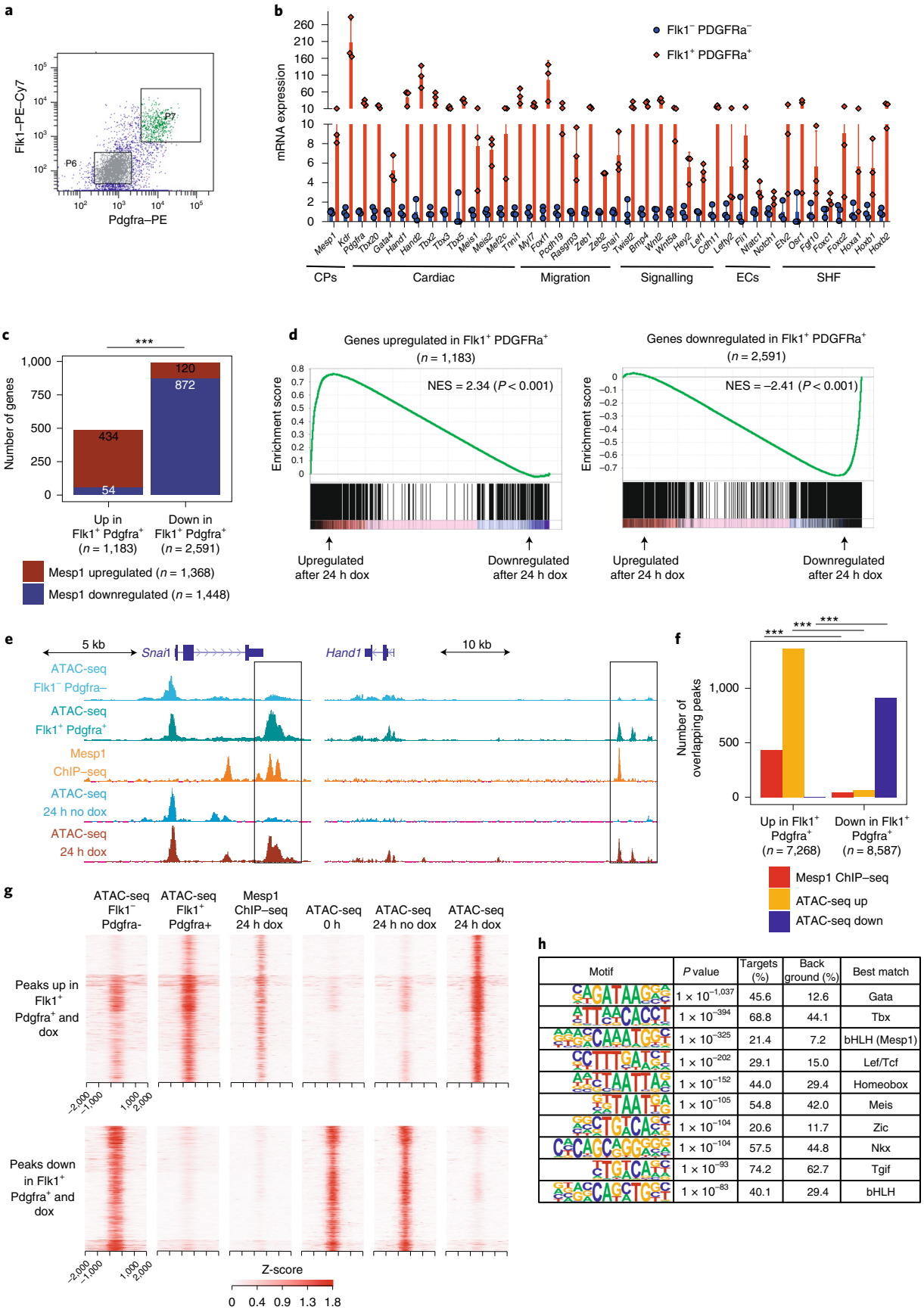
Fig. 3 | Characterization of Mesp1-bound enhancers and prediction of putative transcriptional cofactors. a, Genomic regions (shown with genome browser views on top) containing the four Mesp1 ChIP–seq bound enhancers that were deleted using CRISPR–Cas9n (black boxes). e1, e2, enhancers of *Myocd*. **b**, Levels of expression, determined using quantitative PCR with reverse transcription (RT–qPCR), of Mesp1 target genes in wild-type (WT) and enhancer-KO (eKO) cell lines with and without Mesp1 induction. Data represent the mean ± s.e.m. of the fold change normalized to the WT without dox treatment; $n=3$ independent experiments. A two-sided Student's t-test was used to compare gene expression between WT and KO cells. **c**, Motif enrichment analysis of all Mesp1 ChIP–seq peaks ($n=2{,}011$). 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 in the Homer database[22]. P values were calculated using a binomial test. **d**, Locus containing the *Pdgfra* proximal enhancer (top). The CCATTTG motif of this enhancer was replaced by CGCTAGC (bottom). **e**, *Pdgfra* expression, determined using RT–qPCR, in cell lines with homozygous WT, eKO or mutated (motif KO) Mesp1-bound proximal *Pdgfra* enhancer. Data represent the mean ± s.e.m. of the fold change normalized to the WT with no dox treatment; $n=3$ (*Hoxb1*- and *Hand1*-enhancer KOs) and 4 (two KO and two WT cell lines for Myocd-enhancer KO) independent experiments. The RT–qPCR data in the WT and KO cell lines were compared using a two-sided Student's t-test. **f**, Histogram of Pdgfra protein expression quantified by FACS on day 4 of differentiation following dox addition. **g**, Levels of Pdgfra expression after Mesp1 induction in the WT, eKO and mutated enhancer (motif KO) cell lines. Data represent the mean ± s.e.m.; $n=5$ independent experiments. A two-sided Student's t-test was used to compare the values of the geometric mean fluorescence relative to the WT with dox treatment. **h**, Expression values of TFs that have their motif enriched in Mesp1 ChIP–seq peaks or ATAC-seq peaks. RPM, reads per million; $n=2$ independent experiments.



c

Motif	*P* value	Targets (%)	Background (%)	Best match
De novo	$1 \times 10^{-1,317}$	87.7	12.0	bHLH (Mesp1)
	1×10^{-227}	81.2	46.3	bHLH
	1×10^{-177}	57.8	27.5	bHLH
	1×10^{-74}	29.8	13.9	Zic
	1×10^{-70}	54.55	35.0	Sox
	1×10^{-43}	28.9	16.4	Tcf–Lef
	1×10^{-37}	10.0	3.5	Oct4–Sox2–Tcf–Nanog
	1×10^{-37}	9.6	3.3	2× bHLH
	1×10^{-32}	86.5	75.6	Nanog
	1×10^{-21}	27.6	18.8	Gata
	1×10^{-19}	23.1	15.3	Fox
	1×10^{-18}	19.1	12.3	Znf





Motif	*P* value	Targets (%)	Back ground (%)	Best match
	$1 \times 10^{-1,037}$	45.6	12.6	Gata
	1×10^{-394}	68.8	44.1	Tbx
	1×10^{-325}	21.4	7.2	bHLH (Mesp1)
	1×10^{-202}	29.1	15.0	Lef/Tcf
	1×10^{-152}	44.0	29.4	Homeobox
	1×10^{-105}	54.8	42.0	Meis
	1×10^{-104}	20.6	11.7	Zic
	1×10^{-104}	57.5	44.8	Nkx
	1×10^{-93}	74.2	62.7	Tgif
	1×10^{-83}	40.1	29.4	bHLH

Fig. 4 | Validation of Mesp1 target genes and enhancer remodelling in the presence of endogenous Mesp1. a, FACS plot of Flk1$^+$Pdgfra$^+$ (Mesp1-expressing, P7) and Flk1$^-$Pdgfra$^-$ (P6) cell isolation on day 4 of PSC differentiation in the absence of Mesp1 overexpression. RNA-seq and ATAC-seq experiments were performed in triplicate on these two populations. **b**, Examples of Mesp1 target genes with strong enrichment of their expression in Flk1$^+$Pdgfra$^+$ versus Flk1$^-$Pdgfra$^-$ cells. Data are the fold change in expression over the bulk. The error bars represent the s.e.m.; $n=3$ independent experiments. EC, endocardial cells. SHF, second heart field. **c**, Number of Mesp1-induced upregulated and downregulated genes that were notably enriched or depleted in Flk1$^+$Pdgfra$^+$ compared with Flk1$^-$Pdgfra$^-$ cells. Two-tailed *z*-test, $z=13.007$. **d**, Gene-set enrichment analysis illustrating the enrichment of the Mesp1-induced upregulated genes in Flk1$^+$Pdgfra$^+$ cells (left) and downregulated genes in Flk1$^-$Pdgfra$^-$ cells. The nominal *P* values represent the statistical significance of the enrichment score relative to a null distribution calculated through a permutation test; NES, normalized enrichment score. **e**, Representative locus (with genome tracks) showing the opening of Mesp1-bound enhancers in Flk1$^+$Pdgfra$^+$ cells and closed in Flk1$^-$Pdgfra$^-$ cells on day 4 of differentiation. **f**, Number of Mesp1 ChIP–seq peaks as well as peaks that were selectively opened (up) or closed (down) after Mesp1 induction within peaks enriched in the Flk1$^+$Pdgfra$^+$ or Flk1$^-$Pdgfra$^-$ cell populations. A two-tailed *z*-test was performed for all three comparisons and all three were notably different between peaks up and down in Flk1$^+$Pdgfra$^+$ cells; $z=20.2$ (Mesp1 ChIP–seq), 39.6 (ATAC-seq up) and −28.5 (ATAC-seq down). **c**,**f**, $^{***}P<0.00001$. **g**, Heatmaps illustrating peaks that were both enriched or depleted in Flk1$^+$Pdgfra$^+$ cells and following 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 were calculated using a binomial test[22].

ChIP–seq[32] showed that these core pluripotency TFs bound around 60% of the enhancers that were indirectly repressed by Mesp1 (Extended Data Fig. 4a–c).

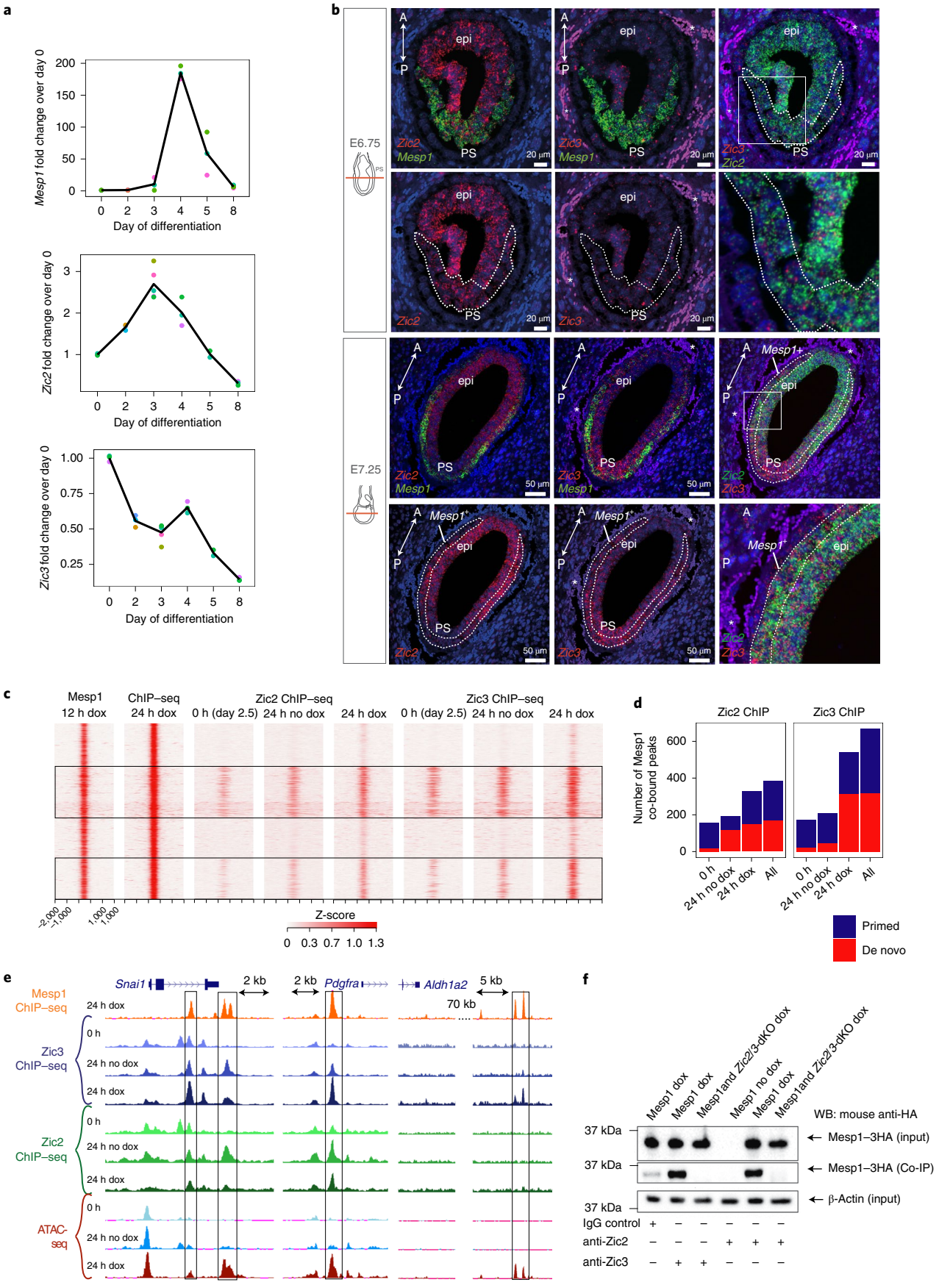
Chromatin landscape and transcriptional regulation at physiological levels of Mesp1. To assess whether these findings are relevant in the absence of Mesp1 overexpression, we performed RNA-seq and ATAC-seq on Flk1 and Pdgfra double-positive and double-negative cells at day 4 of PSC differentiation (Fig. 4a); these cells 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 notably enriched in Flk1$^+$Pdgfra$^+$ cells. On the other hand, 60% of the genes downregulated by Mesp1 were depleted in Flk1$^+$Pdgfra$^+$ cells (Fig. 4b–d). Through ATAC-seq on fluorescence-activated cell sorting (FACS)-isolated Flk1$^+$Pdgfra$^+$ and Flk1$^-$Pdgfra$^-$ cells, we found that 44% of the Mesp1-bound sites were opened in the Flk1$^+$Pdgfra$^+$ cells, 50% of which (435/2,011) were notably enriched in the Flk1$^+$Pdgfra$^+$ cells (Fig. 4e–g). Motif discovery on all peaks enriched in Flk1$^+$Pdgfra$^+$ cells revealed strong enrichment of Gata, T-box and Mesp1 motifs (Fig. 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 naive pluripotent state[35]—and performed Mesp1 ChIP–seq, RNA-seq and ATAC-seq. We found that only 0.4% (6/1,368) of the Mesp1 target genes were notably upregulated 24 h after Mesp1 overexpression in 2i. In 2i conditions, Mesp1 notably bound to 13% of its binding sites identified during PSC differentiation, although low levels of Mesp1 binding occurred at many of its binding sites. The proportion of primed versus de novo binding sites in 2i was similar to that 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 remodelling were unaffected by 2i conditions, Mesp1 overexpression did not result in gene upregulation (Extended Data Fig. 5a–g). Together, these data reveal the importance of the cellular context in Mesp1 binding, chromatin remodelling and transcription regulation mediated by Mesp1.

Zic2 and Zic3 are enriched at Mesp1-bound enhancers. We performed TF motif discovery at Mesp1-bound enhancers to uncover the TFs that cooperate with Mesp1 to regulate its binding, chromatin remodelling and regulation of target gene expression. 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 mice and humans 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].

Both RT–qPCR and western blotting experiments showed that Zic2 and Zic3 are expressed during PSC differentiation in vitro, at the same time as *Mesp1* (Fig. 5a and Extended Data Fig. 6a). Single-molecule RNA-FISH (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).

