Induced in vivo transdifferentiation of vertebrate muscle into early endoderm-like cells

P. Duc Dong (ducdong@sbpdiscovery.org)
Sanford Burnham Prebys Medical Discovery Institute https://orcid.org/0000-0002-8755-2777
Joseph Lancman
Human BioMolecular Research Institute
Clyde Campbell
Iowa State University
Raquel Espin-Palazon
Iowa State University
Jonatan Matalonga
Sanford Burnham Prebys Medical Discovery Institute
Jiaye He
Morgridge Institute for Research. University of Wisconsin-Madison,
Alyssa Graves
Morgridge Institute for Research. University of Wisconsin-Madison,
Rashmi Mishra
Sanford Burnham Prebys Medical Discovery Institute
Jun Yin
Sanford Burnham Prebys Medical Discovery Institute
Chengjian Zhao
Sanford Burnham Prebys Medical Discovery Institute
Xin-Xin Zeng
Sanford Burnham Prebys Medical Discovery Institute
Jan Huisken
Morgridge Institute for Research https://orcid.org/0000-0001-7250-3756
David Traver
University of California, San Diego https://orcid.org/0000-0002-9652-7653

Article

Keywords: vertebrate muscle, in vivo transdifferentiation, cell lineage restriction

Posted Date: October 26th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-90546/v1

Additional Declarations: There is **NO** Competing Interest.

Abstract

The extent to which differentiated cells, while remaining in their native microenvironment, can be reprogrammed to assume a different identity will reveal fundamental insight into cellular plasticity and impact regenerative medicine. To investigate in vivo cell lineage potential, we leveraged the zebrafish as a practical vertebrate platform to determine factors and mechanisms necessary to induce differentiated cells of one germ layer to adopt the lineage of another. We discovered that ectopic co-expression of Sox32 and Oct4 in several non-endoderm lineages, including skeletal muscle, can specifically trigger an early endoderm genetic program in a cell-autonomous manner. Gene expression, live imaging, and functional studies reveal that the endoderm-induced muscle cells lose muscle gene expression and morphology, while specifically gaining endoderm organogenesis lineage markers, such as the pancreatic specification genes, hhex and ptf1a, via a mechanism resembling normal development. Endoderm induction by a pluripotent defective form of Oct4, endoderm markers appearing prior to loss of muscle cell morphology, and a lack of mesoderm, ectoderm, dedifferentiation, and pluripotency gene activation, together, suggests that reprogramming is endoderm specific and occurs via direct transdifferentiation. Our work demonstrates that within a living vertebrate animal, differentiated cells can be induced to directly adopt the identity of a completely unrelated cell lineage, while remaining in a distinct microenvironment, suggesting that differentiated cells in vivo may be more amenable to lineage conversion than previously appreciated. This discovery of extensive lineage potential of differentiated cells, in vivo, challenges our understanding of cell lineage restriction and may pave the way towards new in vivo sources of replacement cells for degenerative diseases such as diabetes.

Introduction

In animals, from flat worms to humans, nearly all cell lineages develop from one of three germ layers, established during the earliest stages of development. Despite several hundred millions of years of evolution, the type of specialized cells (i.e., neuronal, muscle, or pancreatic cells) that can develop from each of the distinct germ layers (ectoderm, mesoderm, or endoderm, respectively), remain well conserved among highly divergent animals, suggesting that lineage identities are restricted to specific germ layers. This paradigm is consistent with the prevailing interpretation of the Waddington's epigenetic landscape[1] model suggesting that during animal development, a cell's lineage potential becomes increasingly restricted as they differentiate. Furthermore, lineage restriction within germ layers appears to also apply to induced *in vivo* cell lineage reprogramming of vertebrate cells, as directed lineage transdifferentiation has only been shown among cell types from the same germ layer[2]. However, in the invertebrate roundworm *Caenorhabditis elegans*, it was demonstrated that two specific non-endoderm cell lineages, can be artificially converted into intestinal cells, which are normally of endoderm origin[3,4]. It remains unclear whether *in vivo* intergerm layer lineage conversion is unique to these cell types or to this and other invertebrates[5]. The ability to directly reprogram a differentiated vertebrate cell identity *in vivo*, without spatial or lineage limitations, would indicate that the developmental origin of a cell and its natural

microenvironment does not absolutely limit its potential identity, challenging the dogma of *in vivo* developmental lineage restriction and opening potential new avenues for regenerative medicine.

