Identifying the molecular basis for functional divergence of duplicated SOX factors controlling endoderm formation and left-right patterning in zebrafish

Simaran Johal[1], Randa Elsayed[2], Kristen A. Panfilio[1,3,4], Andrew C. Nelson[1*]

1. School of Life Sciences, Gibbet Hill Campus, University of Warwick, Coventry, CV4 7AL, UK

2. Warwick Medical School, Gibbet Hill Campus, University of Warwick, Coventry, CV4 7AL, UK

3. Department of Molecular Genetics, Institute of Biology, University of Hohenheim, Garbenstr. 30, 70599 Stuttgart, Germany

4. Institute for Zoology: Developmental Biology, University of Cologne, Zülpicher Straße 47b, 50674 Cologne, Germany

*Corresponding author
Email: A.Nelson.1@warwick.ac.uk (ACN)

Abstract

Endoderm, one of three primary germ layers of vertebrate embryos, makes major contributions to the respiratory and gastrointestinal tracts and associated organs, including liver and pancreas. Placement and patterning of these organs relies on the left-right organiser – known as Kupffer's Vesicle (KV) in zebrafish. The transcription factors Sox32 and Sox17 are members of the zebrafish SoxF subfamily. *Sox32* and *sox17* arose from a duplication of ancestral *Sox17* in the teleost lineage. Sox32 induces *sox17* expression in the early embryo and is required for the specification of endoderm and KV progenitors. Zebrafish Sox17 is implicated in KV morphogenesis. In mammals, *Sox17* is vital for endoderm organ formation and can induce endoderm progenitor identity. Phenotypic evidence therefore suggests functional similarities between zebrafish Sox32 and Sox17 with mammalian SOX17. We sought to explore the functional differences and potential similarities between these proteins in the early zebrafish embryo. Our results indicate that, unlike Sox32, human SOX17 cannot induce endoderm specification in zebrafish. Furthermore, using hybrid protein functional analyses, we show that Sox32 specificity for the endoderm gene regulatory network is linked to evolutionary divergence in its HMG domain from its paralogue Sox17. Additionally, changes in the C-terminal regions of Sox32 and Sox17 underpin their differing target specificity and divergence in mediating differential gene regulatory programmes. Finally, we establish that specific conserved peptides in the C-terminal domain are essential for the role of Sox17 in establishing correct organ asymmetry. Overall, our results provide novel insights into vertebrate endoderm development, left-right patterning, and the evolution of SoxF transcription factor function.

Introduction

Sox family transcription factors (TFs) are evolutionarily conserved proteins with roles in cell fate decisions in a range of developmental processes (Sarkar and Hochedlinger, 2013). Mutations in Sox factors have been linked to an array of developmental defects and congenital diseases in humans (Angelozzi and Lefebvre, 2019). Sox TFs belong to the superfamily of High Mobility Group (HMG) domain containing TFs. Historically, membership of the Sox family has been dictated by having >50% conservation of the DNA-binding HMG domain with mouse SRY – the founding Sox TF (Pevny and Lovell-Badge, 1997, Gubbay et al., 1990, Sinclair et al., 1990). Classically, Sox factors are further characterised into subfamilies, whereby members of a subfamily exhibit 75-80% homology within their HMG domain (Wegner, 2010, Bowles et al., 2000). One subfamily is SoxF, which in mammals consists of SOX7, SOX17 and SOX18. However, SoxF is expanded to include Sox32 in teleosts.

Most duplicated *sox* genes in the teleost lineage arose through a whole genome duplication event that occurred 226-316 million years ago (mya) (Voldoire et al., 2017). An exception to this is *sox32,* which is hypothesised to have emerged through tandem duplication of *Sox17* (Voldoire et al., 2017). While we and others have speculated that zebrafish Sox32 may function in similar processes to mammalian SOX17 (Figiel et al., 2021, Lilly et al., 2017), this has not been explored at the molecular level. Furthermore, it is unknown whether functions of the ancestral gene have been split among the two putative paralogues (sub-functionalisation) or whether they have functionally diverged post-duplication.

Sox32 is considered the master-regulator of endoderm identity in zebrafish. Loss-of-function mutations in *sox32* lead to a profound loss of endoderm progenitors and the consequent absence of the respiratory tract and gut, and its associated organs including the liver and pancreas. Instead, presumptive endoderm cells differentiate into mesoderm (Alexander et al., 1999, Dickmeis et al., 2001). Conversely, *sox32* overexpression (OE) through mRNA injection during early development is sufficient to respecify presumptive mesodermal cells to endodermal fates, resulting in ectopic endoderm formation (Kikuchi et al., 2001). Furthermore, transplantation studies show that cells from *sox32* mRNA-injected donor embryos transplanted at mid/late-blastula stages result in preferential incorporation into developing endoderm tissues such as the gut tube and pharyngeal endoderm in wild type hosts, and endoderm restoration in Sox32-deficient hosts (Kikuchi et al., 2001, Aoki et al., 2002, Stafford et al., 2006).

Conversely, Sox17 null mutant mice are embryonic lethal and show narrower expression domains of definitive endoderm markers along the embryonic gut, indicating that loss of Sox17 results in depletion of the definitive endoderm (Kanai-Azuma et al., 2002). Furthermore, ectopic expression of SOX17 in mouse and human embryonic stem cells causes differentiation towards the extraembryonic and definitive endoderm fates, respectively (Niakan et al., 2010, McDonald et al., 2014, Séguin et al., 2008). Thus, phenotypic evidence suggests functional similarity between SoxF subfamily members, with both mammalian SOX17 and zebrafish Sox32 inducing endoderm cell identity within their respective species. However, a molecular analysis is lacking.

In zebrafish, Sox32 is also required for correct specification of the Dorsal Forerunner Cells (DFCs) – progenitors of the left-right (LR) organiser, which is essential for correct positioning of endoderm-derived organs within the body cavity (Essner et al., 2005, Alexander et al., 1999). Specifically, the DFCs give rise to the zebrafish laterality organ, known as Kupffer's Vesicle (KV) (Cooper and D'Amico, 1996, Melby et al., 1996, Warga and Kane, 2018, Essner et al., 2005). KV function involves cilia-mediated fluid flow in an anti-clockwise direction, without which laterality defects arise affecting organs including the brain, heart, and pancreas (Essner et al., 2005, Kramer-Zucker et al., 2005). This occurs due to

abnormalities in the conserved unilateral expression of Nodal. In teleosts, this Nodal expression is of a teleost-specific Nodal - *spaw* (Long et al., 2003, Montague et al., 2018). The downstream target of Sox32 – *sox17* – is implicated in KV function (Aamar and Dawid, 2010). Similarly, mammalian Sox17 is implicated in the establishment of LR patterning in mouse as a secondary effect of an inability to maintain the gut endoderm in Sox17$^{-/-}$ mutants (Kanai-Azuma et al., 2002). An outstanding question is which regions of the Sox17 protein mediate this key function and consequent organ patterning and placement.

Building on previous phenotypic observations, and to gain novel insights at the molecular level, here we examine whether SoxF factors exhibit functional conservation across species and identify functional domains within the zebrafish Sox32 and Sox17 proteins that are required for their respective developmental roles in endoderm specification and LR patterning.

We have performed molecular dissection and hybrid protein functional analyses of zebrafish Sox32 and Sox17 and drawn functional comparisons with human SOX17 (hSOX17). Our molecular analyses indicate that Sox32 shares some limited functional similarities with both hSOX17 and Sox17. However, Sox17 and hSOX17 differ from each other in terms of molecular capabilities in early zebrafish embryos. Furthermore, hSOX17 cannot induce endoderm specification within the zebrafish embryo, unlike Sox32. Sox32 specificity for the zebrafish endoderm gene regulatory programme appears to be linked to evolutionary divergence in its HMG domain after duplication of the ancestral *Sox17*. Further divergence of Sox32- and Sox17-mediated gene regulatory programmes are driven by changes in the C-terminal regions of Sox32 and Sox17, which underpin their differing target specificity. Finally, the C-terminal domain (CTD) of Sox17 is essential for its role in organ asymmetry. Overall, our analyses reveal that Sox32 and Sox17 functional differences are driven by evolutionary divergence within their HMG and C-terminal domains. Furthermore, we identify the C-terminal domain as vital for the role of Sox17 in the correct placement of organs within the vertebrate body plan.

Results

hSOX17 cannot induce endoderm identity in zebrafish embryos

The ability of human SOX17 (hSOX17) to induce endoderm identity in human embryonic stem cells and of Sox32 to induce ectopic endoderm in early zebrafish embryos raises the question of whether hSOX17 and zebrafish Sox32 and Sox17 have similar molecular capabilities. Whether zebrafish Sox17 can induce endoderm gene expression as Sox32 can is also unknown. To test this in the context of the early zebrafish embryo we injected equimolar quantities of mRNA encoding hSOX17, Sox32 or Sox17 at the one-cell stage, followed by RNA-seq analysis at 6 hours post-fertilisation (hpf), the stage by which endoderm has been specified in wild type embryos (Figure 1A).

Sox32 OE elicited widespread changes in gene expression compared to Sox17, with hSOX17 affecting relatively few genes (Figure 1B-C, S1). We observed a small but significant overlap in upregulated transcripts between Sox32 and hSOX17, and Sox32 and Sox17, but not between hSOX17 and Sox17 (Figure 1C-F). Overall, this suggests Sox32 is functionally more similar to both hSOX17 and Sox17 than hSOX17 and Sox17 are to each other in the context of the early zebrafish embryo.

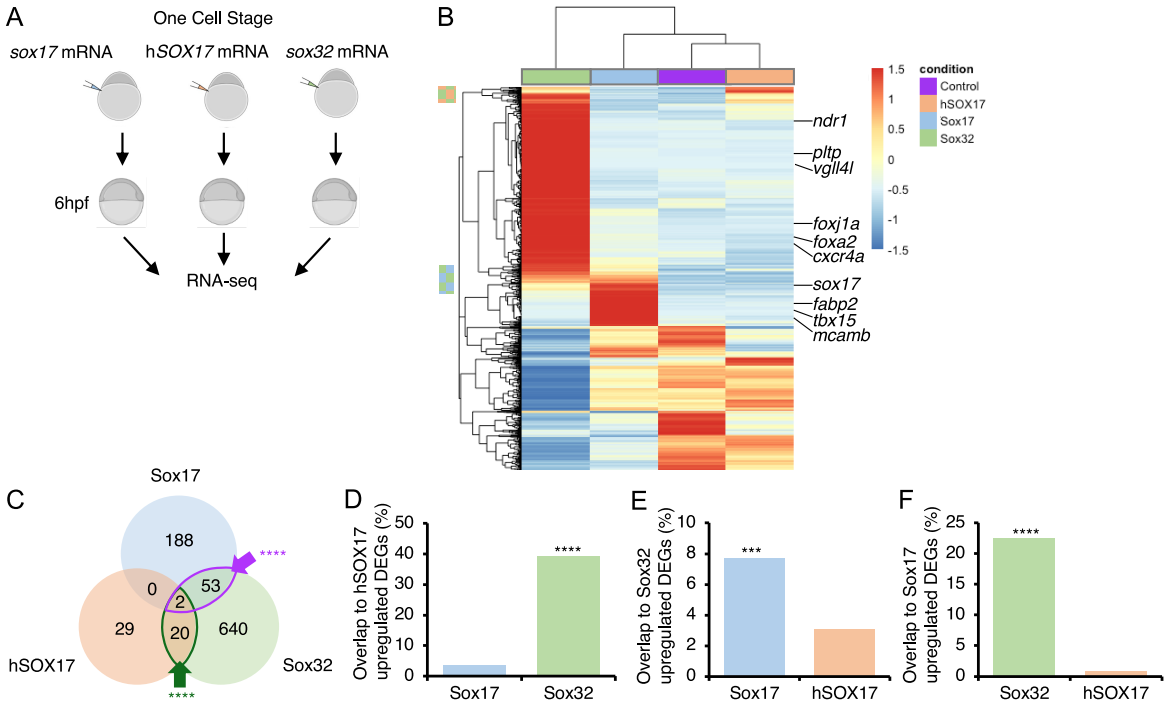

Figure 1: Sox32, Sox17 and hSOX17 are capable of inducing largely distinct gene sets in the early zebrafish embryo. A) Experimental schematic. **B)** Heatmap of all differentially expressed genes (DEGs), displayed as median counts from N=3 per condition. Colour-coded blocks to the left indicate genes induced by both hSOX17 and Sox32, and by Sox17 and Sox32. **C)** Venn diagram showing overlap of genes significantly upregulated by each TF ($P < 0.05$). Overlaps significantly greater than expected according to Fisher's Exact test are indicated. **** $P < 0.0001$. **D-F)** Bar charts indicating percentage overlap of upregulated DEGs to (D) hSOX17, (E) Sox32, and (F) Sox17 upregulated DEGs. Bars representing sets of upregulated genes showing significantly greater overlap to the factor on the y-axis compared to the other bar are indicated with asterisks. Statistical tests were carried out using Fisher's Exact test. * $P<0.05$, *** $P<0.001$, **** $P <0.0001$.

To examine the pathways and processes influenced by OE of these TFs we performed functional and anatomical annotation analysis of the differentially expressed genes (DEGs). This revealed largely distinct results for each TF, further supporting the notion that they are functionally distinct in this context (Figure S2-S3). As expected, transcripts upregulated by Sox32 OE were significantly enriched for terms associated with endoderm development (Figure S3). In comparison, transcripts upregulated by Sox17 OE show a modest enrichment for endoderm functional annotations. Furthermore, Sox17 OE does not result in the enrichment of cell identity terms consistent with increased specification of endoderm, such as "nondorsal involuted anterior" as Sox32 does (Figure S3). Meanwhile, genes upregulated after hSOX17 ectopic expression did not yield any significantly enriched biological or anatomical terms, while downregulation effects suggest mis-regulation of Wnt signalling (Figure S2), but no clear endoderm-related target DEGs.