Fig. 5 | Zic2 and Zic3 bind to a fraction of Mesp1-bound enhancers. a, Expression profiles of *Mesp1* (top), *Zic2* (middle) and *Zic3* (bottom) throughout in vitro differentiation of PSCs, as measured by RT–qPCR. **b**, In vivo images of smRNA-FISH of *Zic3* and *Zic2* together with *Mesp1* in transverse sections of gastrulating embryos at E6.75 and E7.25. PS, primitive streak; A, anterior; P, posterior; and epi, epiblast. Dotted lines indicate *Mesp1* expression domains. Background fluorescence found in the decidua that differs from signal found in the embryo is indicated with an asterisk. Images represent $n=4$ stage-matched embryos from three independent litters. Magnified views (X 2.7) of the regions in the white boxes (right) are shown beneath the respective images. On the left, level of the sections in E6.75 or E7.25 embryos. **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 analyses were performed in two biologically independent experiments; Zic2 ChIP–seq was performed once. **d**, Representation of the dynamic binding of Zic2 and Zic3 within Mesp1-bound enhancers, separated in primed versus de novo peaks based on the ATAC-seq data presented in Fig. 2. **e**, Representative examples of enhancers co-occupied by Mesp1, Zic2 and Zic3 (black box), illustrating the diversity of dynamic binding of Zic2 and Zic3 within Mesp1 ChIP–seq peaks. Genome tracks are shown on top. **f**, Co-immunoprecipitation (Co-IP) using antibodies to Zic2 and Zic3 or IgG control antibodies, followed by western blotting (WB) using anti-HA, revealing the presence of Mesp1–3HA protein. Blot is representative of $n=3$ independent experiments.


a
Mesp1 fold change over day 0
Day of differentiation
Zic2 fold change over day 0
Day of differentiation
Zic3 fold change over day 0
Day of differentiation
b
E6.75
E7.25
epi
PS
Zic2
Mesp1
Zic3
Mesp1+
20 μm
50 μm
c
Mesp1 12 h dox
ChIP–seq 24 h dox
Zic2 ChIP–seq
Zic3 ChIP–seq
0 h (day 2.5)
24 h no dox
24 h dox
Z-score
d
Zic2 ChIP
Zic3 ChIP
Number of Mesp1 co-bound peaks
0 h
24 h no dox
24 h dox
All
Primed
De novo
e
Snai1
Pdgfra
Aldh1a2
2 kb
5 kb
70 kb
Mesp1 ChIP–seq
Zic3 ChIP–seq
Zic2 ChIP–seq
ATAC-seq
f
Mesp1 dox
Mesp1 dox
Mesp1and Zic2/3-dKO dox
Mesp1 no dox
Mesp1 dox
Mesp1and Zic2/3-dKO dox
WB: mouse anti-HA
37 kDa
Mesp1–3HA (input)
Mesp1–3HA (Co-IP)
β-Actin (input)
IgG control
anti-Zic2
anti-Zic3

To validate our bioinformatic prediction that Mesp1, Zic2 and Zic3 bind 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 h 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 the Mesp1-bound enhancers (19% for Zic2 and 29% for Zic3), in agreement with the bioinformatic prediction (Fig. 5d,e). 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, 83% (127/153) of the Zic3 and Mesp1 co-binding sites had their chromatin opened before Mesp1 binding (primed peaks), whereas the other 17% had their chromatin opened following Mesp1 overexpression (Fig. 5d,e and 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, 60% (345/577) of the Mesp1 and Zic3 co-bound peaks were bound by Zic3 and presented chromatin opening only following Mesp1 overexpression (group 3; Extended Data Fig. 6d). Together, these data show the temporality and dynamic nature of the distinct groups of Mesp1 and Zic3 co-binding sites and their impact on chromatin remodelling.

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

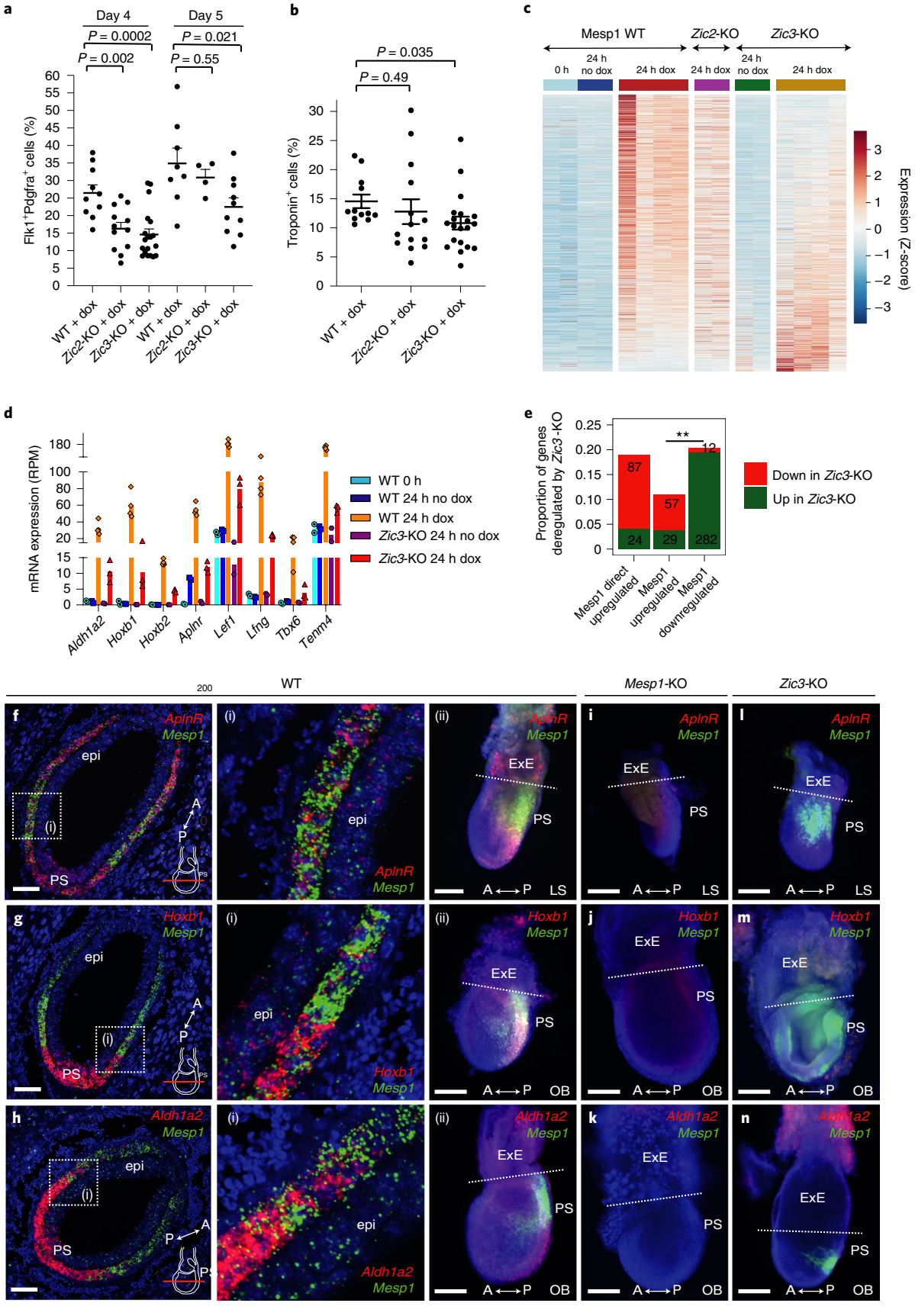
Zic3 regulates Mesp1 transcriptional activity at a subset of mesodermal enhancers. To assess whether Zic2 and Zic3 regulate Mesp1 function, we deleted *Zic2* or *Zic3* using CRISPR–Cas9n in PSCs allowing dox-induced Mesp1 overexpression in the context of *Zic2* or *Zic3* knockout (KO). Although Zic2 and Zic3 have been proposed to regulate pluripotency[44,45], we did not find changes in the expression of core pluripotency TFs when *Zic2*- and *Zic3*-KO PSCs were cultured in LIF/2i medium (Extended Data Fig. 7a). We then assessed whether the deletion of either *Zic2* or *Zic3* 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]. Following Mesp1 overexpression, the number of Flk1$^+$Pdgfra$^+$ CPs was slightly decreased in the *Zic2*- and *Zic3*-KO PSCs on day 4 and in *Zic3*-KO PSCs on 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 cardiomyocytes was slightly decreased in the *Zic3*-KO cell lines but no significant difference was found in the *Zic2*-KO cells (Fig. 6b and Extended Data Fig. 7b).

To unravel the mechanisms by which Zic TFs regulate the transcriptional activity of Mesp1, we performed RNA-seq on *Zic2*- and *Zic3*-KO cells 24 h after Mesp1 overexpression. Single *Zic2*-KO cells only presented 11 genes that were notably differentially expressed following Mesp1 expression compared with WT cells. In contrast, 15% (87/588) of directly upregulated Mesp1 target genes were notably downregulated in *Zic3*-KO cells following Mesp1 overexpression, including several genes known to be important for specification of the different heart lineages (for example, *Aldh1a2*, *Hoxb1*, *Aplnr* and *Tenm4*; refs. [46–51]; Fig. 6c–e). Moreover, several genes controlling pluripotency and other cell lineages (for example, *Nodal*, *Nanog*, *Sox2* and *Foxa2*) were upregulated in *Zic3*-KO cells (Extended Data Fig. 7c). These data show that Zic3 alone regulates the expression of some of the 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 the regulation of 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. From these genes, we selected a pan marker of CPs, *Aplnr*[49–51], as well as *Alhd1a2* and *Hoxb1*, two known markers of the posterior second heart field, 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. Through smRNA-FISH of these three genes in WT embryos at E7.25, we observed that they co-localize with *Mesp1* in the mesodermal cells leaving the primitive streak and migrating towards the anterior pole of the embryo (Fig. 6f–h). The expression of these genes was strongly downregulated in *Mesp1*-null embryos (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. Using CRISPR–Cas9n, we generated *Zic3*-KO mouse lines by injecting Cas9 protein and guide RNA directly into zygotes (Extended Data Fig. 8a). We obtained several founders with a large deletion or insertion, thereby disrupting the *Zic3* gene within its first exon, upstream of the DNA-binding-domain coding sequence.

Fig. 6 | Mesp1 and Zic3 co-regulate gene expression during mouse gastrulation. a, Proportion of Flk1$^+$Pdgfra$^+$ CPs following Mesp1 overexpression (dox-induced) in WT, *Zic2*-KO and *Zic3*-KO cells on days 4 and 5 of PSC differentiation; $n = 10$ (day 4) and 4 (day 5) biologically independent experiments. **b**, Proportion of Troponin$^+$ cardiomyocytes following Mesp1 overexpression in WT, *Zic2*-KO and *Zic3*-KO cells on day 10 of PSC differentiation; $n = 12$ (WT), 14 (*Zic2*-KO) and 20 (*Zic3*-KO) biologically independent experiments. **a**,**b**, Data are the mean ± s.e.m.; unpaired two-sided Student's *t*-test. **c**, Heatmap of the expression of Mesp1-upregulated genes in WT, *Zic2*-KO and *Zic3*-KO cells. Each column represents one sample. **d**, Levels of expression of genes downregulated in the *Zic3*-KO cell lines that were directly upregulated by Mesp1. Expression levels were measured through RNA-seq. RPM, reads per million; $n = 2$ (no dox) and 4 (dox). **e**, Number (indicated in the bars) and proportion of genes regulated by Mesp1 that were up- or downregulated in the *Zic3*-KO cell lines, determined using DESeq2 (ref. [65]) with cut-offs of a twofold change and $P_{adj} < 0.05$. $^{**}P = 0.00214$; two-tailored *z*-test, $z = 3.0679$; $n = 4$ biologically independent experiments. **f**–**h**, Images showing smRNA-FISH of *Mesp1* with *Aplnr* (**f**), *Hoxb1* (**g**) and *Aldh1a2* (**h**) in E7.5 embryos. Images are representative of four embryos per genotype from different litters. The co-expression of *Mesp1* with *Aplnr* (**f**), *Hoxb1* (**g**) and *Aldh1a2* (**h**) was assessed in sections (main images and (i), magnified views of the regions in the main images outlined by a white box (X3.4 for **f**, X4.3 for **g**,**h**)) and whole mounts (ii) of WT embryos. Scale bars, 50 µm (main images, left) and 200 µm (ii). **i**–**n**, Expression of *Aplnr* (**i**,**l**), *Hoxb1* (**j**,**m**) and *Aldh1a2* (**k**,**n**) in *Mesp1*-KO (**i**–**k**) and *Zic3*-KO (**l**–**n**) embryos at E7.5 at the late streak or no-bud stages. **i**,**l**, Decreased *Aplnr* expression was found in 6/6 *Mesp1*-KO embryos (**i**) and 9/11 *Zic3*-KO embryos (**l**). **j**,**m**, *Hoxb1* expression was decreased in 4/4 *Mesp1*-KO embryos (**j**) and 4/5 *Zic3*-KO embryos (**m**). **k**,**n**, *Aldh1a2* expression was decreased in 4/4 *Mesp1*-KO embryos (**k**) and 5/8 *Zic3*-KO embryos (**n**). Dotted lines in **f**–**h** (ii) and **i**–**n** delimit the embryonic and extra-embryonic (ExE) regions. A, anterior; P, posterior; epi, epiblast; PS, primitive streak; LS, late streak stage; OB, no-bud stage. Scale bars, 200 µm.


a
Day 4
Day 5
P = 0.0002
P = 0.021
P = 0.002
P = 0.55
Flk1+Pdgfra+ cells (%)
WT + dox
Zic2-KO + dox
Zic3-KO + dox
b
P = 0.035
P = 0.49
Troponin+ cells (%)
c
Mesp1 WT
Zic2-KO
Zic3-KO
0 h
24 h no dox
24 h dox
Expression (Z-score)
d
mRNA expression (RPM)
WT 0 h
WT 24 h no dox
WT 24 h dox
Zic3-KO 24 h no dox
Zic3-KO 24 h dox
Aldh1a2
Hoxb1
Hoxb2
Aplnr
Lef1
Lfng
Tbx6
Tenm4
e
Proportion of genes deregulated by Zic3-KO
**
Down in Zic3-KO
Up in Zic3-KO
Mesp1 direct upregulated
Mesp1 upregulated
Mesp1 downregulated
WT
Mesp1-KO
Zic3-KO
AplnR Mesp1
Hoxb1 Mesp1
Aldh1a2 Mesp1
epi
ExE
PS
A
P
LS
OB

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) and hypoplasia of the right ventricle (5/32). Moreover, the *Zic3*-KO embryos showed a thinner compact myocardial layer at E14.5, in agreement with the decrease in cardiomyocyte differentiation found in *Zic3*-KO PSCs. On the other hand, the endocardial layer was unaffected in *Zic3*-KO hearts (Extended Data Fig. 8b–f). *Hoxb1*, *Aldh1a2* and *Aplnr* were found to be downregulated during gastrulation in *Zic3*-KO embryos (4/5 for *Hoxb1*, 5/8 for *Aldh1a2* and 9/11 for *Apln*) through smRNA-FISH of these genes (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 and 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* and *Zic3* double-KO (*Zic2/3*-dKO) cell lines in Mesp1-inducible PSCs. Western blotting 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 LIF/2i medium and sustain the expression of core pluripotency TFs (Extended Data Fig. 9a). The specification of Flk1$^+$PDGFRa$^+$ CPs and their terminal differentiation into cardiomyocytes 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, Zic2 and Zic3, we performed RNA-seq on *Zic2/3*-dKO cells 24 h after Mesp1 induction. Strikingly, 53% (310/588) of all direct target genes upregulated by Mesp1 and 49% (675/1,368) of genes directly and indirectly upregulated by Mesp1 were strongly downregulated following Mesp1 overexpression in the *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 and Zic3 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% of the Mesp1 binding sites, enriched for the late Mesp1-binding peaks (Fig. 7g and Extended Data Fig. 9c, d).