In vivo **platform to identify endoderm inducing factors**

To investigate whether differentiated vertebrate cells can be lineage converted *in vivo* into unrelated cell types (across germ layers and in distinct microenvironments), we leveraged the zebrafish embryo. This animal model is highly amenable to transgenic modification, and due to its rapid embryonic development, functionally differentiated muscle cells are already present at 24 hpf (hours post fertilization), as indicated by a beating heart and body movements. We aimed to induce endoderm lineages from differentiated, non-endoderm-derived cells and in tissues not closely localized to gut endoderm derived tissues. Unlike differentiated cells originating from the mesoderm and ectoderm, which are often intermingled, differentiated cells of the endoderm lineage remain largely contiguous with the gut tube, facilitating the identification of ectopically induced endoderm cells in non-endoderm derived tissues.

In contrast to previous *in vivo* reprogramming strategies to induce a specific mature subtype of endoderm cell, such as a beta cell[6], or a pluripotent intermediate, our novel strategy is to directly induce early endoderm-like cells, which are defined by their ability to give rise to a variety of endoderm lineages With this approach, we predict that the induced early-endoderm cells will have the functional potential to progress developmently towards various endoderm lineages and ultimately adopt distinct endoderm cell fates, such as pancreatic progenitors. Based on our understanding of normal endoderm development, zygotes were injected with DNA expression constructs containing candidate genes implicated in endoderm specification, including Foxa2, Foxa3, Gata5, Oct4, Sox17, and Sox32[7]. These transgenes are driven by the heat shock inducible *hsp70l* promoter[8], allowing for temporally controlled, transient, and spatially mosaic expression in differentiated cells throughout the entire body of the fish (Fig. 1A-A”). With this approach, transgene expression levels will vary extensively between cells due to variable transgene dose. Therefore, this *in vivo* platform allows for highly diverse conditions of transgene expression, in a variety of different cell types, and at different states of differentiation, thereby enhancing the probability of discovering factors with lineage conversion potential.

Forced coexpression of Sox32 and Oct4 cell-autonomously induces early endoderm genes in mesoderm and ectoderm cell types

With heat-shock induction of transgenes beginning at 24 hpf, we found that expression of any factor alone failed to significantly induce ectopic expression of *sox17* (*sox17:GFP*), the earliest definitive endoderm marker in fish and mammals[9] (e.g. zebrafish *sox32* (*hsp70l:H2BmCherry-P2A-sox32*) or mouse *Oct4* (*pou5f3*; *hsp70l:mCherryCAAX-P2A-Oct4*); Fig. 1B-C”)[10–12]. However, co-injection of both sox32 and mouse Oct4 constructs resulted in obvious ectopic *sox17:GFP* expression in approximately 50% of animals by 48 hpf, in multiple regions throughout the embryo, including the myotomes, epidermis, and spinal cord (Fig. 1D-D”). Because forced coexpression zebrafish Oct4 with zebrafish Sox32 inefficiently induced *sox17* expression, mouse Oct4 was exclusively used for the remainder of this study.