Beyond endoderm development, analysis of transcripts upregulated by Sox17 revealed significant enrichment for terms associated with cardiovascular development, such as "Regulation of Sprouting Angiogenesis", "Endothelial Tube Morphogenesis" and "Endocardium" (Figure S2). This is consistent with *sox17* expression patterns during cardiovascular development in both zebrafish and mammals (Chung et al., 2011, Saba et al., 2019) (Figure S4B-C), and established *Sox17* loss-of-function phenotypes in mice (Kanai-Azuma et al., 2002, Saba et al., 2019). Additionally, transcripts upregulated by Sox17, indicate enrichment for DFC expression, supporting evidence indicating a role for Sox17 in LR asymmetry (Aamar and Dawid, 2010)

Thus, functional enrichment analysis shows that Sox17 has different roles in early developmental processes than its paralogue Sox32 and orthologue hSOX17, and hSOX17 cannot induce zebrafish endoderm specification.

In comparison to its paralogue Sox17, Sox32 has a divergent HMG domain and longer C-terminal helices

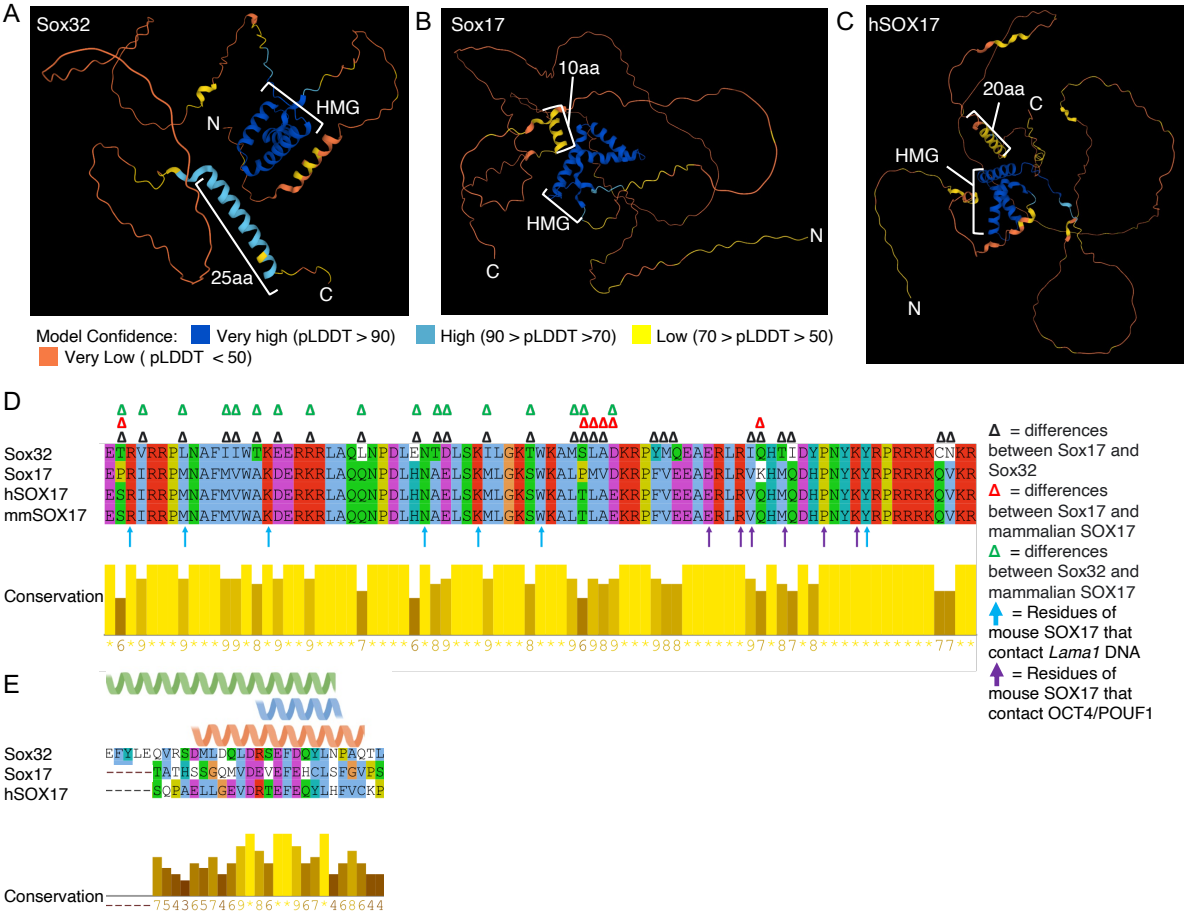

Figure 2: AlphaFold predicts C-terminal helical domains within Sox32, Sox17 and hSOX17. (A-C) Alphafold models coloured by model confidence as shown in the key. DNA-binding HMG domains and equivalently positioned C-terminalC-terminal helix of different lengths are annotated. **(D)** Alignment of HMG domain of Sox32, Sox17, hSOX17 and *Mus musculus* (mm)Sox17.Alignment carried out using Clustal Omega in Jalview, and colour coded based on conserved amino acid properties. Conservation levels calculated according to Analysis of Multiply Aligned Sequences (AMAS) displayed. Delta symbols denote differences in amino acids as per the key. Arrows indicating mapped contact sites between mouse SOX17 and the *Lama1* enhancer (Blue), SOX and OCT4/POU1 (Purple). (Ng et al., 2012, Palasingam et al., 2009) . **(E)** Alignment of CTD helical domain region for Sox32, Sox17 and hSOX17. Amino acids within Alphafold predicted helical domains annotated: Green = Sox32; blue = Sox17; peach = hSOX17.

To determine which structural motifs may explain the functional differences between Sox32, Sox17 and hSOX17 we analysed AlphaFold structural predictions and performed protein sequence alignments (Figure 2A-E). 3D structural models indicate a very high confidence prediction in the centre of all three proteins corresponding to the HMG domain. Although the HMG domains of the three proteins show a high degree of conservation, there are multiple differences. The HMG domains of Sox17 and hSOX17 are most similar, with 92% conservation in amino acid identity, while the Sox32 HMG domain differs from both Sox17 and hSOX17 with 67% and 69% conservation, respectively (Figure 2D). Interestingly, 44% of amino acid differences between Sox32 and Sox17 also correspond to changes in

chemical properties that may influence the 3D folding of the protein, resulting in exposure of different residues to binding partners. Overall, this indicates that the Sox32 HMG domain shows divergence from its paralogue and orthologue counterparts, reflecting a higher rate of evolutionary change in this domain after the original Sox17/Sox32 duplication event.

Another notable structured region within Sox32 is a 25 aa long predicted C-terminal helix (Figure 2A). C-terminal helices were also predicted in Sox17 (Figure 2B) and hSOX17 (Figure 2C), albeit the confidence of the prediction is lower, and they are shorter than that of Sox32. Amino acid sequence alignments of the putative helices suggest the overall level of conservation is low, though they do encompass a seven amino acid peptide region previously recognised to be conserved throughout the SoxF subfamily, including with Sox7 and Sox18 (Sinner, 2004, Kikuchi et al., 2001) (Figure S5B).

Overall, we conclude that the Sox32 HMG domain shows a higher degree of divergence from the common ancestor than Sox17 or hSOX17. Additionally, we identify a new high confidence putative C-terminal domain (CTD) helix in Sox32 that encompasses the seven amino acid peptide region (EFDQYLN) previously shown to be required for Sox32 induced endoderm specification (Zhao et al., 2013).

Sox32 specificity for the endoderm gene regulatory network is driven by HMG domain divergence

In mammals, ectopic expression of hSOX17 in embryonic stem cells (ESCs) results in differentiation to endodermal states (Séguin et al., 2008). However, we show that expression of hSOX17 in the zebrafish embryo does not induce endoderm specification, and nor does zebrafish Sox17. Therefore, Sox32 has a unique property allowing it to engage the zebrafish endoderm gene regulatory network (GRN). We hypothesise that differences between the Sox32 HMG domain compared to hSOX17 and Sox17 may confer this function (Figure 2D). To examine this, we tested the ability of a hybrid protein (HMG Switch) consisting of the Sox32 N- and CTDs flanking the Sox17 HMG domain to induce gene expression via mRNA injection at the one-cell stage followed by RNA-seq at 6hpf as before (Figure 3A).

The HMG domain of SOX factors is considered the DNA-binding domain, therefore it is predicted that deletion of the HMG domain will block Sox32-induced effects on the transcriptome, as seen by the inability of Sox32ΔHMG to induce Sox32 target genes (Figure 3B). HMG Switch-induced genes show significant overlap with both Sox32 and Sox17-induced genes (Figure 3C). To dissect whether the HMG switch is more similar to Sox32 or Sox17 in the genes it induces, pairwise overlaps between the three factors were analysed. If the HMG Switch were more similar to Sox32 we would expect to see a significantly higher overlap between HMG Switch DEGs and Sox32 DEGs than HMG Switch DEGs to Sox17 DEGs. However, we see no significant difference between overlaps of Sox32 or Sox17 DEGs with HMG Switch DEGs (Figure 3D), therefore the HMG Switch is not more similar to either Sox32 or Sox17 than the other. While genes induced by the HMG Switch show overlaps with Sox32 and Sox17-induced genes, 58.8% of total HMG Switch-induced genes are not induced by either Sox32 or Sox17. Furthermore, 58.9% of total Sox32-induced genes are not induced by HMG Switch (Figure 3C). Taken together this indicates that amino acid differences distinguishing Sox32 and Sox17 HMG domains confer significant functional differences. Notably, the Sox17, HMG Switch intersect with Sox17 encompasses the majority of Sox17-induced genes, suggesting the HMG domain is the most important domain for induction of Sox17 targets. However, 40 genes are nevertheless induced by Sox17 but not the HMG Switch version, indicating the Sox17 HMG domain is not the sole influence on Sox17 target gene expression, and other domains of the protein are important for a subset of targets (Figure 3C).

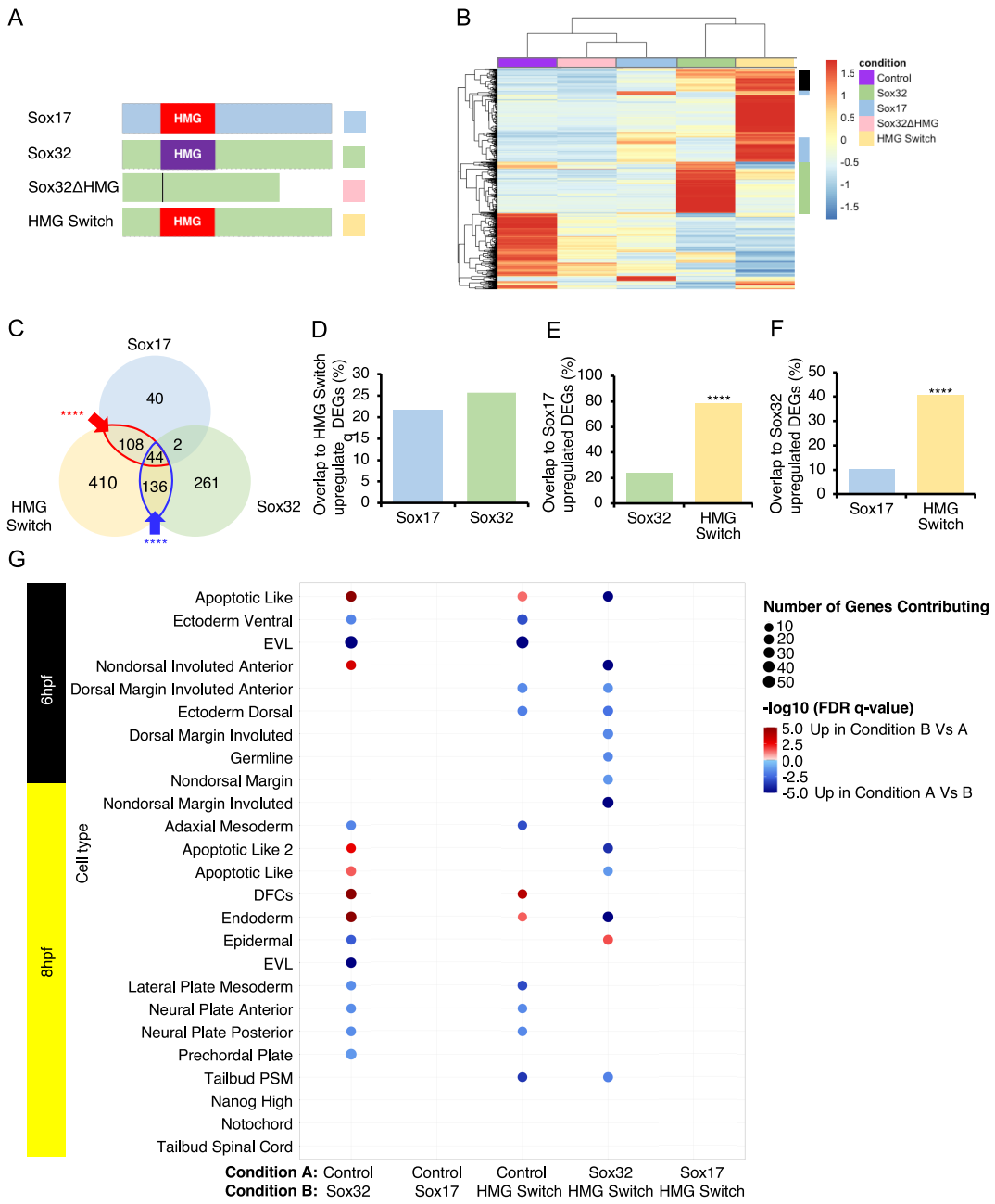

Figure 3: Differences between the Sox32 and Sox17 HMG domains are critical to the ability of Sox32 to induce endoderm and DFC gene expression. A) Protein models depicting constructs injected at the one-cell stage followed by RNA-sequencing at 6hpf. **B)** Heatmap of all (1572) DEGs, depicted as median counts from N=2 per condition. Green sidebar: induction of genes requiring the Sox32 HMG domain. Blue sidebar: induction of genes requiring the Sox17 HMG domain. Black sidebar: genes induced by Sox32 with either Sox32 or Sox17 HMG domain, referred to as "HMG indifferent". **C)** Overlap of upregulated DEGs. Overlaps greater than expected according to Fisher's Exact test are indicated - **** P < 0.0001. **D-F)** Percentage overlap of upregulated DEGs to (D) HMG Switch, (E) Sox17, (F) Sox32 upregulated DEGs. Bars representing sets of upregulated genes showing significantly greater overlap to the factor on the y-axis compared to the other bar are indicated with asterisks. Statistical tests were carried out using Fisher's Exact test. **** P < 0.001. **G)** Gene Set Enrichment Analysis (GSEA) carried out against cell type markers defined by scRNA-seq at 6 and 8 hpf (Wagner et al., 2018).