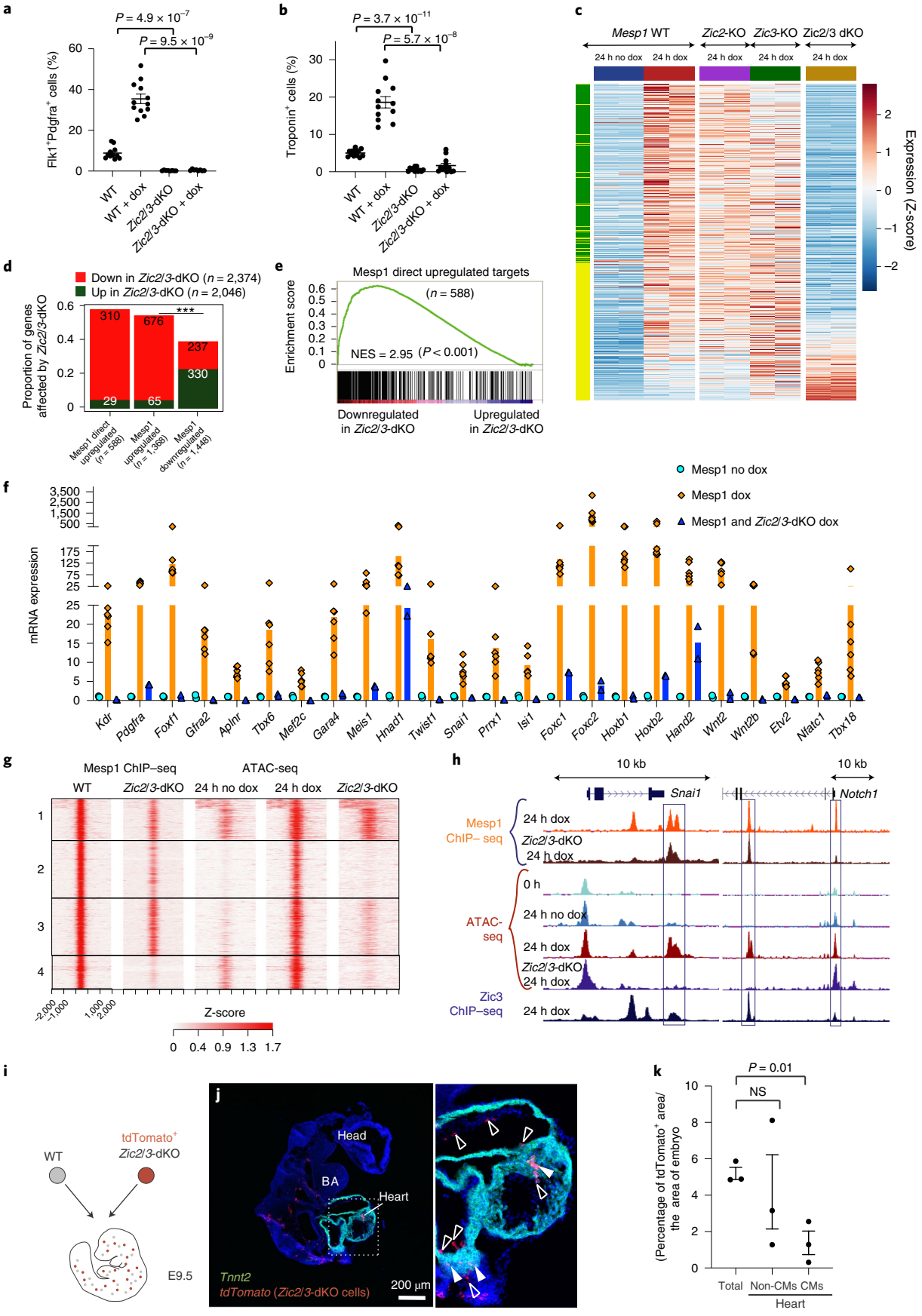
Chromatin profiling of *Zic2/3*-dKO PSCs following Mesp1 overexpression by ATAC-seq revealed a major defect of Mesp1-induced chromatin remodelling, with the signal of 56% (2,130/3,839) of the enhancers remodelled following Mesp1 overexpression and 48% of Mesp1-bound enhancers being strongly decreased (Fig. 7g,h and Extended Data Fig. 9e). Many of these enhancers were still bound by Mesp1 in the absence of Zic2 and Zic3 but their chromatin regions were no longer opened by Mesp1 binding (group 3), showing that Zic2 and Zic3 are important for the regulation of Mesp1-mediated chromatin opening. Motif discovery on the 3,558 downregulated 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 *Zic2/3* in tdTomato-expressing PSCs, injected these cells in WT embryos of the same genetic background and analysed the chimeric embryos at E9.5 when the heart is formed (Fig. 7i). We analysed three embryos from two litters of E9.5 chimeric embryos with a significant percentage of chimerism. In similar chimeric experiments, the contribution of WT tdTomato$^+$ cells to the different embryonic lineages showed no difference in their differentiation potential compared with WT cells[53,54]. We found a decrease in the contribution of tdTomato$^+$ *Zic2/3*-dKO cells to the heart of chimeric embryos with *Zic2/3* dKO (Tomato$^+$) that presented at least 5% of tdTomato$^+$ cells. Although the chimerism of tdTomato$^+$ cells in the embryo was 5.2 ± 0.3%, only 1.4 ± 0.6% of tdTomato$^+$ cells were found in cardiomyocytes, showing a significant decrease of *Zic2/3*-dKO cells in cardiomyocytes. The contribution to the endocardium was more variable, precluding drawing strong conclusions due to the low number of chimeric mice analysed (Fig. 7i–k). These data further support a role for Zic2 and Zic3 during specification of CPs and cardiomyocyte differentiation. Together, these data indicate that Zic2 and Zic3 act redundantly to regulate Mesp1 binding and Mesp1-induced chromatin remodelling, which are crucial for CP specification and differentiation.

Discussion

During mouse gastrulation, cells transit from a pluripotent state to a lineage-committed state within 24 h. 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

Fig. 7 | Zic3 and Zic2 are essential regulators of Mesp1 activity. a, Proportion of Flk1$^+$Pdgfra$^+$ WT and *Zic2/3*-dKO cells on day 4; $n = 12$ (WT and WT + dox) and 8 (*Zic2/3*-dKO and *Zic2/3*-dKO + dox) biologically independent experiments. **b**, Proportion of Troponin$^+$ cardiomyocytes in WT and *Zic2/3*-dKO cells on day 10; $n = 12$ biologically independent experiments. **a**,**b**, Data are the mean ± s.e.m.; unpaired two-sided Student's *t*-test. **c**, Expression of Mesp1-upregulated genes in WT and *Zic2/3*-dKO cells. Each column represents one sample. **d**, Number (indicated in the bars) and proportion of Mesp1 regulated genes up- or downregulated in *Zic2/3*-dKO cells. ***$P < 0.00001$; two-tailed *z*-test, $z = 16.166$; $n = 2$ biologically independent experiments. **e**, Gene-set enrichment analysis showing the expression of genes directly upregulated by Mesp1 that were up- or downregulated in *Zic2/3*-dKO cells 24 h after Mesp1 induction. The nominal *P* values represent the statistical significance of the enrichment score relative to a null distribution calculated through a permutation test; NES, normalized enrichment score. **f**, Examples of genes downregulated in *Zic2/3*-dKO cells. Data are the fold change relative to no dox treatment; $n = 2$ (WT no dox and WT + dox) and 5 (*Zic2/3*-dKO) biologically independent experiments. **g**, Heatmap of Mesp1 ChIP–seq and ATAC-seq peaks in WT and *Zic2/3*-dKO cells 24 h after Mesp1 induction; $n = 2$ (WT) and 3 (*Zic2/3*-dKO) biologically independent experiments. 1, 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 *Zic2/3*-dKO cells; and 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 following Mesp1 induction is reduced (black boxes). Genome tracks are shown on top. **i**, Schematic overview of the chimera experiment. **j**, Image of smRNA-FISH on a section of a chimeric embryo at E9.5 showing *tdTomato* and *Tnnt2* expression. The image is representative of three independent embryos. The empty arrowheads show *Tnnt2*$^-$ *Zic2/3*-dKO *tdTomato*$^+$ cells corresponding to endocardial cells. The filled arrowheads show Tnnt2+ Zic2/3-dKO tdTomato+ cells. BA, branchial arch. A magnified view of (X3.4) the region in the white dotted box in the main image is provided (right). **k**, Percentage of Tomato$^+$ chimerism in the whole embryo (total) as well as cardiomyocytes and non-cardiomyocytes, as measured by the ratio between the area of *tdTomato*$^+$ pixels and the area of 4,6-diamidino-2-phenylindole (DAPI)-stained nuclei. Data are the mean ± s.e.m. Two-tailed paired Student's *t*-test; NS, not significant; $n = 10$, 12 and 13 slides examined from three biologically independent embryos.


a
P = 4.9 × 10−7
P = 9.5 × 10−9
Flk1+Pdgfra+ cells (%)
WT
WT + dox
Zic2/3-dKO
Zic2/3-dKO + dox
b
P = 3.7 × 10−11
P = 5.7 × 10−8
Troponin+ cells (%)
c
Mesp1 WT
Zic2-KO
Zic3-KO
Zic2/3 dKO
24 h no dox
24 h dox
Expression (Z-score)
d
Down in Zic2/3-dKO (n = 2,374)
Up in Zic2/3-dKO (n = 2,046)
Proportion of genes affected by Zic2/3-dKO
***
310
29
676
65
237
330
Mesp1 direct upregulated (n = 588)
Mesp1 upregulated (n = 1,368)
Mesp1 downregulated (n = 1,448)
e
Mesp1 direct upregulated targets
Enrichment score
(n = 588)
NES = 2.95 (P < 0.001)
Downregulated in Zic2/3-dKO
Upregulated in Zic2/3-dKO
f
Mesp1 no dox
Mesp1 dox
Mesp1 and Zic2/3-dKO dox
mRNA expression
Kdr
Pdgfra
Foxf1
Gfra2
Aplnr
Tbx6
Mef2c
Gata4
Meis1
Hnad1
Twist1
Snai1
Prrx1
Isl1
Foxc1
Foxc2
Hoxb1
Hoxb2
Hand2
Wnt2
Wnt2b
Etv2
Nfatc1
Tbx18
g
Mesp1 ChIP–seq
ATAC-seq
WT
Zic2/3-dKO
24 h no dox
24 h dox
Z-score
h
10 kb
Snai1
Notch1
Mesp1 ChIP– seq
ATAC-seq
Zic3 ChIP–seq
0 h
i
WT
tdTomato+ Zic2/3-dKO
E9.5
j
Head
BA
Heart
Tnnt2
tdTomato (Zic2/3-dKO cells)
200 μm
k
P = 0.01
NS
(Percentage of tdTomato+ area/ the area of embryo
Total
Non-CMs
CMs
Heart

of cardiovascular lineages during embryonic development. Using Mesp1 ChIP–seq, H3K27Ac and H3K4me1 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]. Apart from a few exceptions[59–63], the cofactors that are required to regulate TF pioneer activity remain unknown.

Using bioinformatic predictions and functional validation by in vitro and in vivo loss-of-function experiments, we identified Zic2 and Zic3 as essential transcriptional cofactors that regulate Mesp1 TF activity and the expression of 50% of its direct cardiovascular target genes. Although CRISPR–Cas9n deletion of *ZIC2* in human PSCs leads to defective cardiomyocyte 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%), including 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 left/right asymmetry defects in *Zic3*-KO embryos remains unclear, as deletion of *Zic3* before gastrulation (but not in CPs or in the node) causes 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, which suggests 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 dKO cell lines demonstrated that Mesp1 overexpression can no longer activate many of its key enhancers that regulate the expression of many essential cardiovascular genes in *Zic2*/*3*-dKO mutants. These data demonstrate that Zic2 and Zic3 act redundantly and are essential for cardiac-mesoderm formation by regulating the TF activity of Mesp1. Consistent with an early role of Zic2 and Zic3 in regulating Mesp1-mesoderm specification, our chimeric embryo experiments between WT and *Zic2*/*3*-dKO show that cardiomyocytes were decreased in *Zic2*/*3*-dKO cells.

Mechanistically, Zic2 and Zic3 physically interact with Mesp1. Zic2 and Zic3 bind a significant number of Mesp1-bound enhancers previous to Mesp1 expression. In the absence of Zic2 and Zic3, Mesp1 presents a decrease in its binding affinity at 87% of its binding sites, showing that Zic2 and Zic3 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 and Zic3, Mesp1 can no longer open the chromatin regions and upregulate gene expression, showing that Zic2 and Zic3 are important for regulation of the pioneer activity of the transcriptional complex induced by Mesp1. Zic2 and Zic3 are also recruited by Mesp1 expression at 60% of Zic and Mesp1 co-bound enhancers. Finally, at a minor proportion of Mesp1 binding sites, Zic2 and Zic3 do not control Mesp1 binding and chromatin opening but instead promote transcriptional activation, illustrating the different mechanisms by which Zic2 and Zic3 control the transcriptional activity of Mesp1. Our data uncover a novel role for Zic2 and Zic3 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.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41556-022-00947-3.

Received: 5 January 2021; Accepted: 25 May 2022;
Published online: 11 July 2022

References

1. Xin, M., Olson, E. N. & Bassel-Duby, R. Mending broken hearts: cardiac development as a basis for adult heart regeneration and repair. *Nat. Rev. Mol. Cell Biol.* **14**, 529–541 (2013).
2. Harvey, R. P. Patterning the vertebrate heart. *Nat. Rev. Genet.* **3**, 544–556 (2002).
3. Meilhac, S. M. & Buckingham, M. E. The deployment of cell lineages that form the mammalian heart. *Nat. Rev. Cardiol.* **15**, 705–724 (2018).
4. Protze, S. I., Lee, J. H. & Keller, G. M. Human pluripotent stem cell-derived cardiovascular cells: from developmental biology to therapeutic applications. *Cell Stem Cell* **25**, 311–327 (2019).
5. Hoffman, J. I. & Kaplan, S. The incidence of congenital heart disease. *J. Am. Coll. Cardiol.* **39**, 1890–1900 (2002).
6. Srivastava, D. Making or breaking the heart: from lineage determination to morphogenesis. *Cell* **126**, 1037–1048 (2006).
7. Bruneau, B. G. The developmental genetics of congenital heart disease. *Nature* **451**, 943–948 (2008).
8. Saga, Y. et al. MesP1: a novel basic helix-loop-helix protein expressed in the nascent mesodermal cells during mouse gastrulation. *Development* **122**, 2769–2778 (1996).
9. Saga, Y. et al. MesP1 is expressed in the heart precursor cells and required for the formation of a single heart tube. *Development* **126**, 3437–3447 (1999).
10. Kitajima, S., Takagi, A., Inoue, T. & Saga, Y. MesP1 and MesP2 are essential for the development of cardiac mesoderm. *Development* **127**, 3215–3226 (2000).
11. Lescroart, F. et al. Early lineage restriction in temporally distinct populations of Mesp1 progenitors during mammalian heart development. *Nat. Cell Biol.* **16**, 829–840 (2014).
12. Lescroart, F. et al. Defining the earliest step of cardiovascular lineage segregation by single-cell RNA-seq. *Science* **359**, 1177–1181 (2018).
13. Bondue, A. et al. Mesp1 acts as a master regulator of multipotent cardiovascular progenitor specification. *Cell Stem Cell* **3**, 69–84 (2008).
14. David, R. et al. MesP1 drives vertebrate cardiovascular differentiation through Dkk-1-mediated blockade of Wnt-signalling. *Nat. Cell Biol.* **10**, 338–345 (2008).
15. Lindsley, R. C. et al. Mesp1 coordinately regulates cardiovascular fate restriction and epithelial–mesenchymal transition in differentiating ESCs. *Cell Stem Cell* **3**, 55–68 (2008).
16. Chan, S. S. et al. Mesp1 patterns mesoderm into cardiac, hematopoietic, or skeletal myogenic progenitors in a context-dependent manner. *Cell Stem Cell* **12**, 587–601 (2013).
17. Devine, W. P., Wythe, J. D., George, M., Koshiba-Takeuchi, K. & Bruneau, B. G. Early patterning and specification of cardiac progenitors in gastrulating mesoderm. *eLife* **3**, e03848 (2014).
18. Zhang, Q. et al. Unveiling complexity and multipotentiality of early heart fields. *Circ. Res.* **129**, 474–487 (2021).
19. Ivanovitch, K. et al. Ventricular, atrial and outflow tract heart progenitors arise from spatially and molecularly distinct regions of the primitive streak. Preprint at *bioRxiv* https://doi.org/2020.2007.2012.198994 (2020).
20. Creyghton, M. P. et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl Acad. Sci. USA* **107**, 21931–21936 (2010).
21. Shlyueva, D., Stampfel, G. & Stark, A. Transcriptional enhancers: from properties to genome-wide predictions. *Nat. Rev. Genet.* **15**, 272–286 (2014).
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* **38**, 576–589 (2010).