Upon closer examination of sox17:GFP-positive cells within the myotomes, we found that most have multiple nuclei and an elongated and striated cell morphology (see Fig. 1D'). These cells also coexpressed the transgene reporters for both *sox32* (nuclear mCherry) and *Oct4* (membrane mCherry), indicating that differentiated skeletal muscle cells can be cell-autonomously induced by Sox32 and Oct4 to express *sox17*. Because *sox17* is also normally expressed in a sub-population of endothelial cells during development, we examined endothelial markers in muscle cells of the myotome co-expressing *Oct4* and *sox32* and found no ectopic induction of *fli1a:GFP* or *kdrl:GFP* (Supplement Fig. 1A-D'), suggesting that the ectopically induced *sox17* expression does not indicate endothelial identity, but more likely endoderm. Further, forced coexpression of both *sox32* and *Oct4* can also cell-autonomously induce ectopic *foxa3* (*foxa3:GFP*), a gene normally expressed in a broad subset of the endoderm (Fig. 1E-E'). Consistent with normal endoderm development, we find that ectopic *foxa3:GFP* expression appeared later than *sox17*. These results indicate that Sox32 and Oct4 can induce the earliest genetic markers of the endoderm lineage.

To increase the number of cells expressing both Oct4 and Sox32, the coding sequence for each factor, in addition to mCherry (or H2B::mCherry), was placed within the same polycistronic construct, downstream of the *hsp70l* promoter (*hsp70:Oct4-P2A-mCherry-P2A-sox32*). Using this combined expression construct, we find approximately 70% of animals dispay ectopic sox17:GFP expression with 30% of these animals exhibiting 1–3 sox17:GFP expressing muscle cells and 40% showing between 3 and 20 cells (Supplement Fig. 2A-A"). Moreover, we consistently found ectopic *sox17:GFP* in p63 expressing cells (epidermal) and Elavl3/4 expressing cells (neuronal) suggesting that cells originating from the surface ectoderm and neural ectoderm, respectively, can also be induced by Sox32 and Oct4 to express *sox17* (Supplement Fig. 2B-C"). These findings demonstrate that sox17:GFP can be induced in cells derived from mesoderm, surface ectoderm, and neural ectoderm.

Next, we sought to broadly investigate the cell types that could be induced to up-regulate endoderm associated markers. We performed single cell RNA-seq on FACs sorted H2B::mCherry + cells isolated from injected 48–50 hpf zebrafish larvae (Supplement Fig. 3A). The Seurat (v. 3.1) package for R was used for sequence data analysis with standard quality control (QC) parameters applied to exclude low quality cells and dublets[13]. Cells passing QC were then clustered and annotated by label transfer using the zebrafish single cell atlas to assign cell type and germ layer identity[14] (Supplement Fig. 3B, D). 15,344 cells from animals expressing *Oct4, H2B::mCherry*, and *sox32* and 8,592 cells from control animals expressing only *H2B::mCherry* passed QC and were used for down stream analysis. Despite a low representation of the endoderm population compared to those of ectoderm and mesoderm, we found a significant increase in the number of cells annotated as the endoderm lineage (Supplement Fig. 3C; Fischer's exact test: $p < 3.34e\text{-}6$). This finding is consistent with our whole mount expression studies showing the induction of endoderm identity markers. By comparing the number of experimental and control cells expressing any of the endoderm markers established by endoderm developmental genetics studies and particularly from a recently reported mouse endoderm development sc-RNA seq data[15], we were able to address the issue of drop-outs, which is especially problematic for lower expressing endoderm genes. We examined endoderm

gene expression in groups of cells clustered into 9 general cell types and found that a greater number of cells expressed endoderm marker genes in clusters labelled as muscle cell/muscle cell progenitors, macrophages, neural crest as well as those classified as endoderm (Supplement Fig. 3D-F; Fisher's Exact $p < 5.45e\text{-}05$). For example, experimental cells clustered as muscle/muscle progenitor have significantly more cells expressing endoderm related transcripts compared to control (*foxi1*, BH Adjusted Fisher's exact $p < 0.01$; *fabp1b.1*, BH Adjusted Fisher's exact $p < 0.002$; Supplement Fig. 3F). Consistent with our observations shown above, these findings suggest that endoderm factors can be induced in cells originating from non-endoderm populations. Importantly, the cells expressing endoderm genes that were identified in non-endoderm clusters were not annotated as endoderm likely because the reprogramming was incomplete at the time of analysis. Conversely, cells whose identity was more completely reprogrammed would not cluster with their original cell types, but with the endoderm cluster instead, consistent with the increased proportion of the endoderm cluster. However, drop-outs, as well as low level and transient gene expression, limit the ablility to track their transdifferentiation progress using pseudotime analyis.