We further identify a subset of Sox32 and Sox17-induced genes that can be induced by either the Sox32 or Sox17 HMG domain (44) and are therefore HMG domain indifferent. However, there are also subsets which require either the Sox32 HMG domain (261) or the

Sox17 HMG domain (40,410 & 108) (Figure 3C). Interestingly another subset appears to be mediated by the C/NTD of Sox32 (136), as these genes are induced by the HMG Switch and Sox32 – which contain the Sox32 C/NTD's – but not by Sox17 (Figure 3C). Taken together this indicates differences in amino acids within the HMG domain confer target specificity and functional differences between Sox32 and Sox17. However as expected, the domains flanking the HMG domain also influence target gene induction.

To determine whether differences in the HMG domain influence the ability of Sox32 to induce cell identities, functional annotation was carried out using cell markers identified by scRNA-seq (Wagner et al., 2018). This indicated that Sox32 has a stronger capacity to significantly induce markers of endoderm and DFC cell types than the HMG Switch, and this enrichment occurs at the expense of other cells such as the adaxial mesoderm and ventral ectoderm (Figure 3G). This suggests the Sox32 HMG domain shows endodermal target specificity and is required to induce cell fate changes.

Overall, we conclude that the divergence of the Sox32 HMG domain from ancestral Sox17 is critical in conferring endoderm and DFC target specificity within the zebrafish embryo.

The Sox32 25 aa C-terminal helix is required for induction of endoderm and DFC marker genes

We have shown the Sox32 HMG domain is required to induce endoderm and DFC gene expression programmes (Figure 3G). Our analyses also demonstrate the terminal domains of Sox32 are important for target gene induction (Figure 3C). We therefore next explored which peptides outside the Sox32 HMG domain are required for target gene induction.

The Sox32 N-terminal domain (NTD) contains a conserved EKR phosphorylation domain which when phosphorylated in response to FGF signalling leads to attenuation of Sox32 function, antagonising endoderm specification (Poulain, 2006). Furthermore, examination of conservation across *sox32* orthologues and the zebrafish SoxF subfamily revealed the conservation of two peptides within the Sox32 CTD. Specifically, a 23 aa peptide conserved among *sox32* orthologues (Figure S5A) and a downstream 7 aa peptide show complete conservation between all members of the zebrafish SoxF subfamily, except Sox17 (Figure S5B). Previous studies have identified this 7 aa peptide as showing higher conservation between Sox32 and hSOX17 than Sox17 (Sinner, 2004, Kikuchi et al., 2001) and have shown its importance in Sox32 function regarding formation of ectopic endoderm upon OE (Zhao et al., 2013). Furthermore, AlphaFold confidently predicts a 25 aa long CTD helix which partly encompasses these 2 peptides (Figure 2).

To examine the importance of these identified peptides to Sox32 target induction we carried out specific domain deletions. We then analysed the ability of these domain-deleted mRNAs to induce endoderm and DFC cell markers relative to full length (FL)-Sox32 (Figure 4A). All selected markers show induction by FL-Sox32 in RNA-seq experiments (Figure 1B) and spatial expression in either the endoderm only (*foxa2*) (Alexander et al., 1999), endoderm and DFCs (*sox17, cxcr4a*) (Kikuchi et al., 2000, Thisse, 2001) or DFCs only (*vgll4l, ndr1*) (Thisse, 2001, Rebagliati et al., 1998).

We identify the HMG domain as required for induction of all analysed genes (Figure 4B-F), supporting the idea that deletion of the HMG domain from Sox32 blocks or reduces target induction (Figure 3B). We find the NTD of Sox32 is largely dispensable in its ability to induce many endoderm and DFC cell markers, as Sox32ΔNTD can significantly induce all analysed genes compared to the control, and it has the capacity to induce all genes to a level similar to that of FL-Sox32, the exception being *vgll4l* where there is a significant reduction in induction by Sox32ΔNTD compared to FL-Sox32 (Figure 4B-F). Intriguingly, the ability of Sox32ΔNTD to significantly induce *vgll4l* compared to the control coupled with significant reduction compared to FL-Sox32 indicates a likely hypermorphic effect (Figure

4E). Furthermore, Sox32ΔNTD shows a significantly greater induction of *ndr1* compared to both the control and full-length Sox32 (Figure 4F). Analysis of the CTD deletions indicated that the 23 aa and 7 aa peptides are required for the induction of all analysed genes (Figure 4B-F).

Taken together, these results indicate that Sox32 requires its C-terminal helices for proper induction of its endodermal and DFC targets. Both the 23 aa and 7 aa helices are in part present within the CTD helix predicated by AlphaFold (Figure 2), therefore this helix is required for Sox32 target induction.

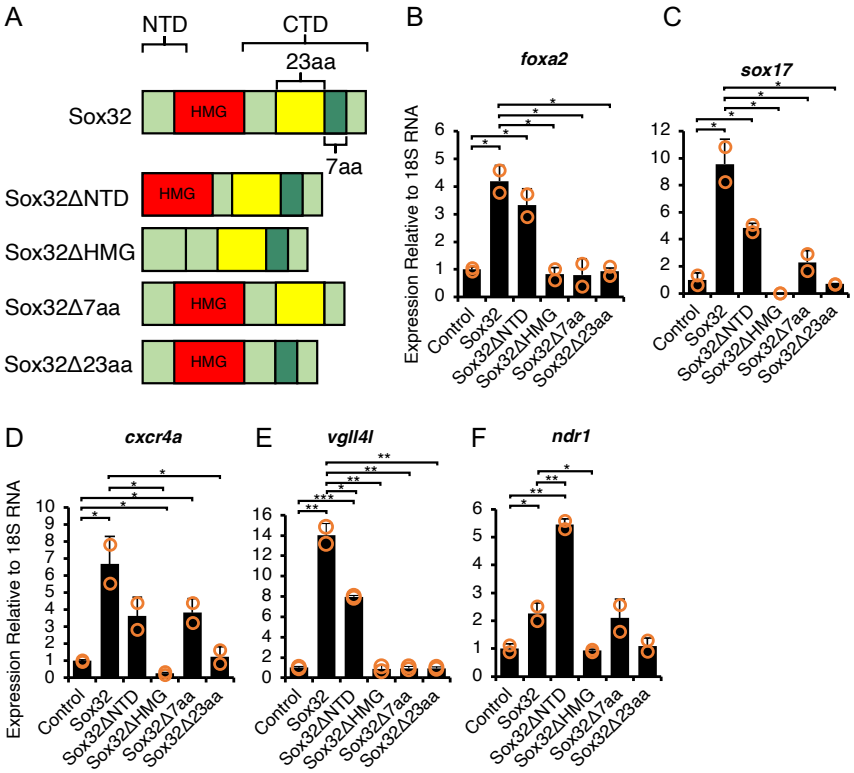

Figure 4: Putative helical peptides highly conserved between Sox32 CTD orthologues are required for induction of downstream endoderm and DFC marker genes. A) Protein models of Sox32 deletions injected at the one-cell stage and RNA extraction followed by qPCR at 6hpf. **B-F)** qPCR analysis of endodermal only (B), endoderm and DFC (C,D) and DFC only (E-F) markers. Statistical tests on biological duplicate samples were carried out using Student's *t*-test. * P<0.05, ** P <0.01, *** P<0.001.

Divergence of C-terminal peptides confers Sox32 and Sox17 differential target specificity

We have shown that divergence in the Sox32 HMG domain confers its endodermal target specificity and that Sox32 requires its CTD helix for proper induction of its endodermal and DFC targets. The Sox32 7 aa region that is part of the Sox32 CTD helix is required for the induction of Sox32 endoderm and DFC targets (Figure 4) and is required for the production of ectopic endoderm upon Sox32 OE (Zhao et al., 2013). This 7 aa region shows complete conservation to other zebrafish SoxF subfamily members Sox7 and Sox18 (Figure S5B) and high conservation to hSOX17. However, the 7 aa peptide is divergent in Sox17 (Figure 5A). This poses a question of whether this divergence between Sox32 and Sox17 confers differential target specificity. To decipher this, constructs were produced where the respective 7 aa CTD regions were deleted from both Sox32 and Sox17, and hybrid constructs were also generated with 7 aa peptides switched between the two TFs.

The AlphaFold structure of Sox32 shows a 25 aa long CTD helix, while the structure of Sox17 indicates a C-terminal helix in an equivalent position, albeit with lower confidence and the Sox17 helix is shorter (Figure 2). To test whether the longer 25 aa Sox32 CTD helical peptide confers specificity for induction of Sox32 targets we also produced a construct with

it replacing the Sox17 7 aa peptide.

Based on injections of each of the resulting mRNAs at the one-cell stage, we find that induction of Sox17 targets requires the Sox17 7 aa peptide and replacement with the Sox32 7 aa fails to elicit FL-Sox17 levels of induction (Figure 5C-E). The replacement of the Sox17

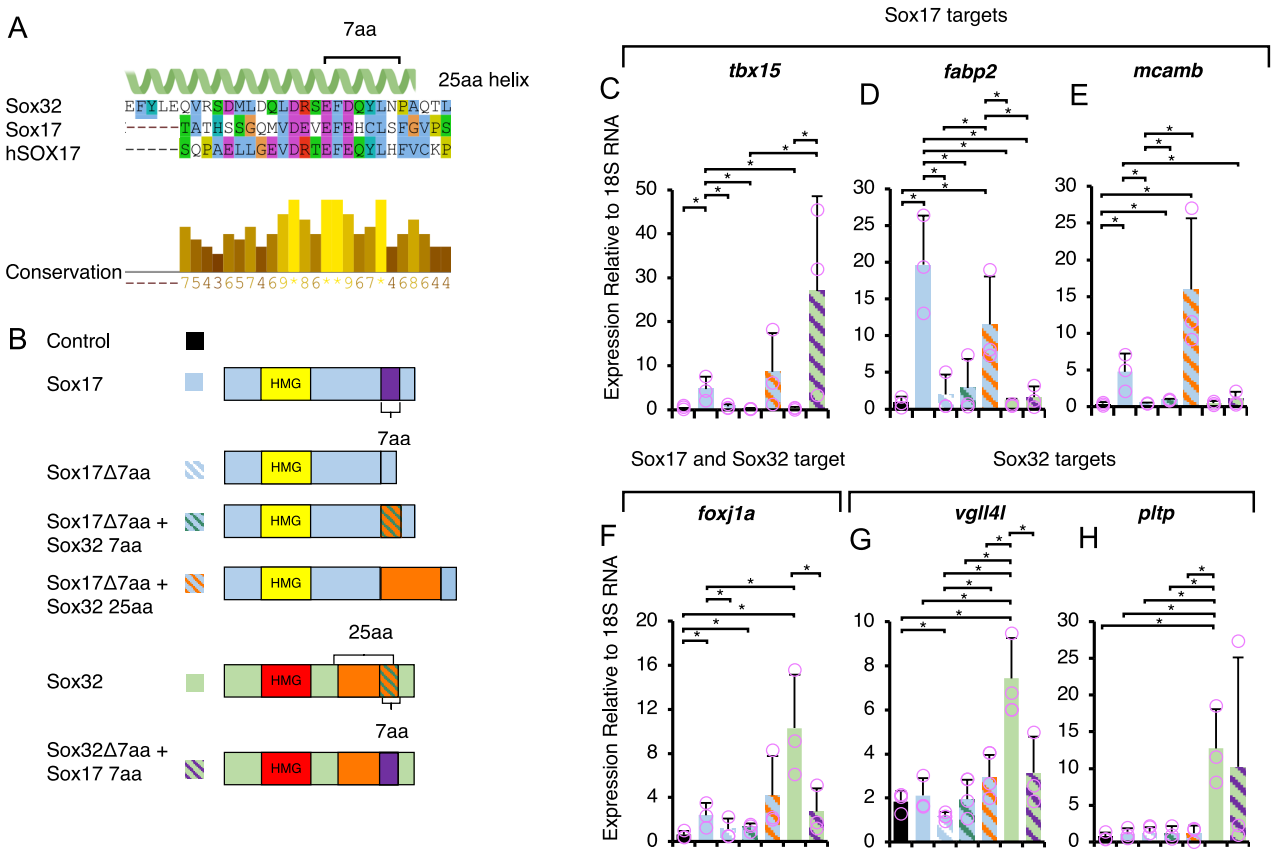

Figure 5: Divergence of a 7 aa CTD putative helical peptides confers differential target specificity of Sox32 and Sox17. A) Alignment of Sox32 CTD helix (25 aa) to Sox17 and hSOX17 with conservation indicated. Alignment carried out using Clustal Omega in Jalview and colour coded based on conserved amino acid properties. Conservation levels calculated according to Analysis of Multiply Aligned Sequences (AMAS) displayed. **B)** Protein models of Sox17 and Sox32 domain deletions and switches injected at the one-cell stage and RNA extraction followed by qPCR at 6hpf. **C-H)** qPCR of Sox17 targets (C-E), Sox17 and Sox32 targets (F) and Sox32 targets (G-H). Statistical tests on biological triplicate samples were carried out using Student's t-test. * P<0.05.