23. Wapinski, O. L. et al. Hierarchical mechanisms for direct reprogramming of fibroblasts to neurons. *Cell* **155**, 621–635 (2013).
24. Soufi, A. et al. Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell* **161**, 555–568 (2015).
25. Paige, S. L. et al. A temporal chromatin signature in human embryonic stem cells identifies regulators of cardiac development. *Cell* **151**, 221–232 (2012).
26. Wamstad, J. A. et al. Dynamic and coordinated epigenetic regulation of developmental transitions in the cardiac lineage. *Cell* **151**, 206–220 (2012).
27. Pikkarainen, S., Tokola, H., Kerkela, R. & Ruskoaho, H. GATA transcription factors in the developing and adult heart. *Cardiovasc. Res.* **63**, 196–207 (2004).
28. George, R. M. & Firulli, A. B. Hand factors in cardiac development. *Anat. Rec.* **302**, 101–107 (2019).
29. Greulich, F., Rudat, C. & Kispert, A. Mechanisms of T-box gene function in the developing heart. *Cardiovasc. Res.* **91**, 212–222 (2011).
30. Kume, T. Novel insights into the differential functions of Notch ligands in vascular formation. *J. Angiogenes. Res.* **1**, 8 (2009).
31. Oda, M. et al. DNA methylation restricts lineage-specific functions of transcription factor Gata4 during embryonic stem cell differentiation. *PLoS Genet.* **9**, e1003574 (2013).
32. Whyte, W. A. et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).
33. Bondue, A. et al. Defining the earliest step of cardiovascular progenitor specification during embryonic stem cell differentiation. *J. Cell Biol.* **192**, 751–765 (2011).
34. Kattman, S. J. 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* **8**, 228–240 (2011).
35. Ying, Q. L. et al. The ground state of embryonic stem cell self-renewal. *Nature* **453**, 519–523 (2008).
36. Gebbia, M. et al. X-linked situs abnormalities result from mutations in ZIC3. *Nat. Genet.* **17**, 305–308 (1997).
37. Ware, S. M. et al. Identification and functional analysis of ZIC3 mutations in heterotaxy and related congenital heart defects. *Am. J. Hum. Genet.* **74**, 93–105 (2004).
38. Cowan, J., Tariq, M. & Ware, S. M. Genetic and functional analyses of ZIC3 variants in congenital heart disease. *Hum. Mutat.* **35**, 66–75 (2014).
39. Purandare, S. M. et al. A complex syndrome of left-right axis, central nervous system and axial skeleton defects in Zic3 mutant mice. *Development* **129**, 2293–2302 (2002).
40. Ware, S. M., Harutyunyan, K. G. & Belmont, J. W. Zic3 is critical for early embryonic patterning during gastrulation. *Dev. Dyn.* **235**, 776–785 (2006).
41. Zhu, L. et al. Identification of a novel role of ZIC3 in regulating cardiac development. *Hum. Mol. Genet.* **16**, 1649–1660 (2007).
42. Elms, P. et al. Overlapping and distinct expression domains of Zic2 and Zic3 during mouse gastrulation. *Gene Expr. Patterns* **4**, 505–511 (2004).
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* **38**, 741–755 (2020).
44. Lim, L. S., Hong, F. H., Kunarso, G. & Stanton, L. W. The pluripotency regulator Zic3 is a direct activator of the Nanog promoter in ESCs. *Stem Cells* **28**, 1961–1969 (2010).
45. Lim, L. S. et al. Zic3 is required for maintenance of pluripotency in embryonic stem cells. *Mol. Biol. Cell* **18**, 1348–1358 (2007).
46. Perl, E. & Waxman, J. S. Retinoic acid signaling and heart development. *Subcell. Biochem.* **95**, 119–149 (2020).
47. Bertrand, N. et al. Hox genes define distinct progenitor sub-domains within the second heart field. *Dev. Biol.* **353**, 266–274 (2011).
48. Ryckebusch, L. et al. Retinoic acid deficiency alters second heart field formation. *Proc. Natl Acad. Sci. USA* **105**, 2913–2918 (2008).
49. Deshwar, A. R., Chng, S. C., Ho, L., Reversade, B. & Scott, I. C. The Apelin receptor enhances Nodal/TGFbeta signaling to ensure proper cardiac development. *eLife* **5**, e13758 (2016).
50. Scott, I. C. et al. The G protein-coupled receptor agtrl1b regulates early development of myocardial progenitors. *Dev. Cell* **12**, 403–413 (2007).
51. Zeng, X. X., Wilm, T. P., Sepich, D. S. & Solnica-Krezel, L. Apelin and its receptor control heart field formation during zebrafish gastrulation. *Dev. Cell* **12**, 391–402 (2007).
52. Klootwijk, R. et al. A deletion encompassing Zic3 in bent tail, a mouse model for X-linked neural tube defects. *Hum. Mol. Genet.* **9**, 1615–1622 (2000).
53. Pijuan-Sala, B. et al. A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* **566**, 490–495 (2019).
54. Guibentif, C. et al. Diverse routes toward early somites in the mouse embryo. *Dev. Cell* **56**, 141–153 e146 (2021).
55. Zaret, K. S. Pioneer transcription factors initiating gene network changes. *Annu. Rev. Genet.* **54**, 367–385 (2020).
56. Zaret, K. S. Pioneering the chromatin landscape. *Nat. Genet.* **50**, 167–169 (2018).
57. Zaret, K. S. & Carroll, J. S. Pioneer transcription factors: establishing competence for gene expression. *Genes Dev.* **25**, 2227–2241 (2011).
58. Iwafuchi-Doi, M. & Zaret, K. S. Cell fate control by pioneer transcription factors. *Development* **143**, 1833–1837 (2016).
59. Liu, Z. & Kraus, W. L. Catalytic-Independent functions of PARP-1 determine Sox2 pioneer activity at intractable genomic loci. *Mol. Cell* **65**, 589–603 (2017).
60. Swinstead, E. E. et al. Steroid receptors reprogram FoxA1 occupancy through dynamic chromatin transitions. *Cell* **165**, 593–605 (2016).
61. Donaghey, J. et al. Genetic determinants and epigenetic effects of pioneer-factor occupancy. *Nat. Genet.* **50**, 250–258 (2018).
62. Mayran, A. et al. Pioneer and nonpioneer factor cooperation drives lineage specific chromatin opening. *Nat. Commun.* **10**, 3807 (2019).
63. Cernilogar, F. M. et al. Pre-marked chromatin and transcription factor co-binding shape the pioneering activity of Foxa2. *Nucleic Acids Res.* **47**, 9069–9086 (2019).
64. Sutherland, M. J., Wang, S., Quinn, M. E., Haaning, A. & Ware, S. M. Zic3 is required in the migrating primitive streak for node morphogenesis and left-right patterning. *Hum. Mol. Genet.* **22**, 1913–1923 (2013).
65. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
66. Tarbell, E. D. & Liu, T. HMMRATAC: a hidden Markov ModeleR for ATAC-seq. *Nucleic Acids Res.* **47**, e91 (2019).

Methods

Tetracycline-inducible PSC lines. The Mesp1–3HA stable transgenic PSC line was generated as described previously[67].

CRISPR–Cas9n-induced KO in PSCs. *Zic2*-KO, *Zic3*-KO and *Zic2/3*-dKO PSCs were generated using CRISPR–Cas9n following a previously described protocol[67,68]. Briefly, two pairs of guide RNAs per gene, targeting two regions of the gene separated from 364 bp to 2381 bp, were designed using a CRISPR–Cas9n online tool (Benchling Biology Software (2019–2020), https://benchling.com; Supplementary Table 1). The vectors pX330-U6-Chimeric_BB-CBh-hSpCas9 (PX330) and pSpCas9n (BB)-2AGFP (PX461) were obtained from Addgene (cat. nos 42230 and 48140). The guide RNAs were cloned into a modified short version of PX330 from which the Cas9 cassette was removed. Four plasmids containing the guide RNAs for single KO or eight plasmids for dKO were co-transfected with PX461 (Cas9n) into Mesp1–3HA PSCs using Lipofectamine 2000 (ThermoFisher, 11661089) to generate a big deletion either upstream or encompassing the DNA-binding domain of the targeted gene(s). Transfected GFP$^+$ cells were individually FACS sorted into 96-well plates 48 h after transfection using a FACSAria system (BD Biosciences). After 7–12 d, the colonies were passaged and then screened by PCR using primers flanking the expected deletion site (Supplementary Table 2). *Zic2* and *Zic3* single-KO and dKO clones were selected based on homozygous PCR profiles and confirmed by Sanger sequencing.

Enhancer KO in PSCs using CRISPR–Cas9n. The same strategy described in the previous section was used to generate enhancer-KO cell lines. Eight guide RNAs were inserted into the shortened version of PX330 to generate deletions of between 200 and 480 bp 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 WT Cas9 (Supplementary Table 1). Single-stranded DNA containing 40-bp homology arms on each side of the motif—which was modified to CGCTAGC, a NheI restriction site—was synthetized by Eurogentec. Mesp1-inducible PSCs were transfected with PX330 containing the guide RNA and the single-stranded DNA template, and then FACS sorted into single cells based on GFP expression as described earlier. 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]. Following the generation of KO cell lines by CRISPR–Cas9n, the PSC cell lines were cultured in feeder-free conditions 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. The RT–qPCR was performed as described previously[12,13]. The qPCR primers are listed in Supplementary Table 3.

RNA-seq. RNA extraction was performed using a Qiagen RNeasy micro kit according to the manufacturer's instructions. Before sequencing, the quality of the RNA was evaluated using a Bioanalyzer 2100 system (Agilent). Indexed complementary DNA 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 system (Illumina).

ChIP–seq. Chromatin immunoprecipitation was performed on Mesp1–3HA-flagged dox-inducible PSC lines as described previously[12]. Briefly, EBs were collected at 12 and 24 h after Mesp1 induction using dox (working concentration 1 µg ml^{-1}), fixed directly with 1% formaldehyde for 7 min at room temperature and quenched with 0.125 M glycine for 5 min. The ChIP was performed according to the manufacturer's instructions (ChIP-IT express kit) with antibodies (Supplementary Table 4). The pulled-down DNA (2–10 ng) was used to construct the sequencing library using a 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-seq. The ATAC-seq was performed following a standard protocol[69]. Briefly, EBs were collected at the same time points and conditions as for the RNA-seq datasets, rinsed with dPBS (ThermoFisher, 14190144) and dissociated with accutase (Sigma, A6964). Cells ($n = 125{,}000$) or FACS-sorted cells were used to perform the ATAC-seq; the ATAC-seq libraries were prepared according to manufacturer's instructions (Nextera DNA sample preparation kit, Illumina) and size selection (200–800 bp) was performed using Ampure XP beads (Beckman) before next-generation sequencing.

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

Western blotting. Western blots were performed as previously described[67]. The antibodies used are listed in Supplementary Table 4.

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

Immunofluorescence analysis. Immunofluorescence on EBs was performed as previously described[13]. The 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.

Standard histological procedures were used on E14.5 hearts[71]. The antibodies used are reported in Supplementary Table 4. Images were acquired using an Axio Zoom.V16 microscope (Zeiss) and the thickness of the cardiomyocyte layer was measured using the Fiji software. Phenotypic analyses of *Zic3*-KO mutant and WT hearts at E14.5 were performed with careful stage matching between controls and the KO embryos (based on the shape of the limbs).

Mouse lines. *Mesp1–cre* mice were previously obtained from Y. Saga[9]. The mice colonies were maintained in certified animal facilities in accordance with European guidelines (7–19 h light cycle, 20–25 °C and 55 ± 15% humidity). The experiments were approved by the local ethical committee (CEBEA) under protocols nos 591N (C.B.) and Apafis 13031 (F.L.), and a national agreement (B1301308; F.L.). 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). The offspring were screened for genomic aberrations in 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 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 an Axio Zoom.V16 microscope (Zeiss). The areas of the right and left ventricles were measured using the Fiji software and a ratio was then calculated.

smRNA-FISH experiments. The smRNA-FISH was performed according to the protocol of the RNAscope multiplex fluorescent v2 assay (ACD-bio. 323110). Briefly, embryos were fixed for 26–30 h 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 16 h in butanol at 4 °C. For smRNA-FISH on sections of E9.5 chimera, the embryos were embedded in histogel (Epredia, HG-4000-012) before paraffin embedding. Tissue sections were cut at a thickness of 7–10 µm. Whole-mount smRNA-FISH was performed as previously described[72]. Probes are reported in Supplementary Table 5. To develop probes, we used Opal dyes from Akoya Bioscience (Opal-520, Opal-570 and Opal-650; 1:100–1:200). The embryos or sections were imaged using an Axio Zoom.V16 (Zeiss) or LSM800 confocal (Zeiss) microscope. When analysing litters from *Mesp1–cre* or *Zic3*-KO lines, the embryos were genotyped by PCR after imaging.

Image analyses after smRNA-FISH experiments. To analyse the expression of *Mesp1* in EBs, images were analysed using 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 $\frac{\Sigma \text{ integrate intensity of selected background regions}}{\Sigma \text{ area of selected background regions}}$. To measure the signal intensity (SI), cells where *Mesp1* expression was detected were selected (region of interest, ROI) and the following formula was used: 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 the *Mesp1*$^+$ pixels and 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 earlier. We then selected pixels with SI > ABI. All positive pixels were selected and their SA was 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 of the

SA of *tdTomato*$^+$ pixels to 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-cardiomyocyte 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*-KO was performed using exactly the same strategy with the same targeting guide RNAs used for *Zic3* KO in Mesp1–3HA inducible PSCs (Supplementary Table 1). Next, *Zic2* KO was performed on this *Zic3*-KO tdTomato-expressing PSC line. Three *Zic2/3*-dKO clones were selected and their genotype was confirmed using PCR and sequencing. These PSCs were injected into WT blastocysts in six 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 to eliminate sequencing adaptors[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]. The counts were normalized for all samples as reads per million (RPM) of mapped reads for each gene. Only genes that had an expression of >1 RPM for both duplicates of one or more conditions were kept for further analysis.

Differentially expressed genes were defined with DESeq2 (v1.34) using a cutoff of $P_{adj}=0.05$ and a minimum of 1.5-fold change in expression between both conditions[65]. The dox and no dox conditions were compared separately at 12 and 24 h; these two lists of differentially expressed genes were then pooled to define their kinetics of expression. Genes with a s.d. higher than 50% of the mean expression were further eliminated. Genes were first classified as up- or downregulated using a threshold of 1.5-fold change ($\frac{\text{mean dox}}{\text{mean no dox}} > 1.5$ or $< 1/1.5$, respectively). To classify upregulated genes into early, constant and late activated, we calculated the slope of increased expression from 0 to 12 h (a) and from 12 to 24 h (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 directions, the slope with the 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 (ref. [76]). When comparing Mesp1-inducible WT cells with *Zic2*-KO, *Zic3*-KO or *Zic2/3*-dKO cells, we used DESeq2 to call differentially expressed genes between cell types and added a cutoff fold-change value of two[65]. For comparing Flk1$^-$Pdgfra$^-$ and 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 medium, taking into account only genes expressed at least 1 RPM.

Analysis of microarray data in Mesp1-expressing cells in vivo. Genes upregulated in GFP (Mesp1$^+$)-expressing cells in vivo were defined by microarray as genes presenting a signal change at least twofold stronger in E6.75 or E7.25 GFP$^+$ cells compared with GFP$^-$ cells. We then classified these genes as early, constant and late depending on the ratio of expression between the 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 were classified as late. All intermediate genes were classified as constant.

Analysis of Mesp1 and H3K27Ac ChIP–seq. Raw sequenced reads were trimmed using Trimmomatic for quality of bases and to eliminate sequencing adaptors[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), and 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 h were compared between both replicates and only the intersection of both were kept for further analyses. Next, 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 the average number of reads at 12 h ($n_{reads,\ 12h}$) / $n_{reads,\ 24h} > 2$, late if this ratio was <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 transcription start site. Ngsplot (v2.63) was used for data visualization[81].

Analysis of H3K4me1, Mesp1 ChIP–seq in 2i conditions and *Zic2/3*-dKO cells. FastQ files were cleaned and trimmed using Trimmomatic[73]. Paired-end reads were aligned using Bowtie2 (ref. [77]). Picard tools were used to remove PCR duplicates (Broad Institute), and samtools to remove mitochondrial reads and low-quality alignments[78]. Peaks were called using MACS2 (ref. [79]) with a threshold $P=10^{-10}$. Where duplicates were available (all of 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 with 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 (ref. [77]). Picard tools were used to remove PCR duplicates (Broad Institute), and samtools to remove mitochondrial reads and low-quality alignments[78]. Peaks were called using MACS2 (ref. [79]) with a threshold $P=10^{-10}$. 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 using DESeq2, using $P_{adj}=0.05$ as the threshold[65]. Peaks were annotated with the same parameters as described earlier 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 cutoff 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 and 0 h of dox treatment using a q-value cutoff of 0.05. De novo peaks were defined as Mesp1 ChIP–seq peaks not presenting a peak at 0 h but presenting one at 24 h. Primed peaks encompassed peaks that presented an ATAC-seq peak at 0 h and undefined Mesp1 peaks, which did not present a peak at either 0 or 24 h 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; up- and downregulated peaks were then defined using DESeq2 with a threshold of $P_{adj}<0.05$ and minimum fold change = 2. For the Flk1$^-$Pdgfra$^-$ and Flk1$^+$Pdgfra$^+$ samples, DESeq2 was used with a cutoff at $P=0.05$ and a minimum fold-change average read enrichment of two.