Differentiated skeletal muscle cells rapidly converted into early endoderm-like cells

Given that genetic and cellular changes induced by ectopic expression of Oct4 and sox32 were consistently observed by both whole mount and scRNAseq expression data, we focused our analysis on skeletal muscle cells in the trunk of the animals, allowing us to genetically and visually track a specific differentiated cell type during reprogramming. Differentiated skeletal muscle cells can be identified based on their elongated shape, striations, and multiple, evenly spaced nuclei, and are found in a consistent, parallel pattern spanning individual myotomes. The appearance of ectopic *sox17*:GFP and *foxa3*:GFP expression in muscle cells with these characteristics suggests that differentiated cells are indeed being induced to activate expression of these early endoderm genes. Importantly, it also suggests that a complete loss of muscle morphology and cell division is not a prerequisite for induction of endoderm genes. However, a subset of skeletal muscle cells expressing *Oct4* and *sox32* transgenes do exhibit a variable loss of muscle morphology, including partial or complete loss of striation and elongated rectangular cell shape, and adopting a more stellate shape (Fig. 1E and J). Interestingly, the nuclei in reporgrammed muscle cells appear to aggregate together towards the center of the cell (Supplement Fig. 4), reminiscent of striated muscle regeneration[16]. Live imaging from 48 to 72 hpf, using light sheet microscopy, shows highly dynamic changes in cell morphology (Supplement Movie 1). In particular, some muscle cells appear to be separating into two individual cells, potentially becoming mononucleated (Fig. 1I-I"; white and yellow arrows, Supplement Movie 2). Further, we observed that cells with induced *sox17*:GFP form highly dynamically cellular extensions protruding in multiple directions, similar to the behavior of normal *sox17* expressing endoderm cells during gastrulation (Supplement Fig. 5)[17]. Intriguingly, live imaging revealed that transgenic mCherry expression can be rapidly cleared, suggesting an active process for protein turnover in cells that are being reprogrammed (Fig. 1J-J"; double arrows, Supplement Movie 3). Consistent with the rapid clearance of transgenic mCherry protein, endogenous Myosin proteins (and transcripts), which are normally abundantly expressed, can become undetectable in

most muscle cells with induced *sox17*:GFP expression (Supplement Fig. 6A-B"). Rapid turnover of structural proteins such as Myosin would explain the rapid loss of the skeletal muscle morphology, as these cells are dense with structural proteins necessary for their shape and striations.

Loss of muscle identity and specific induction the early endoderm genetic program

To investigate whether induced muscle cells with *sox17*:GFP expression can function like early endoderm cells and progress in a development-like manner, we examined later markers of endoderm regionalization and organogenesis. Normal, endogenous endoderm *sox17* mRNA expression is temporally restricted to a short 3-hour time window before the end of gastrulation (~ 7–10 hpf) and is therefore not coexpressed with later endoderm genes[18] such as *foxa2*, *hhex*, and *ptf1a*. Consistent with the early and brief duration of *sox17* transcript expression observed during normal endoderm development, the ectopic *sox17* mRNA expression in the myotomes was limited to within the initial several hours following heat shock induction of the *Oct4* and *sox32* transgenes (Supplement Fig. 7). However, perdurance of GFP allows for tracking of *sox17*:GFP-positive cells beyond the loss of ectopic *sox17* transcript expression. By 48 hpf, a subset of the induced *sox17*:GFP-positive cells in the myotome can be found to coexpress transcripts of the foregut endoderm gene *foxa2* and pancreas specification genes, *hhex* and *ptf1a* (Fig. 1F-H), demonstrating that these induced cells can function like early endoderm cells by differentiating into later, more defined endoderm lineages. In contrast, transcripts specific to muscle such as *myh7* (myosin heavy chain 7) mRNA are lost from most *sox17*:GFP positive muscle cells (Supplement Fig. 6A-B"). These findings suggest that ectopic Oct4 and sox32 expression can inhibit muscle specific transcripts while triggering the endoderm developmental genetic program.