7 aa peptide with the Sox32 25 aa peptide does not allow Sox17 to induce Sox32 targets to the level of FL-Sox32 (Figure 5G-H). Therefore, while the Sox32 CTD helix is necessary for Sox32 induction of endoderm and DFC targets (Figure 4), it is not sufficient in the context of the Sox17 protein. However, replacement of the Sox17 7 aa peptide with the Sox32 25 aa peptide does result in significant induction of Sox17 targets compared to the control (Figure 5C-E). We therefore conclude that the Sox32 25 aa peptide that makes up the Sox32 CTD helix appears to be a potent transactivation domain.

In addition to replacement of the Sox17 7 aa peptide with that of Sox32 not inducing Sox17 targets to the capacity of FL-Sox17 (Figure 5C-E), we also see that replacement of the Sox32 7 aa peptide with the Sox17 7 aa peptide cannot induce *vgll4l* to the same level as FL-Sox32 (Figure 5G). This indicates that both Sox32 and Sox17 require their own intrinsic 7 aa domain for efficient induction of their respective targets. Therefore, we suggest that divergence in C-terminal peptides leads to differential Sox32 and Sox17 target specificity.

Sox17 is necessary for establishing correct brain, heart and pancreas laterality

Bilateral asymmetry occurs for many organ systems and is vital in the correct function of organs across a range of species. Heterotaxia is a syndrome whereby there is an abnormal arrangement of abdominal and thoracic organs across the left-right axis of the body, including the pancreas and heart (Aiello et al., 2007). In humans, heterotaxy affects approximately 1 in 10,000 live births and accounts for ~3% of all congenital cardiac abnormalities (Agarwal et al., 2021, Sutherland and Ware, 2009). Many patients with heterotaxy experience lifelong morbidities and potentially mortality. Consequently, understanding the molecular genetics of LR asymmetry is vital. In zebrafish, the heart is positioned to the left and loops in a dextral manner (Bakkers, 2011). Analysis of looping can be carried out by observing expression of the myocardial marker *myl7* (Yelon et al., 1999). The pancreas is normally positioned on the right, we can investigate pancreas placement by visualising expression of *insulin*, a gene expressed specifically within the endocrine pancreas (Argenton et al., 1999).

In zebrafish and the large majority of vertebrates, the epithalamus is a part of the brain which shows asymmetry (Concha and Wilson, 2001). The epithalamus is part of the dorsal diencephalic conduction system that connects the forebrain to the midbrain, and it is comprised of the pineal complex and a bilateral pair of habenular nuclei (Sutherland, 1982). Asymmetry within the habenular nuclei results in distinct patterns of gene expression and affects connectivity (Bianco and Wilson, 2008). We can study the asymmetry of the habenula nuclei by observing the expression of *kctd12.1*, which is expressed more extensively in the left habenula (Gamse et al., 2005, Gamse et al., 2003).

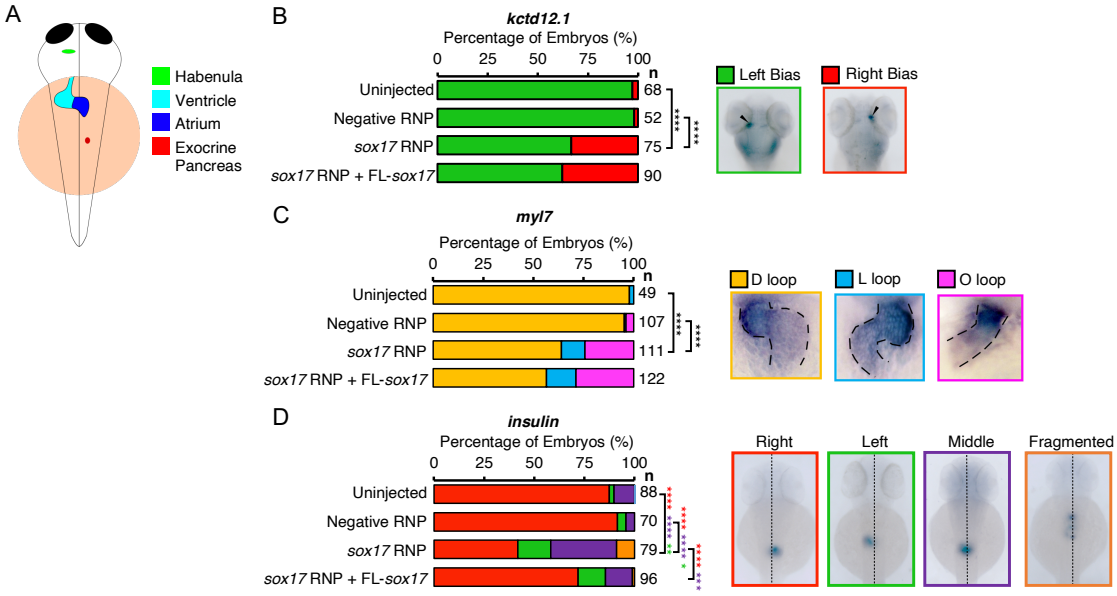

Figure 6: Sox17 CRISPants exhibit left-right asymmetry defects. A) Diagram depicting correct asymmetry of organs (4dpf). The brain habenulae show a left bias. The heart loops in a dextral (D-loop) fashion. The exocrine pancreas is located to the right of the midline. **B-D)** WISH analysing the effect of Sox17 knockdown on dorsal habenula *kctd12.1* asymmetric expression at 4dpf (B), heart looping at 48 hpf indicated by *myl7* (C), and exocrine pancreas placement at 48hpf indicated by *insulin*, and whether abnormal asymmetry can be rescued by co-injection of *sox17* exogenous mRNA along with CRISPR RNPs at the one-cell stage. Statistically significant differences in categorical scoring was inferred using Fisher's Exact test on independent biological triplicate datasets. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$. For B-C) black asterisks indicate significant difference between normal laterality – left bias for B), D-loop for C) and all other phenotypes. For D) red asterisks indicate a significant difference in right-sided pancreas *vs*. everything else, purple asterisks indicate a significant difference between middle-placed pancreas *vs*. everything else and green asterisks indicate a significant difference between left-sided pancreas and everything else. n= the number of embryos analysed.

Sox17 has previously been implicated in the establishment of LR asymmetry. Knockdown of Sox17 using an antisense morpholino (MO) was previously reported to result in abnormal pancreas placement, which could be rescued by co-injection with *sox17* mRNA to which the morpholino did not bind (Aamar and Dawid, 2010). However, an analysis of zebrafish *sox17* function in genetic mutants is lacking. Furthermore, we aim to identify key domains of Sox17 required for correct LR patterning.

To further explore the role of Sox17 in LR asymmetry we used an efficient method of G0 CRISPR (Wu et al., 2018) (Figure S6) to disrupt the *sox17* locus by injection of Cas9-sgRNA complexes at the one-cell stage, followed by analysis of LR patterning using whole mount *in situ* hybridation (WISH) to assess brain habenula asymmetry (*ktcd12.1*), heart looping (*myl7*) and pancreas placement (*insulin*). We found that Sox17 CRISPants showed abnormal brain asymmetry, heart looping and pancreas placement (Figure 6B-D). We find that Sox17 CRISPants have a significant increase in the proportion of embryos exhibiting middle and left-sided pancreas compared to the control (Figure 6D). We co-injected *sox17* mRNA into CRISPants to assess its ability to rescue the observed phenotypes. This resulted in a significant increase in embryos with normal right-sided pancreas placement. This occurred concurrently with a significant decrease in middle placement, while left-sided placement remained unaffected (Figure 6D). This indicates that the middle placement of the pancreas can be rescued by exogenous *sox17* mRNA, however left-sided pancreas, *i.e.*, total reversal cannot.

Neither brain asymmetry (Figure 6B) nor heart looping (Figure 6C) could be rescued by co-injection of exogenous *sox17* mRNA into CRISPants. To test whether exogenous Sox17 itself may be affecting LR patterning, we carried out OE studies through the injection of *sox17* mRNA across a broad range of concentrations. This indicated that Sox17 OE does not alter either brain asymmetry or heart looping (Figure S7). The lack of rescue of abnormal phenotypes is therefore not due to the effect of Sox17 OE.

We note that heart morphogenesis is partly dependent on the adjacent anterior endoderm. For example Sox32 mutants that lack endoderm exhibit *cardia bifida* (Alexander et al., 1999), a condition in which independent left and right heart fields develop due to a failure of coalescing morphogenesis. To determine whether the failure of *sox17* mRNA injection to rescue heart looping was due to loss of anterior endoderm in Sox17 CRISPants, we analysed the expression of anterior endoderm markers *foxa2* and *vgll4l.* Contrary to the reported Sox17-MO phenotype (Aamar and Dawid, 2010), we found that the anterior endoderm is intact in Sox17 CRISPants (Figure S8). Therefore, the lack of rescue of heart looping does not appear to be a secondary effect due to an endodermal phenotype.

In conclusion, we have shown that Sox17 CRISPants have abnormalities in brain asymmetry, heart looping and pancreas placement. Only abnormal middle placed pancreas can be rescued by co-injection of the *sox17* mRNA into CRISPants. The placement of the pancreas is determined by looping of the posterior foregut as a result of lateral plate mesoderm (LPM) migration, in which *sox17* is also expressed (Horne-Badovinac et al., 2003, Chung et al., 2011). Therefore, Sox17 has a role in the process of gut looping that exogenous *sox17* mRNA is sufficient to rescue, which we speculate may be linked to a function in the LPM.

Conserved C-terminal peptides are required in Sox17 for correct pancreas placement

We have confirmed that Sox17 is required for organ asymmetry in zebrafish and that we can rescue pancreas placement through co-injection of *sox17* mRNA into CRISPants. We can therefore use this system to decipher which domains of Sox17 are required for its role in this process. *Sox17* mutant constructs were created (Figure 7A), with domains for deletion selected based on high-level conservation among teleost and mammalian Sox17 orthologues (Figure S9). Four conserved regions were identified and deleted: the Sox17 NTD, and three distinct regions within the CTD of varying lengths of 35 aa, 14 aa and 20 aa

(Figure 7a). The mRNA corresponding to each mutant construct was injected into *sox17* CRISPants at the one-cell stage and their capacity to rescue correct pancreas placement was assessed.

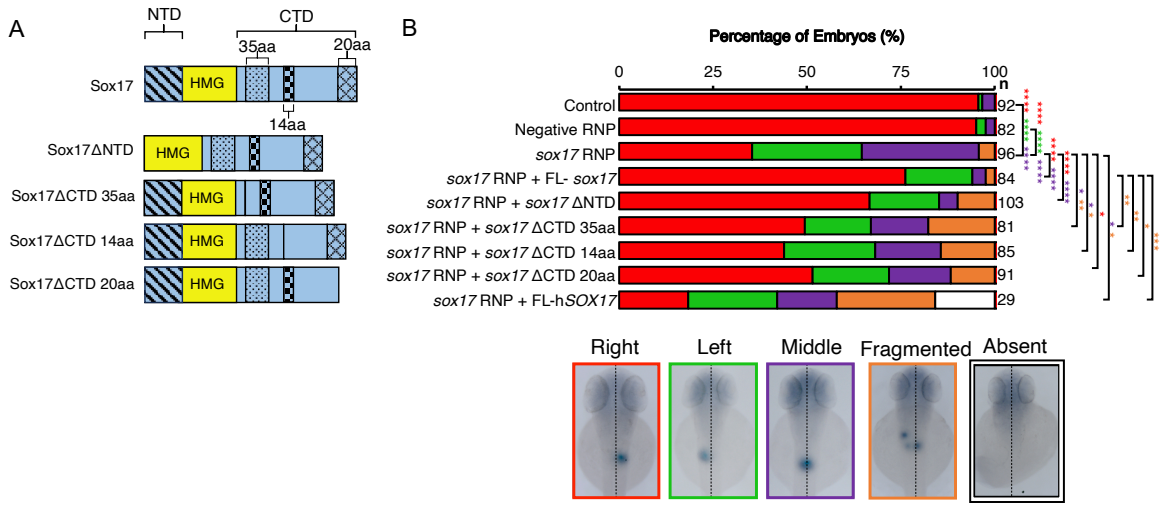

Figure 7: Sox17 CTD peptides are required for rescue of correct pancreas placement in Sox17 CRISPants. A) Diagram depicting protein models of Sox17 mutants. **B)** WISH at 48hpf using *insulin* to assess the ability of Sox17 mutants in rescuing abnormal exocrine pancreas placement by co-injection of *sox17* exogenous mRNA alongside CRISPR RNPs at the one-cell stage. Statistically significant differences in categorical scoring was inferred using Fisher's Exact test on independent biological triplicate datasets. * = P<0.05, ** = P<0.01, *** = P<0.001, **** = P<0.0001. Red asterisks indicate a significant difference in right-sided pancreas *vs.* all other phenotypes, purple asterisks indicate a significant difference between middle-placed pancreas *vs.* all other phenotypes, green asterisks indicate a significant difference between left-sided pancreas *vs.* all other phenotypes and yellow asterisks indicate a significant difference between fragmented *insulin*+ domains *vs.* all other phenotypes. FL refers to full length. n= the number of embryos analysed.

We find that Sox177ΔNTD can rescue pancreas placement in *sox17* CRISPants to a similar degree as full length (FL)-Sox17. Both constructs result in a significant increase in embryos exhibiting normal right-sided placement and a significant decrease in embryos showing middle placement, with left-sided placement unaffected compared to *sox17* CRISPants alone (Figure 7B). This indicates the N-terminal domain of Sox17 is largely dispensable for its function in pancreas placement.