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 24 h of dox treatment. The number of occurrences of each possible bHLH motif were quantified within these footprints and represented as a bar plot.

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 $q=0.05$, given that 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 described earlier.

Statistics and reproducibility. For PSC work, the results shown and statistical tests are from at least three biologically independent experiments performed in at least one cell line for WT cells and two cell lines for the CRISPR–Cas9n-induced KO cells. Analysis and statistical testing of next-generation sequencing results were performed in two biological independent experiments, unless stated otherwise 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 the data structure (categorical versus quantitative and normal versus non-normal distribution).

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

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

Data availability

The next-generation-sequencing 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 next-generation-sequencing data and microarray data that were re-analysed here are available under the accession codes GSE41361, GSE44288 and GSE59033. Source data are provided with this paper. All other data supporting the findings of this study are available from the corresponding authors on reasonable request.

References

67. Chiapparo, G. et al. Mesp1 controls the speed, polarity, and directionality of cardiovascular progenitor migration. *J. Cell Biol.* **213**, 463–477 (2016).

68. Ran, F. A. et al. Double nicking by RNA-guided CRISPR-Cas9n for enhanced genome editing specificity. *Cell* **154**, 1380–1389 (2013).
69. Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* **109**, 21.29.1–21.29.9 (2015).
70. Sambrook, J. & Russell, D. W. Identification of associated proteins by coimmunoprecipitation. *CSH Protoc.* **2006**, pdb.prot3898 (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.* **406**, 247–258 (2015).
72. de Soysa, T. Y. et al. Single-cell analysis of cardiogenesis reveals basis for organ-level developmental defects. *Nature* **572**, 120–124 (2019).
73. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
74. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
75. Anders, S., Pyl, P. T. & Huber, W. HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
76. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis* 2nd edn (Springer Publishing Company, Incorporated, 2009).
77. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
78. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
79. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
80. McLean, C. Y. et al. GREAT improves functional interpretation of *cis*-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).
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* **15**, 284 (2014).
82. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
83. Li, Z. et al. Identification of transcription factor binding sites using ATAC-seq. *Genome Biol.* **20**, 45 (2019).

Acknowledgements

We thank the ULB animal facility and ULB genomic core facility (F. Libert and A. Lefort). We thank Y. Song for bioinformatic assistance. F.L., X.L. and B.S. were supported by an FNRS aspirant fellowship (B.S.), an EMBO long-term fellowship and the Leducq Foundation. S.G. was funded by a Royal Society Newton International Fellowship (grant no. NIF\R1\181950). C.G. is funded by the Swedish Research Council (grant no. 2017-06278). F.L. thanks the MMG imaging and the animal phenotyping core platforms. We thank R. Kelly for sharing antibodies. Work in the Göttgens laboratory is supported by grants from the Wellcome Trust, Blood Cancer UK, Cancer Research UK, NIDDK and core support grants from the Wellcome Trust to the Wellcome–MRC Cambridge Stem Cell Institute. Work in F.L.'s laboratory was supported by the INSERM ATIP-Avenir programme. C.B. is an investigator of WELBIO. Work in C.B.'s laboratory was supported by the FNRS, ULB foundation, European Research Council (ERC), foundation Bettencourt Schueller (C.B. and F.L.) and Leducq Foundation as part of the network '22q11.2 deletion syndrome: novel approaches to understand cardiopharyngeal pathogenesis'. C.B. and B.G. acknowledge support from the Fondation Philippe Wiener–Maurice Anspach.

Author contributions

X.L., B.S., F.L. and C.B. designed the experiments, performed data analyses and wrote the manuscript. X.L. and B.S. performed most of the biological experiments. B.S. performed bioinformatic analyses of all sequencing data. C.D. performed FACS. E.P. and F.L. performed the immunofluorescence and RNAscope experiments on mouse embryos. Y.A. generated the *Zic3*-KO mouse lines. B.S., S.Z., E.B. and F.L. described and analysed the *Zic3*-mutant phenotypes. Chimeric embryos were generated by C.G., M.-L.N.T. and W.M., and processed and analysed by F.L. and B.G. Knockout cell lines for the chimeric embryos were generated by B.S. and C.G. C.P. provided technical support. E.C. and F.F. provided help with some of the next-generation sequencing. S.G. and J.C.M. performed the scRNA-seq analyses for the initial submission. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

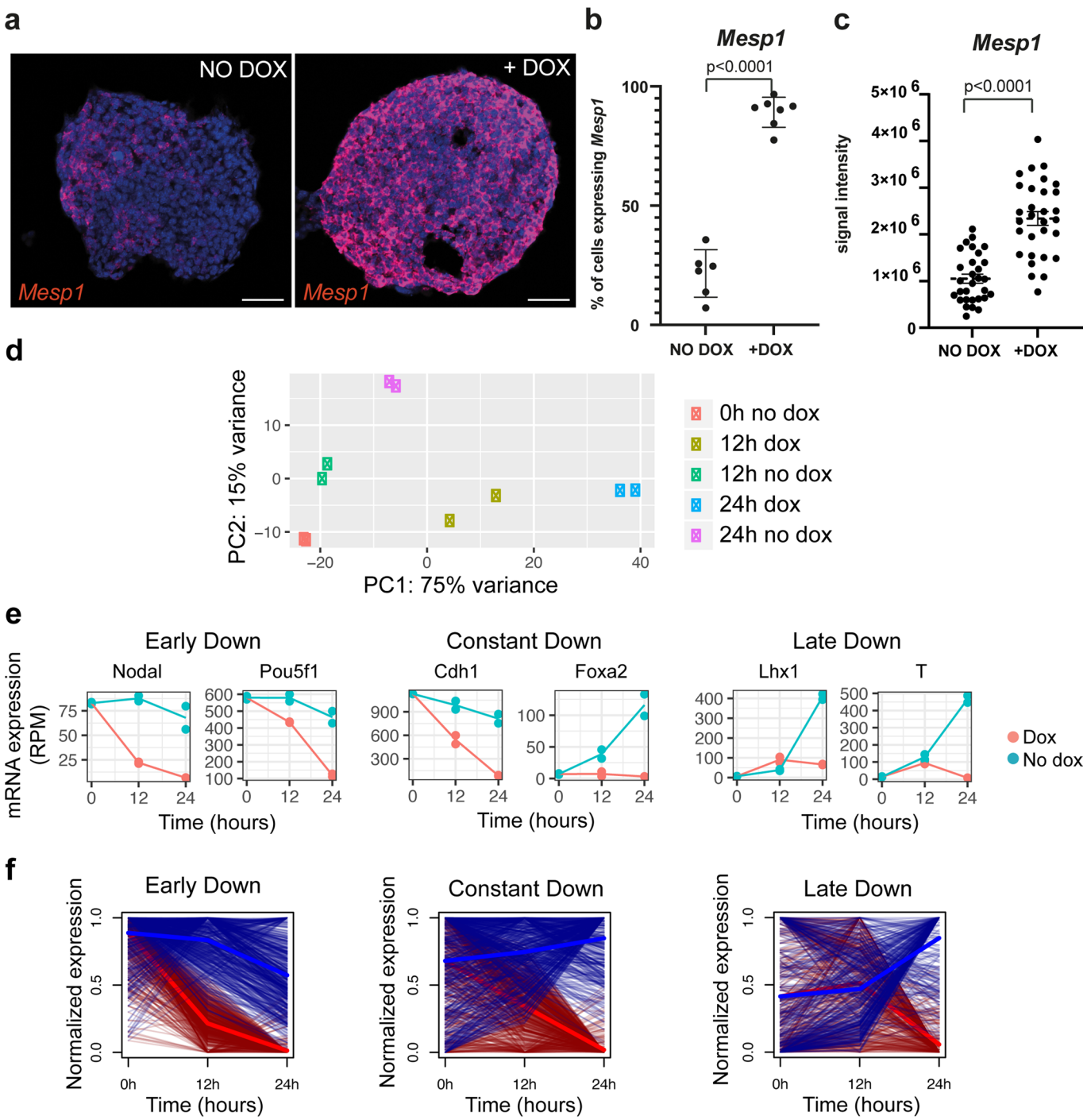
Extended data is available for this paper at https://doi.org/10.1038/s41556-022-00947-3.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41556-022-00947-3.

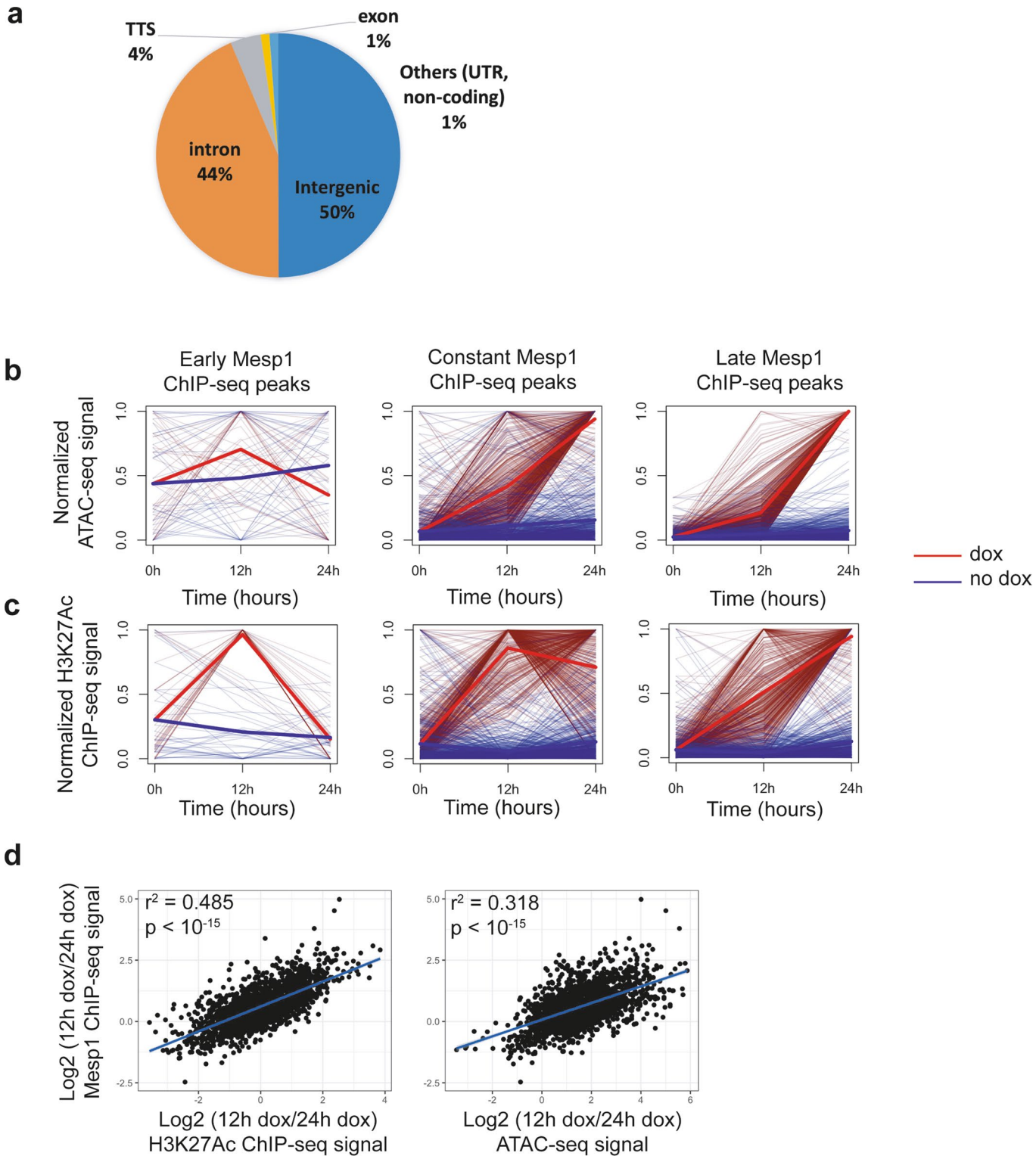
Correspondence and requests for materials should be addressed to Fabienne Lescroart or Cédric Blanpain.

Peer review information *Nature Cell Biology* thanks José Luis de la Pompa and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.



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 24 h 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 (0 h), day 3 (12 h following Mesp1 overexpression) and day 3.5 (24 h 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 TF 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 h and 24 h dox, in order to extract a measure of fold-change between 12 and 24 h. 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.

a

Early peaks (n=35)

Motif	p-value	% targets	% back-ground	Best match
	1e-11	60.0	11.3	bHLH (Mesp1)
	1e-9	68.6	19.7	bHLH
	1e-6	62.9	21.8	Sox
	1e-5	28.6	5.0	Oct4-Sox2-Tcf-Nanog
	1e-4	74.3	39.4	Zeb
	1e-2	31.4	11.3	Zic
	1e-2	31.4	12.3	Oct

Constant peaks (n=1451)

Motif	p-value	% targets	% back-ground	Best match
	1e-961	81.5	8.8	bHLH (Mesp1)
	1e-170	81.3	45.7	bHLH
	1e-54	55.3	35.2	Sox
	1e-53	24.5	10.3	Zic
	1e-38	11.9	3.8	Oct4-Sox2-Tcf-Nanog
	1e-28	28.2	16.4	Tcf/Lef

Late peaks (n=525)

Motif	p-value	% targets	% back-ground	Best match
	1e-297	73.1	7.9	bHLH (Mesp1)
	1e-54	80.2	47.2	bHLH
	1e-23	31.6	14.3	Zic
	1e-22	36.2	17.9	Gata
	1e-18	39.4	22.0	Fox
	1e-18	31.0	15.6	Lef/Tcf
	1e-14	51.2	34.4	Sox

b

De novo peaks (n = 1483)

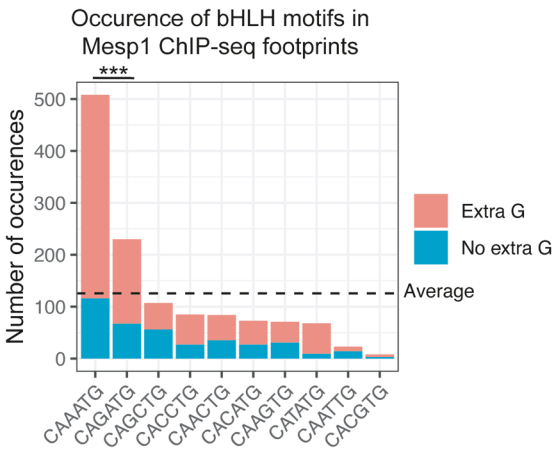
Motif	p-value	% targets	% back-ground	Best match
	1e-1042	89.7	11.6	bHLH (Mesp1)*
	1e-170	81.6	46.5	bHLH
	1e-46	54.6	36.2	Sox
	1e-33	25.6	13.7	Zic
	1e-30	28.9	16.8	Lef/Tcf
	1e-24	87.0	76.3	Nanog
	1e-18	24.3	15.5	Fox
	1e-17	28.7	19.2	Gata

Primed peaks (n = 429)

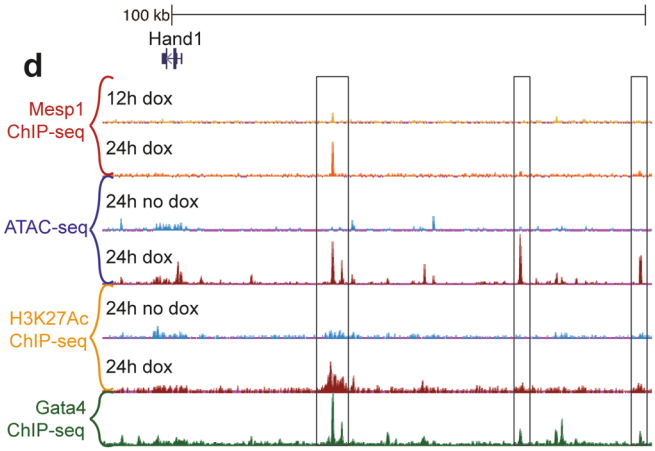
Motif	p-value	% targets	% back-ground	Best match
	1e-214	70.9	8.8	bHLH (Mesp1)**
	1e-46	21.2	3.1	Oct4-Sox2-Tcf-Nanog
	1e-46	45.5	15.8	Zic
	1e-39	50.1	21.1	bHLH
	1e-18	53.4	32.5	Sox
	1e-15	2à.3	8.0	Oct
	1e-13	30.3	15.5	Lef/Tcf

* > ** ($p<1^{-10}$)

c



d

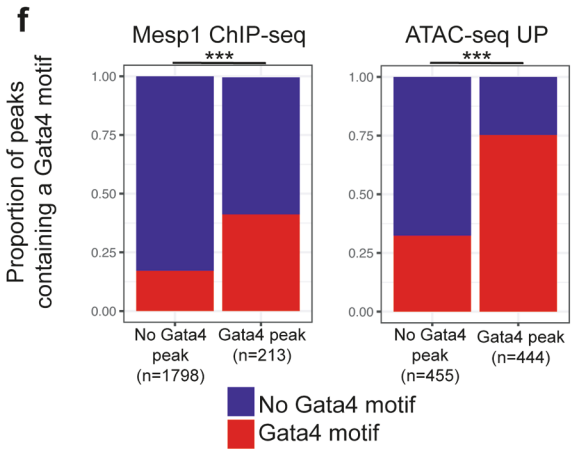


e

ATAC-seq peaks UP no Mesp1 (n = 899)

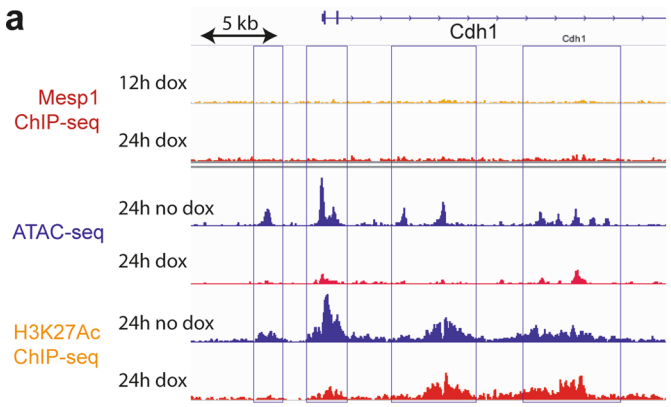
Motif	p-value	% targets	% back-ground	Best match
	1e-213	46.2	7.6	Gata
	1e-43	47.3	25.7	T-box
	1e-36	64.4	43.2	Tgif
	1e-34	21.7	8.3	Lef/Tcf
	1e-34	20.5	7.5	Hand
	1e-34	20.0	7.3	Fox
	1e-30	42.8	25.1	Meis
	1e-28	7.2	1.2	Hox

f



Extended Data Fig. 3 | See next page for caption.