To express transgenes in a specific type of differentiating skeletal muscle cell, the *mylpfa* (*myosin light chain, phosphorylatable, fast skeletal muscle a*) promoter[19] was used to drive Oct4 and Sox32 expression. With a construct containing the *mylpfa* promoter driving mCherry alone, we confirmed that the activity of this promoter is almost exclusively restricted to fast muscle cells in injected embryos (Fig. 2A) and does not induce expression of endoderm genes. With *Oct4*, *sox32*, and *mCherry* all driven by this lineage specific promoter (*mylpfa:Oct4-P2A-mCherry-P2A-sox32*), we found ectopic *sox17*:GFP and *foxa3*:GFP induced only in the fast skeletal muscle layer of the myotomes (Fig. 2B, C), but not in other tissues. Flow cytometry analysis of mCherry positive cells from these injected embryos at 48 hpf reveal that approximately 30% of muscle cells expressing *sox32/Oct4* also ectopically express *sox17*:GFP (Fig. 2D). Given that *sox17*:GFP positive cells continue to arise after 48 hpf and that mCherry expression can also be rapidly lost (see Fig. 1J-J", Supplement Movie 1), the actual level of efficiency is likely higher – although efficiency does vary among individual embryos (see Supplement Fig. 1). In contrast, under the same experimental conditions, no significant induction of endothelial marker *fli1a*:GFP or neural marker *elavl3*:GFP was detected, consistent with specific induction of early endoderm transcripts (Fig. 2D) and suggesting that endoderm is specifically induced.

To more broadly assess transcriptional changes induced in reprogrammed muscle cells, we pooled sorted mCherry-positive cells from fish injected with either the *mylpfa* promoter driving *mCherry* alone (control)

or together with *Oct4* and *sox32*, and carried out qPCR analysis. Consistent with our whole mount histological studies using fluorescent reporter lines and *in situ* hybridization, early and late endoderm genes, *foxa3*, *hnf1ba*, *hnf4a*, and *ptf1a* mRNA expression are all significantly upregulated in sorted muscle cells with *sox32* and *Oct4* transgene expression (Fig. 2E, Supplement Fig. 7). These results suggest that in fast muscle cells, the endoderm specification genes *sox32* and *Oct4* can induce early endoderm-like cells that are functional in their ability to proceed developmentally down multiple genetic programs of various endoderm lineages. We also examined markers of other germ layers to determine the specificity of the endoderm induction. Mesoderm genes, *tbxta* (*T*; *brachyury*) and *meox1*, surface ectoderm gene *p63*, and neural ectoderm genes, *sox1a*, *zic2.2*, and others, are not upregulated, suggesting neither mesoderm or ectoderm lineages are induced (Fig. 2E). However, we found that the muscle genes *myod*, *myhz2*, *tnnt3a*, and *mylpfa* are downregulated, consistent with the loss of muscle cell morphology and myosin protein/mRNA expression described above. Together, these gene expression studies indicate that forced expression of sox32 and Oct4 in fast skeletal muscle cells leads to a loss of the muscle genetic program and to a specific induction of early endoderm-like cells.