Sox17ΔCTD35 aa and Sox17ΔCTD14 aa exhibit a significant decrease in embryos with a middle-placed pancreas compared to embryos injected with *sox17* RNP alone. However, this does not coincide with an increase in normal right-sided placement as seen with FL-Sox17 and Sox17ΔNTD. Instead, there is a significant increase in embryos exhibiting multiple fragmented *insulin*+ domains (Figure 7B). The Sox17ΔCTD20 aa displays a significant increase in embryos with normal right-sided placement compared to embryos injected with *sox17* RNP alone, albeit the proportion of embryos with middle placement does not change. Therefore, it is likely that this construct results in less fragmentation compared to Sox17ΔCTD35 aa and Sox17ΔCTD14 aa, rather than the rescue of pancreas placement (Figure 7B). Altogether, all three domains studied with the CTD of Sox17 are likely required for the establishment of correct pancreas placement.

Sox17$^{-/-}$ mutant mice exhibit an open body wall and failure to turn from a concave to a convex "foetal" position (Kanai-Azuma et al., 2002, Viotti et al., 2012) – characteristics of mutants with LR patterning defects (Hamada et al., 2002). Given the requirement for Sox17 in LR patterning in mammals and zebrafish, we sought to determine whether mammalian SOX17 has the capacity to induce correct pancreas placement through injection of h*SOX17* mRNA into *sox17* CRISPants. We found that hSOX17 is unable to rescue pancreas

placement. Instead, there is a further decrease in right-sided pancreas placement compared to *sox17* RNP in addition to the total absence of the pancreas (Figure 7B). The aberrant LR phenotype in mouse is attributed as a secondary effect to the depletion of gut endoderm (Viotti et al., 2012), while zebrafish Sox17 CRISPants do not exhibit loss of endoderm (Figure S7). Taken together, this indicates that while Sox17 is required for LR asymmetry in mammals and zebrafish, mechanistic variability exists.

Overall, we conclude the Sox17 CTD is required for correct pancreas placement, while the NTD is dispensable. Additionally, our results indicate that peptides in the CTD can influence formation of an intact endocrine pancreas, although whether this reflects aberrant morphogenesis or specification is unclear.

Taken together, our results reveal the molecular basis for functional divergence between Sox17 and Sox32, highlighting key functional peptides controlling both endoderm and DFC specification, and the establishment of pancreas placement during gut morphogenesis.

Discussion

Evolutionary differences between mammalian endoderm-inducer hSOX17 and Sox32 precludes SOX17 from engaging the zebrafish endoderm gene regulatory network

Our results confirm that Sox32 OE results in enrichment of the endodermal cell identity as previously suggested (Kikuchi et al., 2001). Although Sox32 can induce subsets of genes significantly overlapping those induced by Sox17 and hSOX17 (Figure 1C), hSOX17 is unable to induce endoderm specification within the zebrafish embryo (Figure S3). Meanwhile there is no significant overlap in genes induced by both Sox17 and hSOX17 (Figure 1C), suggesting they have acquired distinct biochemical properties during evolution that preclude them from performing similar functions in the early zebrafish embryo. Similarly, Sox32 has acquired biochemical properties after its duplication from the ancestral teleost *sox17* gene which render it largely distinct from both Sox17 and hSOX17, though some limited functional similarity persists in the context of our experiments. However, since Sox32 induces expression of *sox17*, it is likely that gene expression induced by overexpression of both Sox32 and Sox17 can be attributable to Sox32 induction of endogenous *sox17*.

While transcripts induced by ectopic hSOX17 expression do not show enrichment for any biological or anatomical terms, downregulated transcripts are significantly enriched for biological terms linked to the canonical Wnt pathway (Figure S2). hSOX17 contains a conserved β-catenin interaction domain within its C-terminal domain. This domain was first discovered in *Xenopus* Sox17 and shows 100% sequence identity between hSOX17, Sox32 and *Xenopus* Sox17 (Sinner, 2004). However Sox17 shows alterations to the domain (Figure 5A), rather than the complete absence as previously reported (Kikuchi et al., 2001). Additionally, co-occupation of Sox17 and β-catenin at enhancers in *Xenopus* and human pluripotent stem cells results in endoderm specification through activation of endoderm-specific genes and repression of mesoectodermal genes (Mukherjee et al., 2022, Mukherjee et al., 2020). The β-catenin domain shows 100% conservation with Sox32 but not with Sox17. Thus, it is plausible that Sox32 has adopted the function to activate endoderm-specific genes in zebrafish through a sub-functionalisation event.

Divergence of the Sox32 HMG domain is critical to allow engagement with the endoderm gene regulatory network

Consistent with previous reports (Kikuchi et al., 2001) our analyses indicate that ectopic Sox32 expression induces endodermal cell identity at the expense of other cell types, indicating its ability to change cell fate (Figure 3G). Our molecular analyses reveal that the ability of Sox32 to orchestrate the endoderm GRN is dependent on differences between the Sox32 and Sox17 HMG domains. We hypothesise this divergence likely occurred after the duplication event where *sox32* and its paralogue *sox17* emerged. Due to

this divergence in HMG domains, we suggest sub-functionalisation of ancestral *Sox17* functions occurred, whereby in zebrafish *sox32* participates in endoderm specification, while as evidenced by our RNA-seq, *sox17* is implicated in vasculogenesis (Figures S2, S4). However, since the changes in the Sox32 HMG domain do not resemble the protein sequence of mammalian SOX17, it would appear that other changes have occurred in the molecular mechanisms by which Sox32 and SOX17 participate in endoderm specification and development.

Generally, Sox32 and Sox17 are referred to as SoxF subfamily members, however Sox32 has diverged somewhat from other members as indicated by the HMG domain sharing <70% sequence identity to other SoxF subfamily members (Kikuchi et al., 2001). Published studies considering medaka *sox32* (which has a one-to-one orthologue relationship with zebrafish *sox32*) refer to it as the lone member of the SoxK subfamily (Cui et al., 2011, Wilson et al., 2010). We have shown that amino acid differences between the Sox32 and Sox17 HMG domains confer substantial functional differences (Figure 3), arguably supporting the notion of *sox32* being in a subfamily of its own.

An outstanding question is how divergence of the HMG domain confers functional differences. It is possible that the Sox32 and Sox17 HMG domains bind different *cis*-regulatory modules (CRMs), or amino acid differences in the HMG domains alter physiochemical interactions and hence activate different downstream target genes. Contact sites between mouse SOX17 and a *cis*-regulatory enhancer upstream of the *Lama1* transcriptional start site have been mapped (Palasingam et al., 2009). Comparison of contact site amino acids showed no difference to the Sox32 HMG domain (Figure 2D). This may indicate that both the Sox32 and Sox17 HMG domains are capable of engaging similar DNA sequences but participate in different protein complexes, leading to altered gene regulatory profiles.

Sox factors rarely act in isolation. Instead, they interact with binding partners to together engage CRMs and induce or repress target genes. Therefore the ability of Sox32 and Sox17 HMG domains to induce downstream genes could be governed by HMG domain-dependent selectively of partners, as previously discovered with Sox11 and Sox2 (Wiebe et al., 2003) and the capacity of the overall complex to bind CRMs. Pou5f3 is a known interactor of Sox32 in zebrafish and is required for Sox32 induction of endoderm specification, and transcriptional activation of the Sox32 target *sox17* (Zhao et al., 2013, Lunde et al., 2004, Reim et al., 2004). In mammals, Sox17 interacts with POU5f1/OCT4 – the closest homolog of zebrafish Pou5f3. Additional complexity has been discovered in mouse embryonic stem cells, whereby OCT4 partners with SOX2 to drive pluripotency gene expression, upon the increase in SOX17 expression, OCT4 will instead interact with SOX17 to drive expression of endodermal genes (Aksoy et al., 2013). SOX-OCT interacting residues within the C-terminal region of the HMG domain have been mapped (Palasingam et al., 2009). All contact amino acids are conserved between mammalian SOX17 and zebrafish Sox17. However, there are differences in these residues between zebrafish Sox32 and Sox17 (Figure 2D). One of the differences corresponds to a change in amino acid properties, with nonpolar hydrophobic methionine in Sox17 becoming polar threonine in Sox32. While not relevant to the methionine-to-threonine change, the switching of one amino acid residue in the OCT4 interaction region of SOX17 can change the co-operativity of SOX17 to that of SOX2. This one residue switch in SOX17 confers the capacity to reprogram somatic cells into pluripotent ones, as SOX2 would (Jauch et al., 2011). This indicates that a single amino acid difference in the HMG domain could change the affinity of Sox32 and Sox17 with an interacting partner and therefore guide differences in cellular programming.

Furthermore, the requirement for correct partner cooperation in target transcription has been established in human Sox TF function. SOX2 has the capacity to interact with both OCT3 and OCT1, however only the SOX2-OCT3 complex can promote transcriptional activation of the FGF4 enhancer (Yuan et al., 1995). Therefore, differences between the

Sox32 and Sox17 HMG domains may promote differential partner cooperation, resulting in transcription of different targets irrespective of whether the HMG domains can otherwise engage DNA in the same manner.

Divergence of a SoxF C-terminal motif is a key determinant of Sox17 target gene induction

We have shown that differences in the HMG domains of Sox17 and Sox32 are critical for Sox32 function. However, some genes induced by Sox32 OE are also reliant on the terminal domains of Sox32, irrespective of the HMG domain differences. This subset is induced by Sox32 and the HMG Switch synthetic construct but not by wild type Sox17, indicating their activation of target transcription is HMG domain indifferent, but reliant on Sox32 flanking regions. A similar phenomenon has been observed where the SOX1 HMG domain can replace that of SOX9 to bind and activate the δ1-crystallin DC5 enhancer, however replacement of the SOX1 CTD cannot (Kamachi et al., 1999). Therefore, HMG domains can be interchangeable and bind the same CRM, while a specific CTD is required for transcriptional activation. Furthermore, transactivation domains have been identified within the CTD of SOX9, in which truncation results in the impairment of skeletal development (Südbeck et al., 1996). Overall, this demonstrates the importance of the CTD of Sox factors.

Our qPCR indicates that Sox32 requires C-terminal peptides of 23 aa and 7 aa for transactivation of endoderm and DFC target genes (Figure 4) and Sox17 requires the homologous 7 aa peptide for the induction of its target genes (Figure 5C-E). Additionally, we show that Sox32 and Sox17 require their intrinsic 7 aa peptide for transactivation of their respective target genes (Figure 5C-E). The Sox32 and hSOX17 7 aa peptide shows high levels of conservation, and while there are amino acid differences their properties remain similar (Figure 5A). Sox17, however, exhibits multiple differences in amino acid properties compared to Sox32, suggesting these differences are likely to be functionally important. The complete Sox32 7 aa peptide and partial 23 aa peptide (18/23 aa) make up the 25 aa Sox32 C-terminal helix (Figure 2B,E). We hypothesise the Sox32 CTD helix is a potent transactivator as the replacement of the Sox17 7 aa peptide with the Sox32 CTD helix can induce Sox17 targets compared to the control (Figure 5C-E). Although AlphaFold predicts C-terminal helices for both hSOX17 and Sox17, the confidence in the prediction is low (Figure 2A,C). The importance of C-terminal helices in controlling gene transcription activity is exhibited by the FoxA factors (Pani et al., 1992). Additionally, the FoxA C-terminal helix is implicated in *in vivo* chromatin remodelling (Iwafuchi et al., 2020). The changes it exerts on the chromatin landscape occur through binding histone octamers in chromatin resulting in increasing accessibility. Therefore, FoxA factors are categorised as pioneer factors, whose activity is key in cell fate control. An outstanding question is whether Sox32/Sox17 are also able to work in a pioneering manner to change the chromatin landscape.

In addition to the requirement of Sox32 and Sox17 C-terminal 7 aa peptides for transactivation of target genes, they are also important for the role of Sox32 and Sox17 in endoderm specification and pancreas placement, respectively. We show that Sox32Δ7 aa has a reduced capacity to induce endoderm targets compared to full-length Sox32 (Figure 4). This relative gene expression data corroborates spatial expression data that shows Sox32Δ7 aa cannot induce ectopic endoderm at the same capacity as full-length Sox32 within the zebrafish embryo (Zhao et al., 2013). The ability to induce ectopic endoderm upon Sox32 OE at the one-cell stage occurs through the trans-fating of presumptive mesoderm as seen in Figure 3G, where there is an enrichment of the endodermal cell identity at the expense of the adaxial mesoderm, this is in line with previous publications (Aoki et al., 2002, Kikuchi et al., 2001). We also show that the 7 aa C-terminal peptide of Sox17 is required for induction of its targets (Figure 2C-E). This 7 aa peptide is part of the Sox17 CTD 14 aa

deletion which does not have the capacity to rescue aberrant pancreas placement in *sox17* CRISPants, unlike full-length-Sox17 (Figure 7).

Taken together this indicates that divergence of the C-terminal 7 aa peptide in Sox17 shows importance in both conferring target gene transduction specificity and functionality.

Sox17 potentially has multiple roles in establishing organ asymmetry

We have shown via an efficient G0 CRISPR approach, that Sox17 is necessary for correct brain asymmetry, heart looping and pancreas placement (Figure 6). Furthermore, we have shown that co-injection with exogenous *sox17* mRNA is sufficient to rescue abnormal pancreas placement phenotypes in Sox17 CRISPants, consistent with previous experiments with antisense morpholinos (Aamar and Dawid, 2010). However, exogenous *sox17* mRNA is not sufficient to rescue abnormal heart looping or brain asymmetry. This may suggest that Sox17 has different roles within the establishment of organ asymmetry via roles in different cell types.