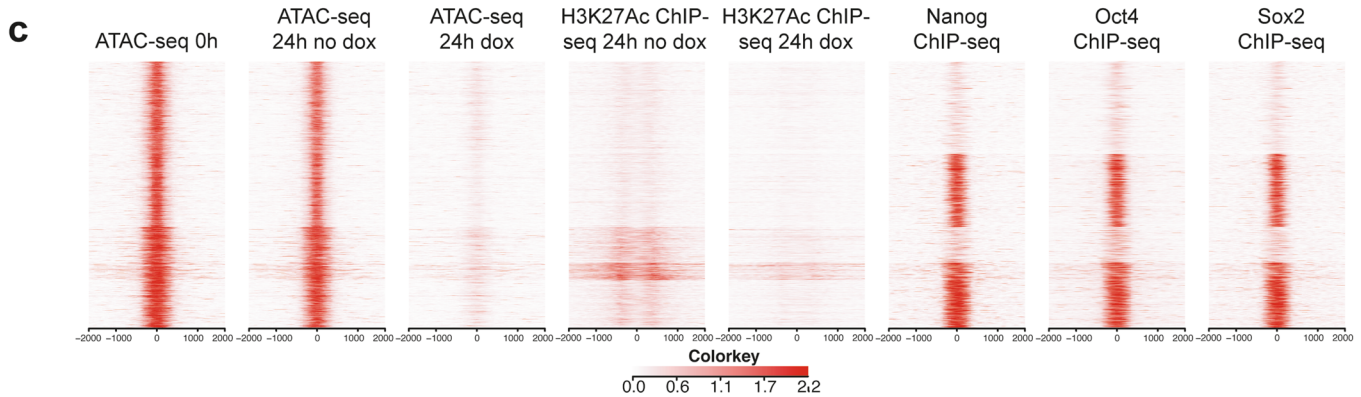
Extended Data Fig. 3 | Motif discovery of enhancers activated by Mesp1. a, Motif discovery performed separately in Mesp1 ChIP-seq peaks classified as early, constant or late peaks. p-values are calculated through a binomial test. **b**, Same pipeline as panel a but separating peaks as *de novo* versus 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 notably over-represented ($p < 0.00001$) 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 24 h 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 notably 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.



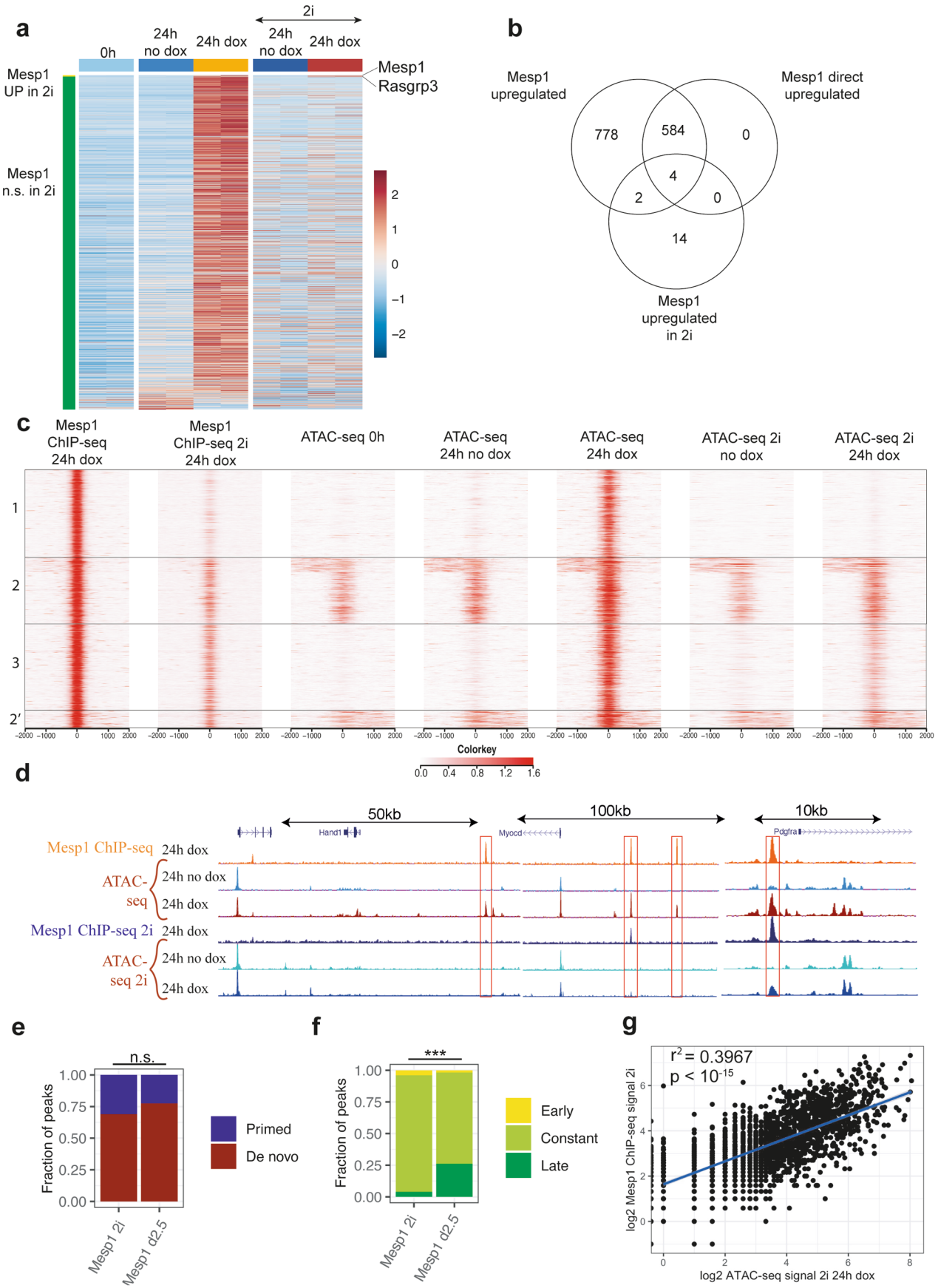
b

ATAC-seq peaks DOWN (n = 3247)

Motif	p-value	% targets	% back-ground	Best match
	1e-496	34.5	5.2	Oct4-Sox2-Tcf-Nanog
	1e-279	53.0	21.9	Sox
	1e-192	32.0	11.0	Zic
	1e-177	57.5	31.3	bHLH
	1e-157	31.1	11.9	Oct
	1e-92	25.1	11.2	NFkB-p65
	1e-51	35.3	22.6	Tead
	1e-48	29.0	17.6	Lef/Tcf
	1e-48	42.4	29.2	Lhx

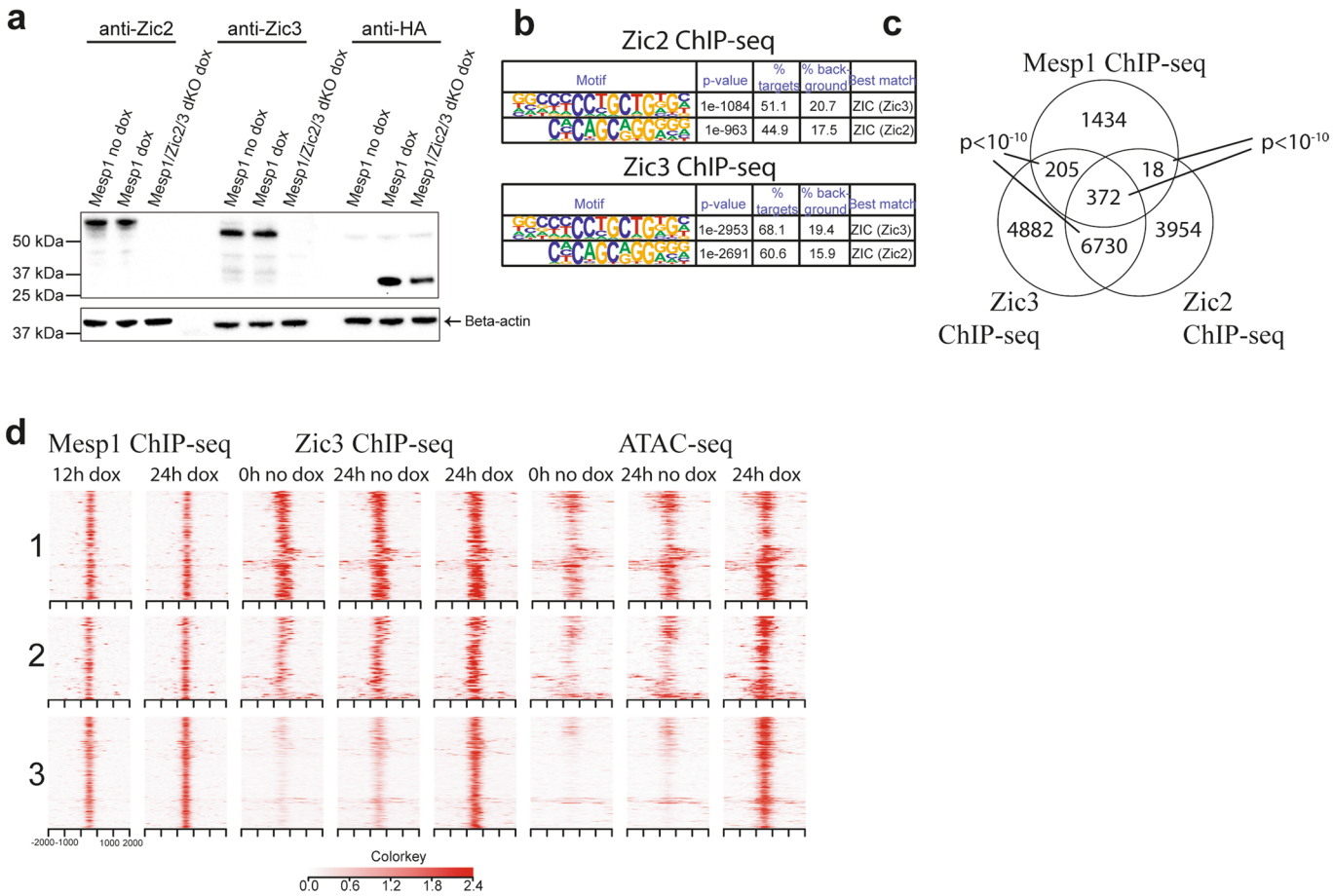


Extended Data Fig. 4 | Repression of the core pluripotency network by Mesp1. a, Representative genomic locus (*Cdh1*) where ATAC-seq and H3K27Ac ChIP peaks found in control (no dox) conditions are absent or smaller in dox conditions, without presenting any Mesp1 binding (blue boxes), suggesting indirect 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].

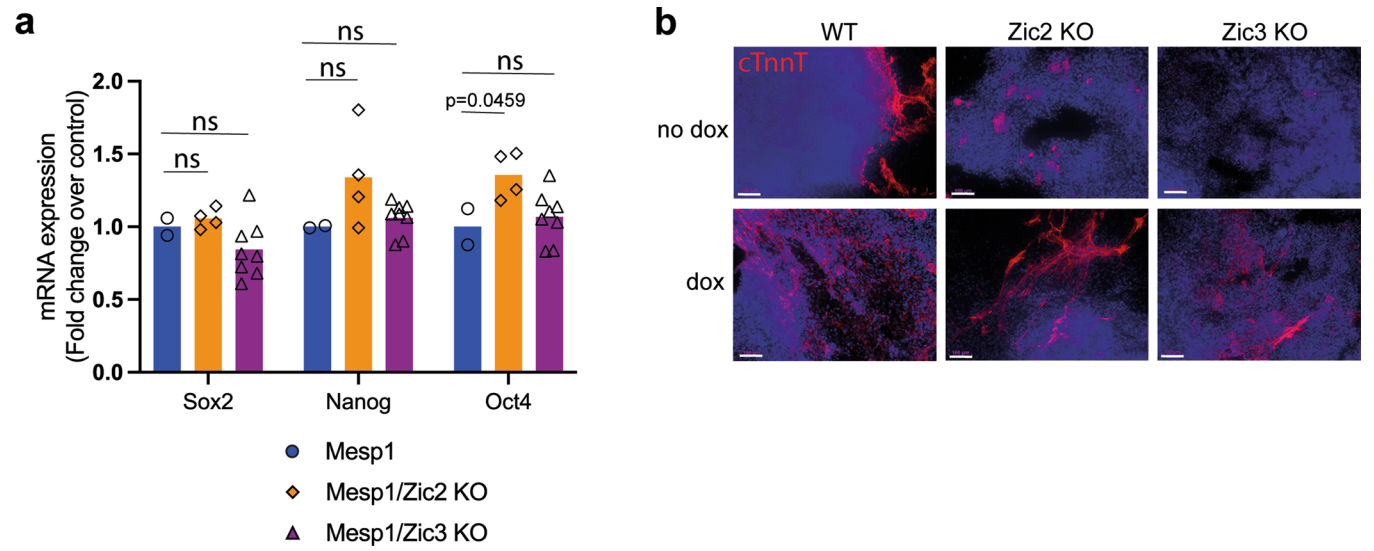


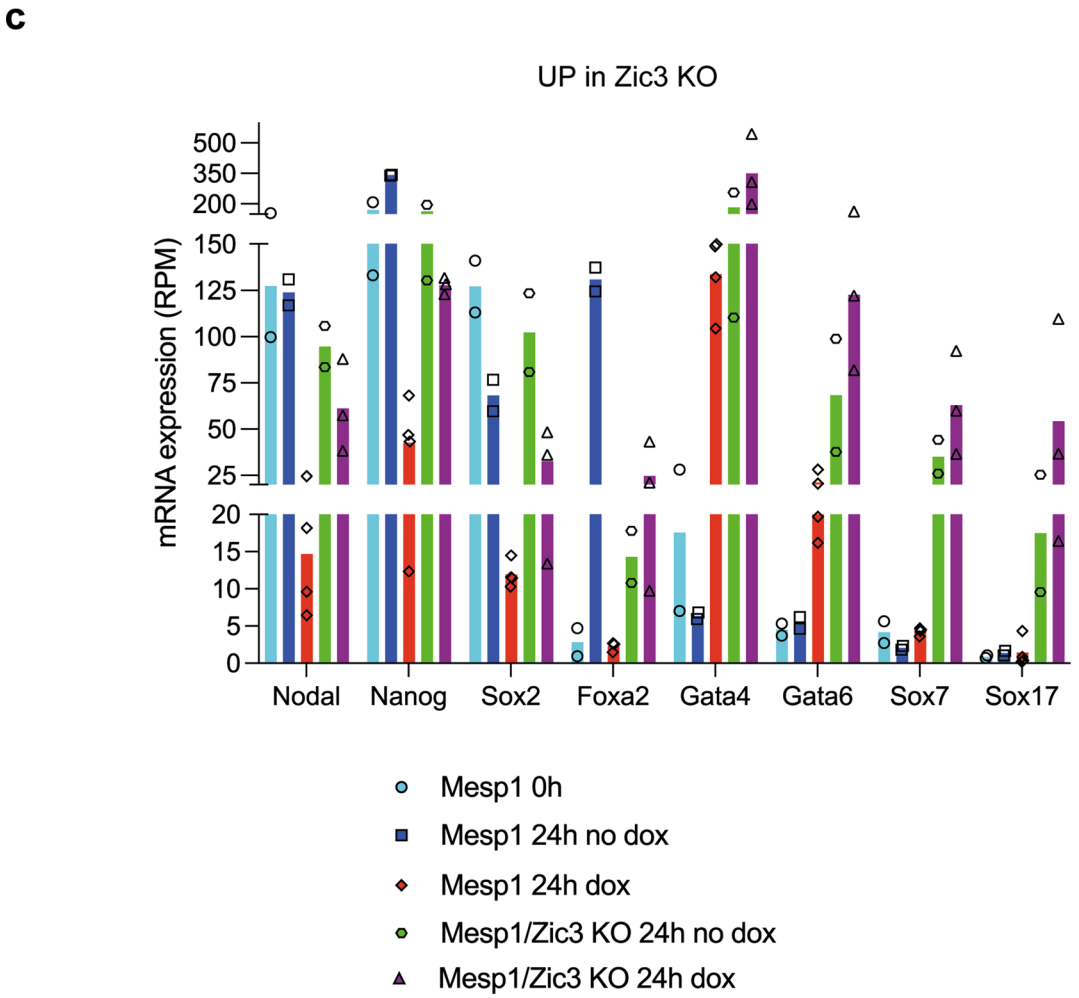
Extended Data Fig. 5 | See next page for caption.