Induced endoderm cells proceed through a developmental mechanism for pancreatic lineage commitment

To explore the mechanism by which the induced endoderm (iEndo) cells proceed to commit to a specific endoderm organogenesis lineage, we focused on reprogrammed muscle cells with ectopic expression of the pancreas specification gene *ptf1a*. Misexpression of *sox32* and *Oct4* can lead to muscle cells with *ptf1a*:GFP by 48–72 hpf in about one in five injected embryos. These *ptf1a*:GFP + cells appear to have only one nuclie and continue to express GFP up to at least 10 dpf, suggesting a persistent lineage conversion. iEndo muscle cells with ectopic *ptf1a*:GFP are most often found in the posterior third of the fish. The low frequency and spatial propensity of induced *ptf1a*:GFP expression led us to posit that only certain iEndo cells with optimal intrinsic (intracellular) and extrinsic (extracellular/microenvironment) conditions will proceed toward a specific endoderm lineage pathway such as liver, intestine, or pancreas. This influence by the microenvironment on pancreas lineage commitment is analogous to the regional specificity observed for induced *in vivo* transdifferentiation of astrocytes into dopaminergic neurons[20]. The skeletal muscle microenvironment has previously been demonstrated to be permissive for human pancreas cell differentiation and function[21–24]. Because we already observed that neuronal genes are not induced by Oct4 and Sox32 (shown above), the ectopic *ptf1a* expression observed is unlikely to represent the normal cerebellar or retinal domains of *ptf1a* expression. Furthermore, because iEndo muscle cells with ectopic *ptf1a* expression can be found to coexpress *sox17*:GFP (see Fig. 1H), they are more likely to be pancreatic endoderm.

The temporal sequence of endoderm genes expressed in iEndo cells (*sox17* intially, followed by *foxa2* and *hhex*; Fig. 1F-H) appears to recapitulate that of endogenous pancreatic endoderm lineage commitment. This observation led us to functionally assess whether induction of ectopic *ptf1a* expression occurred via a genetic mechanism also required for endogenous pancreas development. FGF signaling was shown to be necessary for foregut endoderm cells to adopt ventral pancreatic endoderm

lineage in normal zebrafish development and in human stem cells differentiation models[25,26]. Further, FGF signaling is highest posteriorly where ectopic *ptf1a*:GFP-positive muscle cells are most often observed[27]. Compared to DMSO treated controls (Fig. 3C-C”), blocking FGF receptor tyrosine kinase signaling with the inhibitor SU5402 prevents endogenous foregut endoderm (compare Fig. 3B”, D”, inset), as well as iEndo muscle cells (Fig. 3D-D”), from expressing *ptf1a*:GFP, suggesting that FGF signaling is required for both normal and induced pancreatic lineages. Note that SU5402 had no obvious effect on neural ptf1a:GFP expression (compare Fig. 3”A with 3B” and 3D”, white arrows), consistent with a pancreas specific loss of *ptf1a*:GFP. Moreover, inhibition of FGF signaling does not prevent ectopic *sox17*:GFP (Fig. 3E-F”), showing that induction of early endoderm cells was not hindered by loss of FGF signaling. These findings suggest that iEndo cells do not spontaneously express *ptf1a*, but rather proceed through a step-wise genetic program resembling normal, Fgf signaling dependent, pancreas lineage commitment.