The process of zebrafish heart morphogenesis and looping has been extensively studied and reviewed (Bakkers, 2011). While we have shown that an intact *sox17* locus is necessary for zebrafish heart looping, there is a lack of phenotypic rescue by exogenous *sox17* mRNA. This lack of rescue cannot be attributed to the injection of exogenous *sox17* itself (Figure S7), or as a secondary consequence of an endodermal phenotype (Figure S8). Our RNA-seq analyses show Sox17-induced genes are enriched for heart-related anatomical terms, including the endocardium and ventricular myocardium (Figure S2). Additionally, we have observed GFP+ cells in the heart within the *sox17:EGFP* line (Figure S4A, Supplementary Video 1). The endocardium is formed from vascular endothelial cells (Milgrom-Hoffman et al., 2011). Their precursors – termed angioblasts – reside in the LPM, which is *sox17*+ (Chung et al., 2011, Zhong et al., 2001). It is possible the *sox17*+ LPM cells contribute towards the formation of the endocardium. However, lineage tracing experiments would needed to confirm this. *Sox17* is expressed within the mouse endocardium where it exhibits a cell autonomous role and a secondary non-cell autonomous role in the myocardium for proper heart morphogenesis (Saba et al., 2019). Considering our finding of *sox17:EGFP*+ cells within the zebrafish heart (Figure S8), a conserved role within the endocardium controlling heart morphogenesis is not farfetched.

Sox17 is also expressed within the KV (Schneider et al., 2008, Chung et al., 2011), and morpholino knockdown of Sox17 leads to abnormality in KV morphogenesis and function via defective ciliogenesis (Aamar and Dawid, 2010). The perturbation of cilia function results in aberrant *spaw* expression within the LPM (Kramer-Zucker et al., 2005). We show that Sox17 CRISPants exhibit aberrant *spaw* expression in the LPM, and this cannot be restored to normal left-sided *spaw* expression through co-injection of exogenous *sox17* mRNA (Figure S10). This suggests that we are not able to rescue KV function with exogenous *sox17* mRNA, leading to an inability to rescue correct brain asymmetry and heart looping.

On the other hand, visceral organ laterality and therefore pancreas placement is determined by the directionality of posterior foregut looping that is dependent on the migration of the LPM. This results in exertion of a mechanical force on the endoderm to push the endoderm that is anterior to the pancreatic islet to the left of the midline (Horne-Badovinac et al., 2003). It is noted that randomisation of LPM migration occurs due to aberrant Nodal signalling within the LPM. However, given that *sox17* is expressed in the LPM, we hypothesise exogenous Sox17 is sufficient to independently rescue LPM migration downstream of aberrant Nodal signalling resulting in correct looping of the posterior foregut and therefore pancreas placement. We showed that exogenous *sox17* mRNA was able to rescue middle pancreas placement but not total reversal in *sox17* CRISPants (Figure 6), therefore we propose that Sox17 is required for LPM migration.

Overall, our data suggests Sox17 may control LR asymmetry through roles in multiple cell types. Given investigations in mouse (Saba et al., 2019), it is possible Sox17 specifically contributes to heart development through a role in the endocardium. It likely also controls LR patterning via roles in the KV and LPM (Chung et al., 2011, Schneider et al., 2008). We show exogenous *sox17* mRNA has the capacity to rescue pancreas placement but not brain asymmetry or heart looping. This may be due to exogenous *sox17* mRNA being sufficient to rescue LPM migration resulting in normal pancreas placement, but not sufficient to rescue the KV or endocardium, leading to aberrant brain asymmetry and heart looping. This phenomenon could be a result of mosaicism, whereby exogenous *sox17* mRNA enters a sufficient number of LPM cells, but not those of the KV or endocardium which are more minor cell populations.

Limitations of the study

This study exploits heterologous expression to provide insights into interspecies SoxF factor similarity, presenting potential caveats. While Sox17 and hSOX17 have both evolved from ancestral Sox17, it is possible their co-factors have undergone parallel evolution. This could result in hSOX17 being unable to interact with the zebrafish orthologues of its co-factors. It is also possible that Sox17 and hSOX17 have different effects on the epigenetic landscapes of the zebrafish embryo, thus altering the contexts in which we are assessing their function. This is similarly true of comparisons with Sox32. In spite of these caveats, our findings still indicate key biochemical differences between these related Sox factors. Our results indicate different functional capabilities between Sox17 and hSOX17 in zebrafish early development and that hSOX17 cannot induce endoderm specification in the zebrafish embryo in a manner similar to Sox32, contrary to the hypothesis arising from earlier phenotypic evidence. This indicates that Sox32 has a unique property that enbles its major role in zebrafish endoderm specification.

Conclusion

In summary, we present evidence that Sox32 shares some functional similarities with both hSOX17 and Sox17, however hSOX17 and Sox17 have different molecular capabilities and functions in early zebrafish embryos. We further show that hSOX17 does not induce endoderm specification in the zebrafish embryo as Sox32 does, and the specificity of Sox32 to participate in the zebrafish endoderm GRN emerges from divergence within its HMG domain. We also identify a divergent SoxF C-terminal motif in Sox17 that is vital in distinguishing target gene induction compared to its paralogue Sox32. Finally, through functional dissection we identify multiple conserved motifs within the C-terminal domain required for the role of Sox17 in LR asymmetry.

Our results provide novel findings into the evolution of Sox factors involved in vertebrate endoderm formation and LR asymmetry. This work contributes to understanding of Sox factor evolution and the role of Sox32 in zebrafish endoderm development, and it further dissects the potential role of Sox factors in human heterotaxy.

Experimental Procedures

Animals

AB, *Tg(sox17:EGFP)*^{ha01Tg} and *Tg(kdrl:mCherry*^{S896}*,gata1a:dsRed*^{sd2tg}*)* fish were reared as described (Westerfield, 2000, Mizoguchi et al., 2008, Traver et al., 2003, Chi et al., 2008). All zebrafish studies were conducted in compliance with the United Kingdom Animals (Scientific Procedures) Act 1986, licensed by the UK Home Office and implemented by The University of Warwick.

Confocal Microscopy

48hpf zebrafish embryos were anesthetized with 0.168%(w/v) of Ethyl 3-aminobenzoate methanesulfonate, prior to mounting in 1% lateral view agarose moulds. Moulds were generated from a 3D stamp as designed by (Kleinhans and Lecaudey, 2019). To ensure high-quality imaging, embryos were treated before 24hpf with 0.003% (w/v) 1-Phenyl-2-thiourea (PTU) to prevent pigmentation (Karlsson et al., 2001). Imaging was carried out on the Zeiss LSM 980 with Airyscan 2.

Cloning for *in vitro* production of mRNAs

pCS2+*hSOX17* was cloned from PB-TRE3G-SOX17 (Miller et al., 2018) using XhoI and XbaI restriction sites. PB-TRE3G-SOX17 was a gift from David Vereide (Addgene plasmid # 104541 ; http://n2t.net/addgene:104541 ; RRID:Addgene_104541)

Sox17 and *sox32* deletion and domain switch constructs (Table S1) were generated using pCS2+*sox17* and pCS2+*myc-sox32* by PCR amplification using Q5 polymerase (NEB) using described primers (Table S2) followed by re-circularisation or insertion using T4 ligase (Invitrogen). All constructs were generated using blunt-end cloning, except for Sox17Δ7aa + Sox32 25aa and myc-Sox32ΔHMG + Sox17 HMG where NEBuilder HiFi DNA Assembly (NEB), with Sox17Δ7aa + Sox32 25aa as a gBlock (Integrated DNA Technologies) was used.

mRNA *in vitro* transcription

Capped mRNA was synthesised from pCS2+ plasmids containing genes of interest. Linearisation with NotI was conducted followed by transcription using SP6 RNA polymerase as previously described (Talbot et al., 2022).

Ribonuclear protein (RNP) complex production

To knockdown *sox17* ribonuclear protein (RNP) complexes were generated using a pool of three guide RNAs and Cas9 protein (Integrated DNA Technologies) as described (Essner, 2016). Guide RNAs that do not target the zebrafish genome were used to generate a negative control RNP. Guide RNAs were generated from specific crRNAs (Table S3) and standard tracrRNA (Integrated DNA Technologies) as described (Essner, 2016). crRNAs against *sox17* were predesigned by Integrated DNA Technologies and selected based on predicted on-target and off-target scores from both Integrated DNA Technologies and CHOPCHOP (Labun et al., 2019).

Microinjection

Equimolar amounts of mRNA (calculated based on relative mRNA length) were injected at the one-cell stage. For interspecies RNA-seq seq each embryo was injected with the following amount: myc-*sox32* – 150pg, h*SOX17* – 141pg, *sox17* – 175pg. For HMG Switch RNA-seq each embryo was injected with the following amount: myc-*sox32* – 150pg, myc-*sox32*ΔHMG – 121.5pg, and myc-*sox32*ΔHMG+*sox17*HMG – 148.5pg. For *sox32* mutant qPCR, each embryo was injected with the following amount: myc-*sox32* – 150pg, myc-*sox32*ΔNTD – 127.5pg, myc-*sox32*ΔHMG – 121.5pg, myc-*sox32*ΔCTD23aa – 142.5pg, myc-*sox32*ΔCTD7aa – 147pg. For CTD switch qPCR, each embryo was injected with the following amount: *sox17* – 150pg, *sox17*Δ7aa – 147pg, myc-*sox32* – 128ng, myc-*sox32*Δ7aa+*sox17*-7aa – 128pg, *sox17*Δ7aa+*sox32*-7aa – 150pg, *sox17*Δ7aa+*sox32*-25aa – 155pg. For Sox17 functional dissection, each embryo was injected with the following amount of mRNA: *sox17* – 150pg, *sox17*ΔNTD – 134pg, *sox17*ΔCTD35aa – 139pg, *sox17*ΔCTD14aa – 145pg, *sox17*ΔCTD20aa – 143pg, h*SOX17* – 108pg. For overexpression experiments mRNA was injected in a volume of 1nL. To produce CRISPant embryos RNPs were injected in a volume of 3nL. For experiments involving combined

CRISPR RNP and mRNA injection RNPs and mRNAs were injected in a combined volume of 4nL.

Whole mount *in situ* hybridisation (WISH)

WISH of zebrafish embryos was carried out as described (Thisse and Thisse, 2008). Anti-sense RNA probes for *myl7* (Yelon et al., 1999), *insulin* (Milewski et al., 1998), *vgll4l (Nelson et al., 2014)* and *foxa2* (Odenthal and Nüsslein-Volhard, 1998) were synthesised as described. *Ktcd12.1* RNA probe was generated from a PCR product using primers: F - CTGCCGGACTATTTTCCAGAG, R- TAATACGACTCACTATAGGGAGCTGCACGCGACCATCT and transcribed using T7 polymerase (Promega). Categorical scoring was completed blinded.

Total RNA Extraction

Embryos were lysed at 6 hours post fertilisation (hpf) via addition of 600ul buffer RLT and DTT (2% v/v) to at least fifty embryos per condition, with initial disruption using a plastic pestle. Further disruption was carried out using QIAshredder columns (QIAGEN) and RNA was extracted using the RNeasy Mini Kit (QIAGEN) with on-column DNase treatment (QIAGEN) as per manufacturer's protocol.

qPCR

cDNA was generated from 400ng of total RNA using Random Hexamers (Promega) and Ultrascript Reverse Transcriptase (PCR Bio) as per manufacturer's protocol. qPCR was carried out using relevant primers (Table S4) and Power SYBR Green Master Mix (Thermofisher) on the Stratagene Mx3005P (Agilent Technologies). Raw Ct values were extracted using MxPro software and analysed using the $\Delta\Delta$Ct method (Livak and Schmittgen, 2001), with normalisation to *18S* rRNA (McCurley and Callard, 2008). Statistical differences were inferred using Students t-test with significance set at $P < 0.05$.

RNA-sequencing Library Preparation and Analysis

Total RNA was sent to the Cologne Center for Genomics (CCG) for library preparation (Illumina Stranded TruSeq) and sequencing on the NovaSeq 6000 with a read length of 101 bp.

Quality control of raw data was performed using FastQC. Reads were mapped to danRer11 zebrafish genome and counts per gene (Ensembl version 97) generated using STAR with default settings (Dobin et al., 2013). Differential expression was identified through adjusted P-value < 0.05 using DESeq2(1.36.0) (Love et al., 2014) in R (4.2.0). To identify whether different TFs illicit similar or different effects on the transcriptome, upregulated and downregulated differentially expressed genes (DEGs) were separately overlapped to identify whether TFs induce or repress similar genes. Statistical testing to identify whether the overlap in genes was significantly more/less than expected by chance was conducted using Fisher's exact test (phyper on R). This analysis used a gene universe size defined as all genes detected in RNA-sequencing where samples in all conditions showed counts > 0. Heatmaps were generated on median normalised counts per condition using pheatmap (1.0.12) with 'scale = row' option to display deviation in expression from the average expression of each gene. Gene Ontology (GO) and anatomy term enrichment analysis was carried out on fishEnrichR (Chen et al., 2013, Kuleshov et al., 2016) using autoRIF Z-score, with significantly enriched terms defined by an adjusted P-value < 0.05.
To infer how TFs influence cell fate, Gene Set Enrichment Analysis (GSEA) was carried out using published scRNA-sequencing data (Wagner et al., 2018) as cell identity markers using GSEA (4.3.2) (Subramanian et al., 2005, Mootha et al., 2003).

Conservation Analysis

Protein sequences were obtained from UniProt (The UniProt, 2023) (Table S5) and alignments were conducted aligned using Clustal Omega in JalView (Sievers and Higgins, 2018, Sievers et al., 2011). Jalview uses multiple sequence alignments according to Analysis of Multiply Aligned Sequences (AMAS) (Livingstone and Barton, 1993) to calculate conservation levels. Conserved sequences were compared to 3D structures obtained from AlphaFold; Sox32 (Q90Z46), Sox17 (Q5PQZ5), hSOX17 (Q9H6I2) (Jumper et al., 2021, Varadi et al., 2022).