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 24 h 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 notably 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 (24 h dox), demonstrating a linear correlation between these two variables. Data shown represent two biologically independent replicates.

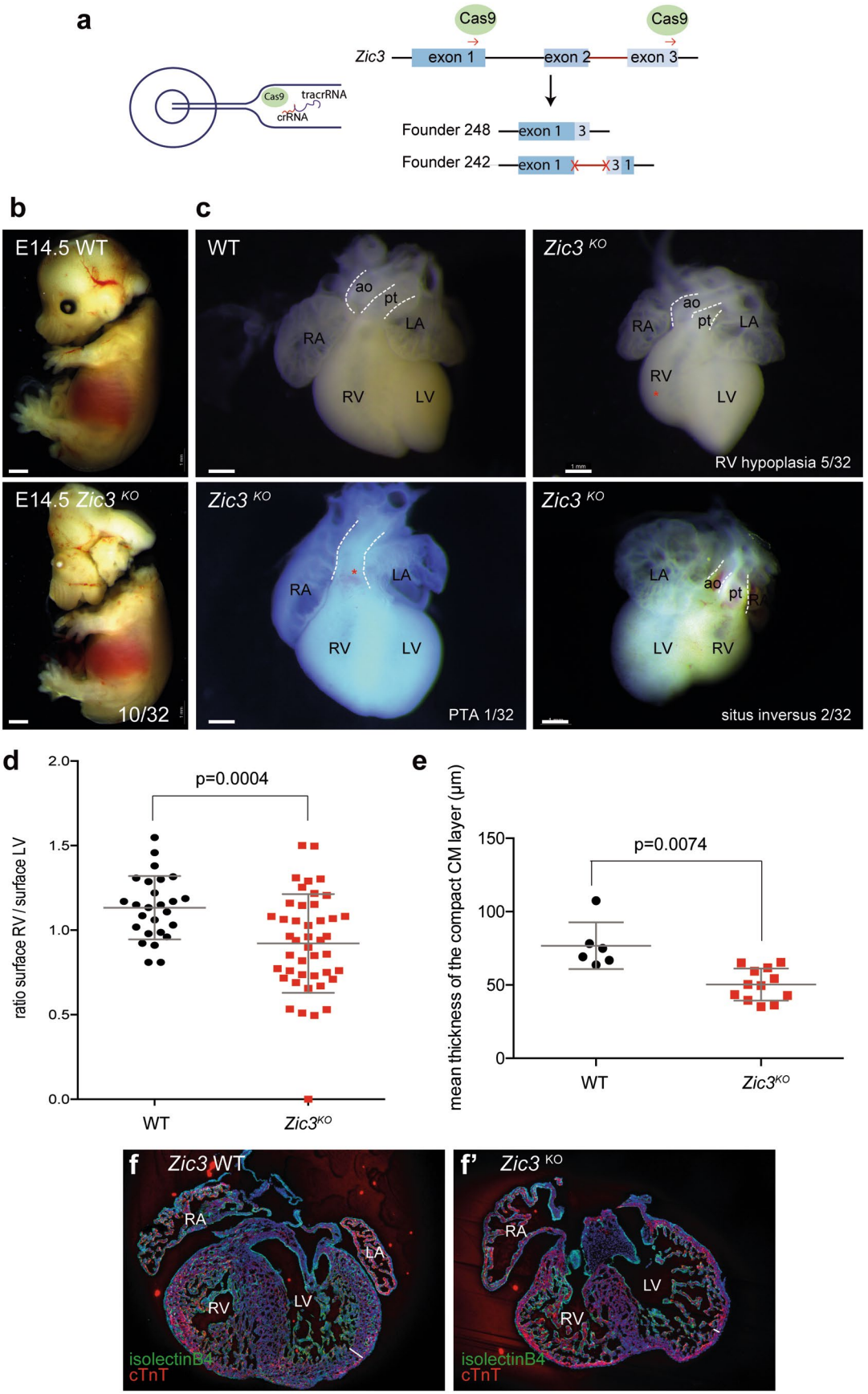


Extended Data Fig. 6 | Zic2 and Zic3 cooperate with Mesp1 and potentially with other mesoderm-inducing TFs. a, Western blot illustrating the expression of Zic2 and Zic3 with and without Mesp1 induction, at day 3.5 (24 h) of PSC differentiation, as well as the lack of Zic2 and Zic3 protein expression 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 h); 2, peaks bound by Zic3 at 24 h no dox; 3, peaks bound by Zic3 at 24 h dox.



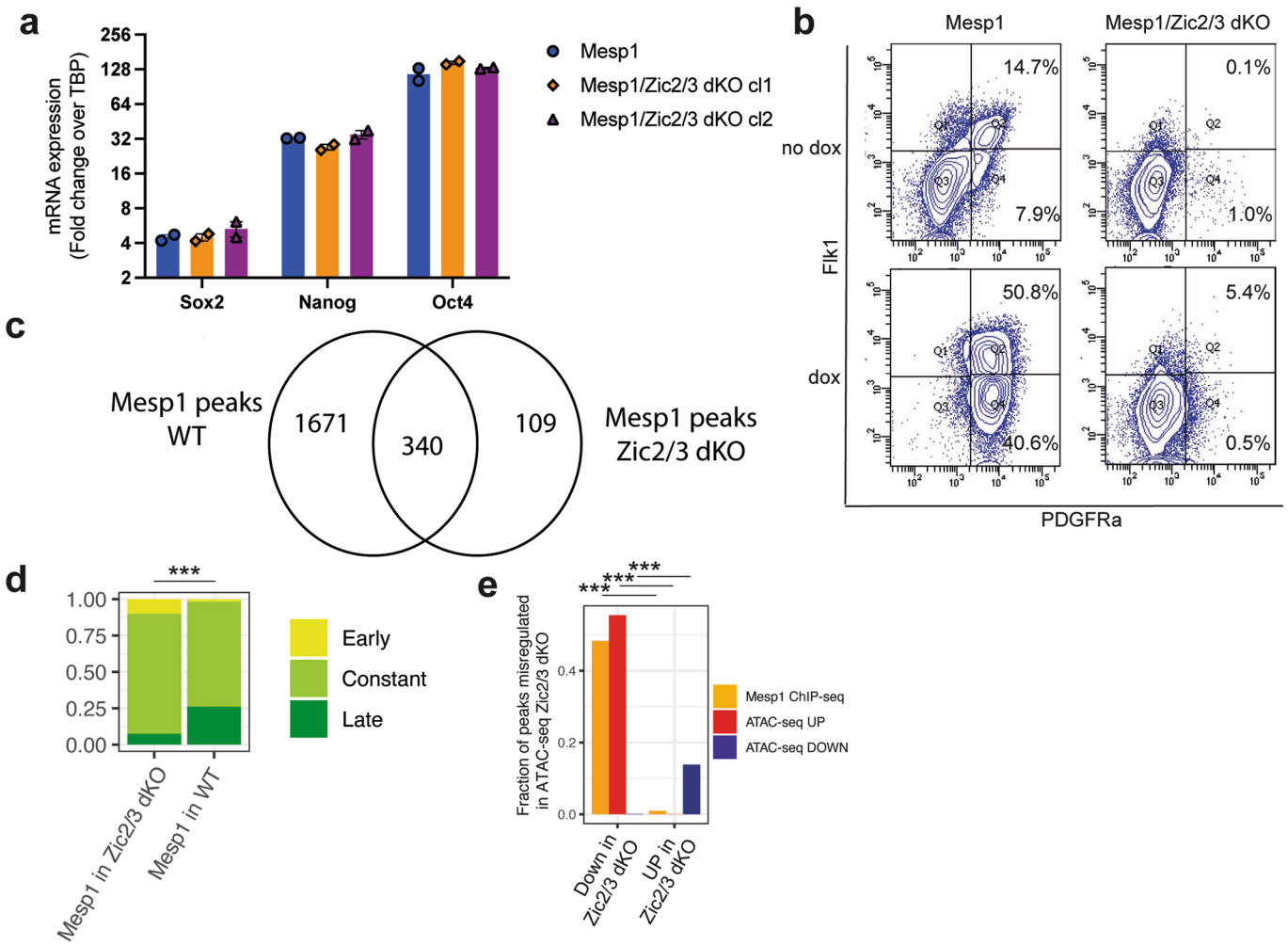


Extended Data Fig. 7 | Zic2 and Zic3 regulate Mesp1-induced CP specification and differentiation. a, mRNA expression of the core pluripotency associated TFs in Mesp1-inducible WT, *Zic2* and *Zic3* KO PSCs in Lif/2i pluripotency conditions, as measured by RT-qPCR. (n = 4 biologically independent replicates 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 notably upregulated in Zic3 KO cells in comparison to WT cells, including known important factors of pluripotency and endoderm differentiation.



Extended Data Fig. 8 | See next page for caption.

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= 1 mm. **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= 1 mm. **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, two-tailed 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= 200 µm. RV, right ventricle; LV, left ventricle; RA, right atrium; LA, left atrium; pt, pulmonary trunk; ao, aorta.



f

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

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

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 *Zic2/3* dKO 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.

nature research

Corresponding author(s): Fabienne Lescroart
Cédric Blanpain

Last updated by author(s): 2022/27/04

Reporting Summary

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

Statistics

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

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

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

Software and code

Policy information about availability of computer code

Data collection	No softawre was used for data collection.
Data analysis	All packages used are available online and the versions and programs are described in the Materials and methods section. For more details we have used Trimmomatic (v0.39), HTseq (v0.11.3), STAR (v2.7.3), DESeq2 (v1.34), Bowtie2 (v2.4.2), Samtools (v1.2.10), Macs2 (v2.4.2), GREAT (v4.0.4), ngs.plot (v2.63), HHMRATAC (v1.2.10), HINT-ATAC (v0.13.1), Homer (v4.11.1).

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

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

NGS Data has been deposited in Gene Expression Omnibus (GEO) and is accessible through GEO Series accession number GSE165107. To review GEO accession GSE165107: Go to https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE165107 , Enter token obkpqysqfrcvdot into the box.

Field-specific reporting

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

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

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

Life sciences study design

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

Sample size	Samples size for each experiment is indicated in the figures or corresponding figure legends. The sample size was chosen based on previous experience in the lab, for each experiment to yield high power to detect specific effects. For smRNA-FISH on wild type and KO embryos, the experiments were performed on at least 3 different litters per genotypes and on n>4 embryos for each genotypes. Embryos were carefully stage-matched for comparison. To match the ethical rules, we have analyzed the maximum number of embryos to see reproducibility and robustness of the experiments. No statistical methods were used to predetermine sample size.
Data exclusions	For the rigourous phenotypic analysis of Zic3 KO embryos, we have compared embryos that were strictly staged-matched when measuring the mean thickness of CM layer. In this situation, we have thus excluded embryos that showed developmental delay on a pre-established basis.
Replication	Reproducibility of findings were assessed via statistical tests, varying on sample size and experiment type. All the experiments were performed in at least 3 biologically independent embryos. All replicates reported in the manuscript and on which statistics are based are biological replicates. No technical replicates were used to calculate statistics. All attempts at replication of the results were successful.
Randomization	Since the wild-type and Mesp1 KO, chimeric embryos were not genotyped for the presence of the Y chromosome, sex-specific differences were minimized by including randomized numbers of male and female embryos. Each experiment contained animals from at least 2 different litters. Although Zic3 gene is found on the X chromozome, we have designed crosses to obtain the same probabilties of obtaining male and female Zic3 KO embryos (although male are hemizygous while females are homozygous).
Blinding	Investigators were blinded to mouse genotypes during experiments, for performing sample analysis, imaging and quantification and genotyping was only performed at the end of the experiment.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

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

Methods

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

Antibodies

Antibodies used

The following primary antibodies were used:
Rabbit anti-HA (cat: ab9110, Abcam, ChIP-grade)
Rabbit anti-H3K27ac (Cat: ab4729,Abcam, ChIP-grade)
Mouse anti-H3K4me1 (cat: 39635, Active Motif)
Rabbit anti Zic3 (cat: ab222124, Abcam)
Rabbit anti Zic2 (cat: epr7790, Abcam)
Rabbit Igg control Chip grade (cat: ab171870, Abcam)
Mouse anti-HA (cat: ab1424, Abcam)
Rabbit anti-HA (cat: ab9110, Abcam)
Rabbit anti-actin (cat: ab8227, Abcam)
Biotin conjugate Rat anti-Flk1 (Cat: 13-5821, 1:100,eBioscience, clone Avas12a1)
PE conjugate Rat anti-PDGFRa (Cat: 12-1401-81, 1:100, eBioscience, clone APA5)
mouse anti-troponin T (clone 13-11, 1:100, Thermo)

Rat anti-VE-cadherin (cat: 11D4.1, 1:200, BDBioscience)
mouse anti-cardiac troponin T (Invitrogen; 1/100)
Isolectin GS-IB4 (cat: I21411, 1/100; Thermo).