In vivo **induced endoderm is independent of a pluripotency mechanism**

Examples of *in vivo* cell lineage plasticity, including natural transdifferentiation in worms, fin regeneration in zebrafish, and maintenance of the neural crest lineage potential in frogs, have implicated a pluripotency mechanism[28–30]. These findings led us to functional assess whether our induction of muscle into endoderm lineage conversion using Oct4 and Sox32 also involves a pluripotency mechanism. The lack of induction of the skeletal muscle progenitor genes *pax3* and *pax7* (Fig. 2G), suggests that lineage conversion does not involve a dedifferentiation mechanism. The appearance of endoderm gene expression prior to loss of muscle morphology suggests that cell division is not required for iEndo cells. Together with a lack of mesoderm and ectoderm gene activation, we posit that lineage conversion of muscle to endoderm does not involve a pluripotent intermediate. Consistently, qPCR expression analysis does not show upregulation of standard pluripotent mRNA markers, including endogenous *oct4*, *myca*, *vasa*, or *nanog* (Fig. 2E). However, it may be possible that Oct4 functions as a pluripotent factor without requiring the iEndo cells to have gone through a detectable pluripotent intermediate. Mammalian Oct4, in combination with other Yamanaka factors, was previously shown to be able to reprogram fish cells in culture to pluripotency[31]. To assess the requirement of Oct4’s pluripotency function for reprogramming muscle into endoderm *in vivo*, we used a modified form of mouse Oct4 which has an amino acid substitution in the linker domain, Oct4(L80A), previously shown to be transcriptionally active but unable to induce pluripotency (Fig. 4A)[32]. As with wild-type Oct4, mis-expression of mutant Oct4(L80A) with Sox32 (*hsp70:Oct4(L80A)-P2A-mCherry-P2A-sox32*) can induce muscle cells to express *sox17*:GFP, as well as lead to cell shape changes (Fig. 4B-B”). Moreover, like wild-type Oct4 iEndo cells, these *Oct4(L80A)* iEndo cells can also proceed to exhibit *ptf1a*:GFP expression and lose myosin expression (Fig. 4C-E”,Supplemental Fig. 6F-F”), suggesting that they can progress towards a pancreas lineage genetic program, while rapidly losing muscle protein expression. These findings demonstrate that a robust pluripotency mechanism via Oct4 is not required for induction of muscle cells into early endoderm-like cells, further supporting our conclusion that *in vivo* lineage conversion across germ layers can be induced directly, independent of a pluripotency mechanism. This finding also

functionally demonstrates that the role of Oct4 in endoderm specification may be distinct from its well recognized role in pluripotency.

Discussion

Within a vertebrate embryo, we demonstrate that differentiated mesoderm-derived skeletal muscle cells can be induced by ectopic expression of just two transcription factors, Sox32 and Oct4, to cell-autonomously trigger an early endoderm-like developmental program. These early endoderm-induced cells can rapidly lose muscle cell morphology and gene expression while progressing through an endoderm lineage program resembling endogenous endoderm development. This lineage conversion process appears to be direct as no evidence was found to support a pluripotent or dedifferentiated intermediate state. As with normal endogenous pancreatic *ptf1a* expression, expression of *ptf1a* in iEndo muscle cells also requires Fgf signaling. Expression of *ptf1a*, in addition to other endoderm organogenesis specification factors, shows that these iEndo cells can function to give rise to distinct endoderm lineage genetic programs. Expression of organogenesis lineage markers also suggests the potential for these iEndo to be further coaxed, with additional factors, into a specific functional differentiated endoderm organ cell type such as pancreatic beta-cells, which would be useful for diabetics.

Waddington's epigenetic landscape model suggests that as cells differentiate during normal development, they become more restricted from adopting other lineage identities. However, requiring only a few transcription (intrinsic) factors, transdifferentiation across germ layers using *in vitro* approaches has proven to be surprisingly easier than the Waddington model would predict, challenging this paradigm of limited lineage potential of differentiated cells[33,34]. Yet, with *in vitro* lineage reprogramming, removal of the cells from their native microenvironment and exposing them to artificial (extrinsic) culture conditions may compromise their lineage stability, facilitating their reprogramming. Importantly, Waddington's model addresses cell lineage constraint in the context of an embryo, where a cell’s normal microenvironment may also be restricting its lineage potential. Although *in vivo* lineage reprogramming during development in fruit flies suggests great plasticity, those efforts were restricted to converting undifferentiated cells within a germ layer [35,36]. Our work, using zebrafish embryos, demonstrates that despite the differentiated muscle cells remaining in their native microenvironment, ectopic expression of only two transcription factors is able to repress muscle identity while triggering the early endoderm genetic program, suggesting that both intrinsic and extrinsic factors maintaining muscle cell identity can be surmounted to induce conversion towards an unrelated lineage identity, as previously predicted[37]. Because we observe that ectopic *sox17* expression is also induced by Oct4 and Sox32 in other differentiated, non-muscle, cell types in other regions of the embryo, it is likely that other cell lineages, in other distinct microenvironments, are also amenable to induced *in vivo* lineage conversion.