Authors' contributions

Conceived and designed the experiments: ACN, SJ, KAP. Performed the experiments: SJ, RE. Analysed the data: SJ. Wrote the paper: SJ. Provided feedback on the paper: ACN, KAP.

Funding

This research was funded in whole or in part by the BBSRC Midlands Integrative Biosciences Training Partnership (BB/M01116X/1). This research was also funded, in part, by the Wellcome Trust through a Wellcome Seed Award in Science to ACN (210177/Z/18/Z); and by funding from the German Research Foundation (Deutsche Forschungsgemeinschaft) through SFB 680 project A12 to KAP. SJ has a PhD studentship funded by the BBSRC Midlands Integrative Biosciences Training Partnership (BB/M01116X/1). RE was funded by the MRC Doctoral Training Partnership in Interdisciplinary Biomedical Research (MR/N014294/1).

Acknowledgments

We thank Rui Monteiro for kindly gifting us the *Tg(kdrl:mCherry,gata1a:dsRed)* fish used in this study. We thank Fiona Wardle and Karuna Sampath for reagents. We further thank Karuna Sampath for valuable discussions and generous access to equipment. We also thank the Warwick zebrafish facility for zebrafish care and, Ian Hands-Portman from the Warwick School of Life Sciences for imaging advice.

References

AAMAR, E. & DAWID, I. B. 2010. Sox17 and chordin are required for formation of Kupffer's vesicle and left-right asymmetry determination in zebrafish. *Developmental dynamics : an official publication of the American Association of Anatomists,* 239**,** 2980-2988.

AGARWAL, R., VARGHESE, R., JESUDIAN, V. & MOSES, J. 2021. The heterotaxy syndrome: associated congenital heart defects and management. *Indian Journal of Thoracic and Cardiovascular Surgery,* 37**,** 67-81.

AIELLO, V. D., ANDERSON, R. H., BÉLAND, M. J., COLAN, S. D., DUCA, D. D., ELLIOTT, M. J., FRANKLIN, R. C. G., JACOBS, J. P., KROGMANN, O. N., KUROSAWA, H., MARUSZEWSKI, B., STELLIN, G., TCHERVENKOV, C. I., WALTERS, H. L., WEINBERG, P. M. & WILLIAM GAYNOR, J. 2007. The nomenclature, definition and classification of cardiac structures in the setting of heterotaxy. *Cardiology in the Young,* 17**,** 1-28.

AKSOY, I., JAUCH, R., ERAS, V., CHNG, W.-B. A., CHEN, J., DIVAKAR, U., NG, C. K. L., KOLATKAR, P. R. & STANTON, L. W. 2013. Sox Transcription Factors Require Selective Interactions with Oct4 and Specific Transactivation Functions to Mediate Reprogramming. *STEM CELLS,* 31**,** 2632-2646.

ALEXANDER, J., ROTHENBERG, M., HENRY, G. L. & STAINIER, D. Y. R. 1999. casanova Plays an early and essential role in endoderm formation in zebrafish. *Developmental Biology*.

ANGELOZZI, M. & LEFEBVRE, V. 2019. SOXopathies: Growing Family of Developmental Disorders Due to SOX Mutations. *Trends Genet,* 35**,** 658-671.

AOKI, T. O., DAVID, N. B., MINCHIOTTI, G., SAINT-ETIENNE, L., DICKMEIS, T., PERSICO, G. M., STRÄHLE, U., MOURRAIN, P. & ROSA, F. D. R. M. 2002. Molecular integration of casanova in the Nodal signalling pathway controlling endoderm formation. *Development,* 129**,** 275-286.

ARGENTON, F., ZECCHIN, E. & BORTOLUSSI, M. 1999. Early appearance of pancreatic hormone-expressing cells in the zebrafish embryo. *Mechanisms of Development,* 87**,** 217-221.

BAKKERS, J. 2011. Zebrafish as a model to study cardiac development and human cardiac disease. *Cardiovascular Research,* 91**,** 279-288.

BIANCO, I. H. & WILSON, S. W. 2008. The habenular nuclei: a conserved asymmetric relay station in the vertebrate brain. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 364**,** 1005-1020.

BOWLES, J., SCHEPERS, G. & KOOPMAN, P. 2000. Phylogeny of the SOX Family of Developmental Transcription Factors Based on Sequence and Structural Indicators. *Developmental Biology,* 227**,** 239-255.

CHEN, E. Y., TAN, C. M., KOU, Y., DUAN, Q., WANG, Z., MEIRELLES, G. V., CLARK, N. R. & MA'AYAN, A. 2013. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics,* 14**,** 128.

CHI, N. C., SHAW, R. M., DE VAL, S., KANG, G., JAN, L. Y., BLACK, B. L. & STAINIER, D. Y. R. 2008. Foxn4 directly regulates tbx2b expression and atrioventricular canal formation. *Genes & Development,* 22**,** 734-739.

CHUNG, M. I. S., MA, A. C. H., FUNG, T.-K. & LEUNG, A. Y. H. 2011. Characterization of Sry-related HMG box group F genes in zebrafish hematopoiesis. *Experimental Hematology,* 39**,** 986-998.e5.

CONCHA, M. L. & WILSON, S. W. 2001. Asymmetry in the epithalamus of vertebrates. *Journal of Anatomy,* 199**,** 63-84.

COOPER, M. S. & D'AMICO, L. A. 1996. A Cluster of Noninvoluting Endocytic Cells at the Margin of the Zebrafish Blastoderm Marks the Site of Embryonic Shield Formation. *Developmental Biology,* 180**,** 184-198.

CUI, J., SHEN, X., ZHAO, H. & NAGAHAMA, Y. 2011. Genome-Wide Analysis of Sox Genes in Medaka (Oryzias latipes) and Their Expression Pattern in Embryonic Development. *Cytogenetic and genome research,* 134**,** 283-294.

DICKMEIS, T., MOURRAIN, P., SAINT-ETIENNE, L., FISCHER, N., AANSTAD, P., CLARK, M., STRÄHLE, U. & ROSA, F. 2001. A crucial component of the endoderm formation pathway, CASANOVA, is encoded by a novel sox-related gene. *Genes & development,* 15**,** 1487-1492.

DOBIN, A., DAVIS, C. A., SCHLESINGER, F., DRENKOW, J., ZALESKI, C., JHA, S., BATUT, P., CHAISSON, M. & GINGERAS, T. R. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England),* 29**,** 15-21.

ESSNER, J. J. 2016. *RE: Zebrafish embryo microinjection: Ribonucleoprotein delivery using the Alt-RTM CRISPR-Cas9 System.*

ESSNER, J. J., AMACK, J. D., NYHOLM, M. K., HARRIS, E. B. & YOST, H. J. 2005. Kupffer's vesicle is a ciliated organ of asymmetry in the zebrafish embryo that initiates left-right development of the brain, heart and gut. *Development,* 132**,** 1247 LP-1260.

FIGIEL, D. M., ELSAYED, R. & NELSON, A. C. 2021. Investigating the molecular guts of endoderm formation using zebrafish. *Brief Funct Genomics*.

GAMSE, J. T., KUAN, Y.-S., MACURAK, M., BRÖSAMLE, C., THISSE, B., THISSE, C. & HALPERN, M. E. 2005. Directional asymmetry of the zebrafish epithalamus guides dorsoventral innervation of the midbrain target. *Development,* 132**,** 4869-4881.

GAMSE, J. T., THISSE, C., THISSE, B. & HALPERN, M. E. 2003. The parapineal mediates left-right asymmetry in the zebrafish diencephalon. *Development,* 130**,** 1059-1068.

GUBBAY, J., COLLIGNON, J., KOOPMAN, P., CAPEL, B., ECONOMOU, A., MÜNSTERBERG, A., VIVIAN, N., GOODFELLOW, P. & LOVELL-BADGE, R. 1990. A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. *Nature,* 346**,** 245-250.

HAMADA, H., MENO, C., WATANABE, D. & SAIJOH, Y. 2002. Establishment of vertebrate left–right asymmetry. *Nature Reviews Genetics,* 3**,** 103-113.

HORNE-BADOVINAC, S., REBAGLIATI, M. & STAINIER, D. Y. R. 2003. A Cellular Framework for Gut-Looping Morphogenesis in Zebrafish. *Science,* 302**,** 662-665.

IWAFUCHI, M., CUESTA, I., DONAHUE, G., TAKENAKA, N., OSIPOVICH, A. B., MAGNUSON, M. A., RODER, H., SEEHOLZER, S. H., SANTISTEBAN, P. & ZARET, K. S. 2020. Gene network transitions in embryos depend upon interactions between a pioneer transcription factor and core histones. *Nature Genetics,* 52**,** 418-427.

JAUCH, R., AKSOY, I., HUTCHINS, A. P., NG, C. K. L., TIAN, X. F., CHEN, J., PALASINGAM, P., ROBSON, P., STANTON, L. W. & KOLATKAR, P. R. 2011. Conversion of Sox17 into a Pluripotency Reprogramming Factor by Reengineering Its Association with Oct4 on DNA. *Stem Cells,* 29**,** 940-951.

JUMPER, J., EVANS, R., PRITZEL, A., GREEN, T., FIGURNOV, M., RONNEBERGER, O., TUNYASUVUNAKOOL, K., BATES, R., ŽÍDEK, A., POTAPENKO, A., BRIDGLAND, A., MEYER, C., KOHL, S. A. A., BALLARD, A. J., COWIE, A., ROMERA-PAREDES, B., NIKOLOV, S., JAIN, R., ADLER, J., BACK, T., PETERSEN, S., REIMAN, D., CLANCY, E., ZIELINSKI, M., STEINEGGER, M., PACHOLSKA, M., BERGHAMMER, T., BODENSTEIN, S., SILVER, D., VINYALS, O., SENIOR, A. W., KAVUKCUOGLU, K., KOHLI, P. & HASSABIS, D. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature,* 596**,** 583-589.

KANAI-AZUMA, M., KANAI, Y., GAD, J. M., TAJIMA, Y., TAYA, C., KUROHMARU, M., SANAI, Y., YONEKAWA, H., YAZAKI, K., TAM, P. P. L. & HAYASHI, Y. 2002. Depletion of definitive gut endoderm in Sox17-null mutant mice. *Development*.

KARLSSON, J., VON HOFSTEN, J. & OLSSON, P.-E. 2001. Generating Transparent Zebrafish: A Refined Method to Improve Detection of Gene Expression During Embryonic Development. *Marine Biotechnology,* 3**,** 522-527.

KIKUCHI, Y., AGATHON, A., ALEXANDER, J., THISSE, C., WALDRON, S., YELON, D., THISSE, B. & STAINIER, D. Y. R. 2001. casanova encodes a novel Sox-related protein necessary and sufficient for early endoderm formation in zebrafish. *Genes and Development*.

KIKUCHI, Y., TRINH, L. A., REITER, J. F., ALEXANDER, J., YELON, D. & STAINIER, D. Y. R. 2000. The zebrafish bonnie and clyde gene encodes a Mix family homeodomain protein that regulates the generation of endodermal precursors. *Genes and Development*.

KLEINHANS, D. S. & LECAUDEY, V. 2019. Standardized mounting method of (zebrafish) embryos using a 3D-printed stamp for high-content, semi-automated confocal imaging. *BMC Biotechnology,* 19**,** 68.

KRAMER-ZUCKER, A. G., OLALE, F., HAYCRAFT, C. J., YODER, B. K., SCHIER, A. F. & DRUMMOND, I. A. 2005. Cilia-driven fluid flow in the zebrafish pronephros, brain and Kupffer's vesicle is required for normal organogenesis. *Development,* 132**,** 1907 LP-1921.

KULESHOV, M. V., JONES, M. R., ROUILLARD, A. D., FERNANDEZ, N. F., DUAN, Q., WANG, Z., KOPLEV, S., JENKINS, S. L., JAGODNIK, K. M., LACHMANN, A., MCDERMOTT, M. G., MONTEIRO, C. D., GUNDERSEN, G. W. & MA'AYAN, A. 2016. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research,* 44**,** W90-W97.

LABUN, K., MONTAGUE, T. G., KRAUSE, M., TORRES CLEUREN, Y. N., TJELDNES, H. & VALEN, E. 2019. CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Research,* 47**,** W171-W174.

LILLY, A. J., LACAUD, G. & KOUSKOFF, V. 2017. SOXF transcription factors in cardiovascular development. *Semin Cell Dev Biol,* 63**,** 50-57.

LIVAK, K. J. & SCHMITTGEN, T. D. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods (San Diego, Calif.),* 25**,** 402-408.

LIVINGSTONE, C. D. & BARTON, G. J. 1993. Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation. *Computer applications in the biosciences : CABIOS,* 9**,** 745-756.

LONG, S., AHMAD, N. & REBAGLIATI, M. 2003. The zebrafish nodal-related gene southpaw is required for visceral and diencephalic left-right asymmetry. *Development (Cambridge, England),* 130**,** 2303-2316.

LOVE, M. I., HUBER, W. & ANDERS, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology,* 15**,** 550.

LUNDE, K., BELTING, H. G. & DRIEVER, W. 2004. Zebrafish pou5f1/pou2, Homolog of Mammalian Oct4, Functions in the Endoderm Specification Cascade. *Current Biology*.

MCCURLEY, A. T. & CALLARD, G. V. 2008. Characterization of housekeeping genes in zebrafish: male-female differences and effects of tissue type, developmental stage and chemical treatment. *BMC Molecular Biology,* 9**,** 102.

MCDONALD, ANGELA C. H., BIECHELE, S., ROSSANT, J. & STANFORD, WILLIAM L. 2014. Sox17-Mediated XEN Cell Conversion Identifies Dynamic Networks Controlling Cell-Fate Decisions in Embryo-Derived Stem Cells. *Cell Reports,* 9**,** 780-793.

MELBY, A. E., WARGA, R. M. & KIMMEL, C. B. 1996. Specification of cell fates at the dorsal margin of the zebrafish gastrula. *Development,* 122**,** 2225 LP-2237.