The secondary antibodies used are:
HRP anti-rabbit (cat: RPN4301, Healthcare)
HRP anti-mouse (cat: RPN4201, Healthcare)
PE-Cy7 anti-streptavidin (cat: BD557598, BD Bioscience)
APC anti-mouse (cat: BD550874, BD Bioscience)
RRX anti-mouse (cat: AB_2340832, Jackson Laboratories)
RRX anti-rat (cat: AB_2340676, Jackson Laboratories)
Alexa Fluor 647 anti-mouse (cat: A-21235, Invitrogen)

All antibodies with dilutions and applications are listed in Supplementary Table 4

Validation

- Rabbit anti-HA:
validation reference for ChiP and Western Blot: https://www.abcam.com/ha-tag-antibody-chip-grade-ab9110.html?productWallTab=ShowAll
- Rabbit anti-H3K27ac:
validation reference for ChiP: https://www.abcam.com/histone-h3-acetyl-k27-antibody-chip-grade-ab4729.html
- Mouse anti-H3K4me1:
validation reference for ChiP: https://www.activemotif.com/catalog/details/39635.html
- Rabbit anti Zic3/ Rabbit anti Zic2 were validated by Wertern Blot in Zic2/Zic3 KO cell lines.
- Mouse anti-HA
validation refrence for Western Blot: https://www.abcam.com/ha-tag-antibody-12ca5-ab1424.html
- Rabbit anti-actin:
validation for western blot: https://www.abcam.com/beta-actin-antibody-ab8227.html
- Biotin conjugate Rat anti-Flk1
validation for Flowcytometry: https://www.bdbiosciences.com/content/dam/bdb/products/global/reagents/flow-cytometry-reagents/research-reagents/single-color-antibodies-ruo/555307_base/pdf/550549.pdf
- PE conjugate Rat anti-PDGFRa:
validation for flowcytometry: https://www.bdbiosciences.com/en-fr/products/reagents/flow-cytometry-reagents/research-reagents/single-color-antibodies-ruo/pe-rat-anti-mouse-cd140a.562776
- mouse anti-troponin T (clone 13-11, 1:100, Thermo):
validation for IF: https://www.thermofisher.com/antibody/product/Cardiac-Troponin-T-Antibody-clone-13-11-Monoclonal/MA5-12960
- Rat anti-VE-cadherin (cat: 11D4.1, 1:200, BDBioscience)
validation for IF: https://www.bdbiosciences.com/en-fr/products/reagents/flow-cytometry-reagents/research-reagents/single-color-antibodies-ruo/purified-rat-anti-mouse-cd144.550548
- Isolectin GS-IB4 (cat: I21411, 1/100; Thermo).
validation from publication: Prados et al. 2018 Cell Death and Disease (PMID: 29540665)

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	Mesp1_3HA dox inducible cell line and all dervied KO clones were generated from A2lox mouse PSC line that originally provided by Dr. Michael Kyba., who established the cell line (PMID: 11955444)
Authentication	The new generated PSC cell lines for this study were not authenticated.
Mycoplasma contamination	All cell lines have been tested negative for Mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified lines were used in this study.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Mesp1–Cre mice were previously obtained from Y. Saga (Saga et al. 1999). Females between 8 to 40 weeks were used for breeding and dissected to collect embryos, 7 to 14 days after the observation of a vaginal plug. Zic3 KO mice were generated in a facility at the Université Catholique de Louvain, then imported in the local campus animal house. Mice colonies were maintained in certified animal facilities in accordance with European guidelines, with a 7h to 19h light cycle. Embryos were collected 7 to 14 days after the observation of a vaginal plug. The room temperature ranged from 20 and 25°C. The relative ambient humidity at the level of mouse cages was 55 per cent +/-15. Each cage was provided with food, water and two types of nesting material.
Wild animals	No wild animals were involved in this study.
Field-collected samples	No field collected samples were used in this study.
Ethics oversight	The experiments done in CB lab were approved by the local ethical committee (CEBEA) under protocols #591N.

Ethics oversight

The experiments in FL lab were approved by the national ethical committee under protocols Apafis #13031 (FL) and work in animal facility is approved by a National agrrement (#B1301308).

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

ChIP-seq

Data deposition

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

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

Data access links

May remain private before publication.

ChIPseq Data has been deposited in Gene Expression Omnibus (GEO) and is accessible through GEO Series accession number GSE165107. To review GEO accession GSE165107: Go to https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE165107 , Enter token obkpqysqfrcvdot into the box.

Files in database submission

Processed file (bigwig file):
Mesp1_ChIP-seq-12h-dox.bw
Mesp1_ChIP-seq-24h-dox.bw
H3K27Ac-12h-nodox.bw
H3K27Ac-12h-dox.bw
H3K27Ac-24h-nodox.bw
H3K27Ac-24h-dox.bw
Zic3-ChIP-seq-subsampled.bw
XLI1_Input_2i.bw
XLI2_2i_Mesp1_ChIP.bw
XLI3_Input_Zic23_dKO.bw
XLI4_Z23-dKO_Mesp1_ChIP.bw
XLI5_Z23-dKO_Zic2_ChIP.bw
XLI6_WT_Zic2_ChIP_d2-5.bw
XLI7_Zic3_d2-5.bw
XLI8_H3K4me_d2-5.bw
XLI10_Input_d2-5.bw
XLI11_Input_d3-5-ND.bw
XLI13_Zic2_ChIP_d3-5-ND.bw
XLI14_Zic3_d3-5-ND.bw
XLI15_Input_D3-5-DOX.bw
XLI16_H3K4Me_d3-5-DOX.bw
XLI17_Zic2_ChIP_d3-5-DOX.bw
XLI18_Zic3_ChIP_d3-5-DOX.bw
Mesp1-12hD-Rep2.bw
Mesp1-24hD-Rep2.bw
Mesp1-2i-Rep2.bw
Mesp1-Zdko-Rep2.bw
Mesp1-Zdko-Rep3.bw
Zic3-d2-5-Rep2.bw
Zic3-24hND-Rep2.bw
Zic3-24hDOX-Rep2.bw

Raw files (fastq data):
Mesp1_3HA.Chip.12hour.fastq.gz
Mesp1_3HA.Chip.24hour.fastq.gz
H3K27ac.Chip.12hour.noDox.fastq.gz
H3K27ac.Chip.12hour.Dox.fastq.gz
H3k27ac.Chip.24hour.noDox.fastq.gz
H3K27ac.Chip.24hour.Dox.fastq.gz
1510-XL07-12-ND-Input.fastq.gz
1506-X-INPUT.fastq.gz
1510-XL09-24-ND-Input.fastq.gz
1510-XL01-24-D-Input.fastq.gz
ZIC3-ChIP_S99_R1_001.fastq.gz
Zic3-input_S98_R1_001.fastq.gz
ZIC3-ChIP_S99_R2_001.fastq.gz
Zic3-input_S98_R2_001.fastq.gz
mm87-ci-XLI1-ESC-2i-input_S29_R1_001.fastq.gz
mm87-ci-XLI1-ESC-2i-input_S29_R2_001.fastq.gz
mm87-ci-XLI2-ESC-2i-HA-ChIP-mesp1_S30_R1_001.fastq.gz
mm87-ci-XLI2-ESC-2i-HA-ChIP-mesp1_S30_R2_001.fastq.gz
mm87-ci-XLI3-Zic2-3-KO-input_S31_R1_001.fastq.gz
mm87-ci-XLI3-Zic2-3-KO-input_S31_R2_001.fastq.gz
mm87-ci-XLI4-Zic2-3-HA-ChIP-mesp1_S32_R1_001.fastq.gz
mm87-ci-XLI4-Zic2-3-HA-ChIP-mesp1_S32_R2_001.fastq.gz
mm87-ci-XLI5-Zic2-3-Zic2-ChIP_S33_R1_001.fastq.gz
mm87-ci-XLI5-Zic2-3-Zic2-ChIP_S33_R2_001.fastq.gz

mm87-ci-XLI6-Ebs-day2-5-Zic2-ChIP_S2_R1_001.fastq.gz
mm87-ci-XLI6-Ebs-day2-5-Zic2-ChIP_S2_R2_001.fastq.gz
ULB-CB-mm87-ci-XLI7-Ebs-day2-5-Zic3-ChIP_S185_R1_001.fastq.gz
ULB-CB-mm87-ci-XLI7-Ebs-day2-5-Zic3-ChIP_S185_R2_001.fastq.gz
ULB-CB-mm87-ci-XLI9-Ebs-day2-5-H3K27ac-ChIP_S187_R1_001.fastq.gz
ULB-CB-mm87-ci-XLI9-Ebs-day2-5-H3K27ac-ChIP_S187_R2_001.fastq.gz
ULB-CB-mm87-ci-XLI10-Ebs-day2-5-Input_S188_R1_001.fastq.gz
ULB-CB-mm87-ci-XLI10-Ebs-day2-5-Input_S188_R2_001.fastq.gz
ULB-CB-mm87-ci-XLI11-Ebs-day3-5-nodox-Input_S189_R1_001.fastq.gz
ULB-CB-mm87-ci-XLI11-Ebs-day3-5-nodox-Input_S189_R2_001.fastq.gz
ULB-CB-mm87-ci-XLI13-Ebs-day3-5-nodox-Zic2_S191_R1_001.fastq.gz
ULB-CB-mm87-ci-XLI13-Ebs-day3-5-nodox-Zic2_S191_R2_001.fastq.gz
ULB-CB-mm87-ci-XLI14-Ebs-day3-5-nodox-Zic3_S192_R1_001.fastq.gz
ULB-CB-mm87-ci-XLI14-Ebs-day3-5-nodox-Zic3_S192_R2_001.fastq.gz
ULB-CB-mm87-ci-XLI15-Ebs-day3-5-dox-Input_S193_R1_001.fastq.gz
ULB-CB-mm87-ci-XLI15-Ebs-day3-5-dox-Input_S193_R2_001.fastq.gz
ULB-CB-mm87-ci-XLI16-Ebs-day3-5-dox-H3K4Me_S194_R1_001.fastq.gz
ULB-CB-mm87-ci-XLI16-Ebs-day3-5-dox-H3K4Me_S194_R2_001.fastq.gz
ULB-CB-mm87-ci-XLI17-Ebs-day3-5-dox-Zic2_S195_R1_001.fastq.gz
ULB-CB-mm87-ci-XLI17-Ebs-day3-5-dox-Zic2_S195_R2_001.fastq.gz
ULB-CB-mm87-ci-XLI18-Ebs-day3-5-dox-Zic3_S196_R1_001.fastq.gz
ULB-CB-mm87-ci-XLI18-Ebs-day3-5-dox-Zic3_S196_R2_001.fastq.gz
ULB-CB-ci-mm87-XLI11-Mesp13HA-2i-24h-dox-HA_S58_R1_001.fastq.gz
ULB-CB-ci-mm87-XLI11-Mesp13HA-2i-24h-dox-HA_S58_R2_001.fastq.gz
ULB-CB-ci-mm87-XLI13-Mesp13HA-day2-5-Input_S59_R1_001.fastq.gz
ULB-CB-ci-mm87-XLI13-Mesp13HA-day2-5-Input_S59_R2_001.fastq.gz
ULB-CB-ci-mm87-XLI14-Mesp13HA-12h-dox-input_S60_R1_001.fastq.gz
ULB-CB-ci-mm87-XLI14-Mesp13HA-12h-dox-input_S60_R2_001.fastq.gz
ULB-CB-ci-mm87-XLI15-Mesp13HA-24h-nodox-input_S61_R1_001.fastq.gz
ULB-CB-ci-mm87-XLI15-Mesp13HA-24h-nodox-input_S61_R2_001.fastq.gz
ULB-CB-ci-mm87-XLI16-Mesp13HA-24h-dox-input_S62_R1_001.fastq.gz
ULB-CB-ci-mm87-XLI16-Mesp13HA-24h-dox-input_S62_R2_001.fastq.gz
ULB-CB-ci-mm87-XLI18-Mesp13HA-2i-24h-dox-input_S64_R1_001.fastq.gz
ULB-CB-ci-mm87-XLI18-Mesp13HA-2i-24h-dox-input_S64_R2_001.fastq.gz
ULB-CB-ci-mm87-XLI2-Mesp13HA-day2-5-zic3_S49_R1_001.fastq.gz
ULB-CB-ci-mm87-XLI2-Mesp13HA-day2-5-zic3_S49_R2_001.fastq.gz
ULB-CB-ci-mm87-XLI3-Mesp13HA-12h-dox-HA_S50_R1_001.fastq.gz
ULB-CB-ci-mm87-XLI3-Mesp13HA-12h-dox-HA_S50_R2_001.fastq.gz
ULB-CB-ci-mm87-XLI5-Mesp13HA-24h-nodox-zic3_S52_R1_001.fastq.gz
ULB-CB-ci-mm87-XLI5-Mesp13HA-24h-nodox-zic3_S52_R2_001.fastq.gz
ULB-CB-ci-mm87-XLI7-Mesp13HA-24h-dox-zic3_S54_R1_001.fastq.gz
ULB-CB-ci-mm87-XLI7-Mesp13HA-24h-dox-zic3_S54_R2_001.fastq.gz
ULB-CB-ci-mm87-XLI8-Mesp13HA-24h-dox-HA_S55_R1_001.fastq.gz
ULB-CB-ci-mm87-XLI8-Mesp13HA-24h-dox-HA_S55_R2_001.fastq.gz

Genome browser session
(e.g. UCSC)

N/A

Methodology

Replicates

Mesp1 CHIP was done in duplicate at two time points in dox conditions, and in duplicate in 2i and Zic2/3 KO conditions. H3K27ac ChIP was done single time at two time points in both no dox and dox condition, H3K4me1 ChIP-seq was done once in a one condition; Zic2 ChIP-seq samples were performed once n 3 different conditions.; Zic3 ChIP-seq samples were performed 2 times in 3 different conditions.

Sequencing depth

Sample, Read Depth, Uniquely Mapped, Read Length, Single/Paired-End
Mesp1-ChIP-seq-12h-dox, 50025815, 33853361, 75nt, single-End;
Mesp1-ChIP-seq-24h-dox, 59553038, 39078399, 75nt, single-End;
H3K27Ac ChIP-seq 0h, 79406221,87912894, 100nt, paired-End;
H3K27Ac-ChIP-seq-12h-no-dox, 35844148, 27313255,75nt,single-End;
H3K27Ac-ChIP-seq-12h-dox, 29354082, 23211995, 75nt, single-End;
H3K27Ac-ChIP-seq-24h-no-dox, 29777789, 15571040, 75nt, single-End;
H3K27Ac-ChIP-seq-24h-dox, 37563862 , 27240731, 75nt, single-End;
Zic3-ChIP-seq-24h-dox, 346233456, 215104788, 100nt, paired-End;
Mesp1 ChIP-seq in 2i, 72437960, 66499524, 100nt, paired-End;
Mesp1 ChIP-seq in Zic2/3 dKO, 69049557, 71681724, 100nt, paired-End;
Zic2 ChIP-seq in Zic2/3 dKO, 67675017, 80209570, 100nt, paired-End;
Zic2 ChIP-seq d2.5, 70873849, 68431382, 100nt, paired-End;
Zic2 ChIP-seq d3.5 no dox,93516744, 104607116, 100nt, paired-End;
Zic2 ChIP-seq d3.5 dox, 97113922, 122625358, 100nt, paired-end;
Zic3 ChIP-seq d2.5,78441541, 75940998, 100nt, paired-End;
Zic3 ChIP-seq d3.5 no dox,74636028, 66294276, 100nt, paired-End;
Zic3 ChIP-seq d3.5 dox,85042130, 72840936, 100nt, paired-End;
H3K4me1 ChIP-seq 24h dox,110615077, 93521226, 100nt, paired-End;
Mesp1 ChIP-seq 12h DOX rep2, , 69752554, 100nt, paired-End;

	Mesp1 ChIP-seq 24h DOX rep2, , 68791082, 100nt, paired-End; Mesp1 ChIP-seq in 2i rep2, , 38487160, 100nt, paired-End; Mesp1 ChIP-seq in Zic2/3 dKO rep2, , 40683540, 100nt, paired-End; Mesp1 ChIP-seq in Zic2/3 dKO rep3, , 115651346 , 100nt, paired-End; Zic3 ChIP-seq d2-5 rep2, , 45904902, 100nt, paired-End; Zic3 ChIP-seq d3-5 no dox rep2, , 62435240, 100nt, paired-End; Zic3 ChIP-seq d3-5 dox rep2, , 52177232, 100nt, paired-End;
Antibodies	Rabbit anti-HA (cat: ab9110, Abcam, ChIP-grade) and Rabbit anti-H3K27ac (Cat: ab4279,Abcam, ChIP-grade) has been validated by the vendor for ChIP.
Peak calling parameters	Macs2 callpeaks using a p-value threshold of 1e-10 or a q-value of 0.05 for endogenous proteins
Data quality	Data quality was assessed via the number of peaks detected at our stringent p-value cutoff
Software	Trimmomatic for removing sequencing adapters, Bowtie2 for alignment, Picardtools and Samtools for quality and duplicate reads filtering, macs2 for calling peaks, Htseq-count for quantifying reads in peaks

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	For staining cell surface markers, day 4 or day 5 EBs were dissociated in 3mM EDTA. Day 10 EBs were dissociated for 1-2 hours at 37 °C by a mixture of collagenase I (1mg/ml), collagenase II (1mg/ml, Sigma C2-BIOC) and collagenase IV (1mg/ml, Sigma C4-BIOC) diluted in PBS.
Instrument	FACS Fortessa, BD
Software	FACS analyses were performed by BD FACS Diva V6.2
Cell population abundance	A minimum of 50,000 cells per sample were analyzed
Gating strategy	Live cells were Hoechst- (< 10^3) , 3X10^3 and above was set as fluorescent signal positive for PE anti-PDGFRa and 10^3 and above was set as fluorescent signal positive for PE-Cy7 anti Flik1, as shown in supplemental Fig. 5C.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.