MILEWSKI, W. M., DUGUAY, S. J., CHAN, S. J. & STEINER, D. F. 1998. Conservation of PDX-1 Structure, Function, and Expression in Zebrafish*. *Endocrinology,* 139**,** 1440-1449.

MILGROM-HOFFMAN, M., HARRELSON, Z., FERRARA, N., ZELZER, E., EVANS, S. M. & TZAHOR, E. 2011. The heart endocardium is derived from vascular endothelial progenitors. *Development,* 138**,** 4777-4787.

MILLER, A. Z., SATCHIE, A., TANNENBAUM, A. P., NIHAL, A., THOMSON, J. A. & VEREIDE, D. T. 2018. Expandable Arterial Endothelial Precursors from Human CD34+ Cells Differ in Their Proclivity to Undergo an Endothelial-to-Mesenchymal Transition. *Stem Cell Reports,* 10**,** 73-86.

MIZOGUCHI, T., VERKADE, H., HEATH, J. K., KUROIWA, A. & KIKUCHI, Y. 2008. Sdf1/Cxcr4 signaling controls the dorsal migration of endodermal cells during zebrafish gastrulation. *Development,* 135**,** 2521-9.

MONTAGUE, T. G., GAGNON, J. A. & SCHIER, A. F. 2018. Conserved regulation of Nodal-mediated left-right patterning in zebrafish and mouse.

MOOTHA, V. K., LINDGREN, C. M., ERIKSSON, K.-F., SUBRAMANIAN, A., SIHAG, S., LEHAR, J., PUIGSERVER, P., CARLSSON, E., RIDDERSTRÅLE, M., LAURILA, E., HOUSTIS, N., DALY, M. J., PATTERSON, N., MESIROV, J. P., GOLUB, T. R., TAMAYO, P., SPIEGELMAN, B., LANDER, E. S., HIRSCHHORN, J. N.,

ALTSHULER, D. & GROOP, L. C. 2003. PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature Genetics,* 34**,** 267-273.
MUKHERJEE, S., CHATURVEDI, P., RANKIN, S. A., FISH, M. B., WLIZLA, M., PARAISO, K. D., MACDONALD, M., CHEN, X., WEIRAUCH, M. T., BLITZ, I. L., CHO, K. W. & ZORN, A. M. 2020. Sox17 and β-catenin co-occupy Wnt-responsive enhancers to govern the endoderm gene regulatory network. *eLife,* 9**,** e58029-e58029.
MUKHERJEE, S., LUEDEKE, D. M., MCCOY, L., IWAFUCHI, M. & ZORN, A. M. 2022. SOX transcription factors direct TCF-independent WNT/β-catenin responsive transcription to govern cell fate in human pluripotent stem cells. *Cell Reports,* 40.
NELSON, A. C., CUTTY, S. J., NIINI, M., STEMPLE, D. L., FLICEK, P., HOUART, C., BRUCE, A. E. E. & WARDLE, F. C. 2014. Global identification of smad2 and eomesodermin targets in zebrafish identifies a conserved transcriptional network in mesendoderm and a novel role for eomesodermin in repression of ectodermal gene expression. *BMC Biology*.
NG, C. K. L., LI, N. X., CHEE, S., PRABHAKAR, S., KOLATKAR, P. R. & JAUCH, R. 2012. Deciphering the Sox-Oct partner code by quantitative cooperativity measurements. *Nucleic Acids Research,* 40**,** 4933-4941.
NIAKAN , J. H. M. R., VOKES, S. A., RODOLFA, K. T., SHERWOOD, R. I., YAMAKI, M., DIMOS, J. T., CHEN, A. E., MELTON, D. A., MCMAHON, A. P. & EGGAN, K. 2010. Sox17 promotes differentiation in mouse embryonic stem cells by directly regulating extraembryonic gene expression and indirectly antagonizing self-renewal. *Genes and development,* 24**,** 312-326.
ODENTHAL, J. & NÜSSLEIN-VOLHARD, C. 1998. fork head domain genes in zebrafish. *Development Genes and Evolution,* 208**,** 245-258.
PALASINGAM, P., JAUCH, R., NG, C. K. L. & KOLATKAR, P. R. 2009. The structure of Sox17 bound to DNA reveals a conserved bending topology but selective protein interaction platforms. *Journal of molecular biology,* 388**,** 619-630.
PANI, L., OVERDIER, D. G., PORCELLA, A., QIAN, X., LAI, E. & COSTA, R. H. 1992. Hepatocyte Nuclear Factor 3β Contains Two Transcriptional Activation Domains, One of Which Is Novel and Conserved with the Drosophila Fork Head Protein. *Molecular and Cellular Biology,* 12**,** 3723-3732.
PEVNY, L. H. & LOVELL-BADGE, R. 1997. Sox genes find their feet. *Current opinion in genetics & development,* 7**,** 338-344.
POULAIN, M. 2006. Zebrafish endoderm formation is regulated by combinatorial Nodal, FGF and BMP signalling. *Development*.
REBAGLIATI, M. R., TOYAMA, R., FRICKE, C., HAFFTER, P. & DAWID, I. B. 1998. Zebrafish Nodal-Related Genes Are Implicated in Axial Patterning and Establishing Left–Right Asymmetry. *Developmental Biology,* 199**,** 261-272.
REIM, G., MIZOGUCHI, T., STAINIER, D. Y., KIKUCHI, Y. & BRAND, M. 2004. The POU domain protein Spg (Pou2/Oct4) is essential for endoderm formation in cooperation with the HMG domain protein casanova. *Developmental Cell*.
SABA, R., KITAJIMA, K., RAINBOW, L., ENGERT, S., UEMURA, M., ISHIDA, H., KOKKINOPOULOS, I., SHINTANI, Y., MIYAGAWA, S., KANAI, Y., KANAI-AZUMA, M., KOOPMAN, P., MENO, C., KENNY, J., LICKERT, H., SAGA, Y., SUZUKI, K., SAWA, Y. & YASHIRO, K. 2019. Endocardium differentiation through Sox17 expression in endocardium precursor cells regulates heart development in mice. *Scientific Reports,* 9**,** 11953.
SARKAR, A. & HOCHEDLINGER, K. 2013. The sox family of transcription factors: versatile regulators of stem and progenitor cell fate. *Cell stem cell,* 12**,** 15-30.

SCHNEIDER, I., HOUSTON, D. W., REBAGLIATI, M. R. & SLUSARSKI, D. C. 2008. Calcium fluxes in dorsal forerunner cells antagonize β-catenin and alter left-right patterning. *Development,* 135**,** 75-84.
SÉGUIN, C. A., DRAPER, J. S., NAGY, A. & ROSSANT, J. 2008. Establishment of Endoderm Progenitors by SOX Transcription Factor Expression in Human Embryonic Stem Cells. *Cell Stem Cell,* 3.
SIEVERS, F. & HIGGINS, D. G. 2018. Clustal Omega for making accurate alignments of many protein sequences. *Protein Science,* 27**,** 135-145.
SIEVERS, F., WILM, A., DINEEN, D., GIBSON, T. J., KARPLUS, K., LI, W., LOPEZ, R., MCWILLIAM, H., REMMERT, M., SÖDING, J., THOMPSON, J. D. & HIGGINS, D. G. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology,* 7**,** 539.
SINCLAIR, A. H., BERTA, P., PALMER, M. S., HAWKINS, J. R., GRIFFITHS, B. L., SMITH, M. J., FOSTER, J. W., FRISCHAUF, A. M., LOVELL-BADGE, R. & GOODFELLOW, P. N. 1990. A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature,* 346**,** 240-244.
SINNER, D. 2004. Sox17 and -catenin cooperate to regulate the transcription of endodermal genes. *Development*.
STAFFORD, D., WHITE, R. J., KINKEL, M. D., LINVILLE, A., SCHILLING, T. F. & PRINCE, V. E. 2006. Retinoids signal directly to zebrafish endoderm to specify insulin-expressing β-cells. *Development,* 133**,** 949-956.
SUBRAMANIAN, A., TAMAYO, P., MOOTHA, V. K., MUKHERJEE, S., EBERT, B. L., GILLETTE, M. A., PAULOVICH, A., POMEROY, S. L., GOLUB, T. R., LANDER, E. S. & MESIROV, J. P. 2005. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences,* 102**,** 15545-15550.
SÜDBECK, P., SCHMITZ, M. L., BAEUERLE, P. A. & SCHERER, G. 1996. Sex reversal by loss of the C–terminal transactivation domain of human SOX9. *Nature Genetics,* 13**,** 230-232.
SUTHERLAND, M. J. & WARE, S. M. 2009. Disorders of left–right asymmetry: Heterotaxy and situs inversus. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics,* 151C**,** 307-317.
SUTHERLAND, R. J. 1982. The dorsal diencephalic conduction system: A review of the anatomy and functions of the habenular complex. *Neuroscience & Biobehavioral Reviews,* 6**,** 1-13.
TALBOT, C. D., WALSH, M. D., CUTTY, S. J., ELSAYED, R., VLACHAKI, E., BRUCE, A. E. E., WARDLE, F. C. & NELSON, A. C. 2022. Eomes function is conserved between zebrafish and mouse and controls left-right organiser progenitor gene expression via interlocking feedforward loops. *Frontiers in Cell and Developmental Biology,* 10.
THE UNIPROT, C. 2023. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Research,* 51**,** D523-D531.
THISSE, B., PFLUMIO, S., FÜRTHAUER, M., LOPPIN, B., HEYER, V., DEGRAVE, A., WOEHL, R., LUX, A., STEFFAN, T., CHARBONNIER, X.Q. AND THISSE, C. (2001) 2001. Expression of the zebrafish genome during embryogenesis (NIH R01 RR15402). *ZFIN Direct Data Submission.*
THISSE, C. & THISSE, B. 2008. High-resolution in situ hybridization to whole-mount zebrafish embryos. *Nature Protocols,* 3**,** 59-69.
TRAVER, D., PAW, B. H., POSS, K. D., PENBERTHY, W. T., LIN, S. & ZON, L. I. 2003. Transplantation and in vivo imaging of multilineage engraftment in zebrafish bloodless mutants. *Nature Immunology,* 4**,** 1238-1246.

VARADI, M., ANYANGO, S., DESHPANDE, M., NAIR, S., NATASSIA, C., YORDANOVA, G., YUAN, D., STROE, O., WOOD, G., LAYDON, A., ŽÍDEK, A., GREEN, T., TUNYASUVUNAKOOL, K., PETERSEN, S., JUMPER, J., CLANCY, E., GREEN, R., VORA, A., LUTFI, M., FIGURNOV, M., COWIE, A., HOBBS, N., KOHLI, P., KLEYWEGT, G., BIRNEY, E., HASSABIS, D. & VELANKAR, S. 2022. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research,* 50**,** D439-D444.

VIOTTI, M., NIU, L., SHI, S.-H. & HADJANTONAKIS, A.-K. 2012. Role of the Gut Endoderm in Relaying Left-Right Patterning in Mice. *PLOS Biology,* 10**,** e1001276-e1001276.

VOLDOIRE, E., BRUNET, F., NAVILLE, M., VOLFF, J. N. & GALIANA, D. 2017. Expansion by whole genome duplication and evolution of the sox gene family in teleost fish. *PLoS ONE,* 12**,** 1-20.

WAGNER, D. E., WEINREB, C., COLLINS, Z. M., BRIGGS, J. A., MEGASON, S. G. & KLEIN, A. M. 2018. Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science,* 360**,** 981 LP-987.

WARGA, R. M. & KANE, D. A. 2018. Wilson cell origin for kupffer's vesicle in the zebrafish. *Developmental Dynamics,* 247**,** 1057-1069.

WEGNER, M. 2010. All purpose Sox: The many roles of Sox proteins in gene expression. *The International Journal of Biochemistry & Cell Biology,* 42**,** 381-390.

WESTERFIELD, M. 2000. The Zebrafish Book. A Guide for The Laboratory Use of Zebrafish (Danio rerio). *Eugene,* 385.

WIEBE, M. S., NOWLING, T. K. & RIZZINO, A. 2003. Identification of Novel Domains within Sox-2 and Sox-11 Involved in Autoinhibition of DNA Binding and Partnership Specificity*. *Journal of Biological Chemistry,* 278**,** 17901-17911.

WILSON, N. K., FOSTER, S. D., WANG, X., KNEZEVIC, K., SCHÜTTE, J., KAIMAKIS, P., CHILARSKA, P. M., KINSTON, S., OUWEHAND, W. H., DZIERZAK, E., PIMANDA, J. E., DE BRUIJN, M. F. T. R. & GÖTTGENS, B. 2010. Combinatorial Transcriptional Control In Blood Stem/Progenitor Cells: Genome-wide Analysis of Ten Major Transcriptional Regulators. *Cell Stem Cell,* 7**,** 532-544.

WU, R. S., LAM, I. I., CLAY, H., DUONG, D. N., DEO, R. C. & COUGHLIN, S. R. 2018. A Rapid Method for Directed Gene Knockout for Screening in G0 Zebrafish. *Developmental Cell,* 46**,** 112-125.e4.

YELON, D., HORNE, S. A. & STAINIER, D. Y. R. 1999. Restricted Expression of Cardiac Myosin Genes Reveals Regulated Aspects of Heart Tube Assembly in Zebrafish. *Developmental Biology,* 214**,** 23-37.

YUAN, H., CORBI, N., BASILICO, C. & DAILEY, L. 1995. Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3. *Genes and Development,* 9**,** 2635-2645.

ZHAO, J., LAMBERT, G., MEIJER, A. H. & ROSA, F. M. 2013. The transcription factor Vox represses endoderm development by interacting with x and Pou2. *Development*.

ZHONG, T. P., CHILDS, S., LEU, J. P. & FISHMAN, M. C. 2001. Gridlock signalling pathway fashions the first embryonic artery. *Nature,* 414**,** 216-220.