Analogous to the commonly used MEFs (mouse embryonic fibroblasts) for *in vitro* lineage reprogramming studies, differentiated cells within the zebrafish embryo were used here as a practical vertebrate *in vivo* discovery platform to identify lineage reprogramming factors and investigate their

mechanisms. This *in vivo* platform allows for transient and mosaic expression of transgenes, at greatly varying levels, in a large number of cell types within a large number of animals, thereby increasing the likelihood of discovering combinations of factors capable of inducing lineage conversion. Testing reprogramming factors on differentiated cells throughout the embryonic fish has key advantages. The rapid development and transparency of the zebrafish embryo together with transgenic fluorescent lineage reporter lines allows for quick assessment of candidate reprogramming factors in live animals. In contrast to MEFs, the particular cell lineage reprogrammed, and its specific microenvironment, can be definitively identified, allowing for the assessment of how intrinsic and extrinsic factors influence the lineage conversion process. But similar to MEFs, differentiated cells at developmental stages are presumably more amenable to cell lineage reprogramming, providing a sensitized background for revealing cell lineage reprogramming factors that would otherwise be difficult to uncover. Additional factors, including epigenetic modifying small molecules, may be necessary for reprogramming more mature or aged differentiated cells, as previously shown in adult mice[38].

Although we find that there is on average a high efficiency (30%) of endoderm induction within pooled samples (Fig. 2D), the variability of reprogramming efficiency we observed among individual embryos indicates great potential in the number of differentiated cells that are amenable to transdifferentiation. Conversely, the high variability of reprogramming efficiency also indicates that there are factors and conditions yet be uncovered that may allow for more consistent and efficient induction of lineage conversion. Uncovering these factors to enhance lineage reprogramming will undoubtedly lead to further mechanistic insight into direct lineage conversion, both *in vivo* and *in vitro*. Testing additional reprogramming factors may also allow for induction of a distinct endoderm lineage such as pancreatic beta cells, which will potentially have biomedical applications. Transplantation of replacement beta cells into various locations in the body[39], including in skeletal muscles in humans[23,24] suggests that beta-cells can function outside their normal microenvironment to help maintain blood glucose levels. The ability to induce cells outside the foregut to adopt endoderm-like identity, as demonstrated here, is a significant step towards ultimately generating replacement beta cells from potentially any cell type, directly in the body of diabetics. Direct *in vivo* lineage conversion to generate replacement cells[40,41] may bypass safety[42] and efficacy risks associated with transplantation of *in vitro* engineered pluripotent cells[43–47]. An unrestricted ability for directly reprogramming any differentiated vertebrate cells *in vivo* into any cell type, in any microenvironment, would greatly expand the potential therapeutic applications of direct, induced *in vivo* lineage conversion for regenerative medicine.

Materials And Methods

Animal husbandry:

Adult zebrafish and embryos were cared for and maintained under standard conditions. All research activity involving zebrafish was reviewed and approved by SBP Medical Discovery Institute Institutional Animal Care and use Committee (IACUC: UO1 DK105554; Expires 01/18/2021) in accordance with Public Health Policy regarding care and use of laboratory animals and comply with the Guide for the Use of

laboratory Animals and the regulations set forth in the Animal Welfare Act and other applicable federal, state and local laws, regulations and policies. The following transgenic lines were used: *Tg(sox17:GFP)^{s870}* [48], *Tg(gut:GFP)^{s854}* [49], *Tg(ptf1a:GFP)^{jh1}* [50], *Tg(elavl3:GFP^{knu3}* [51]; *Tg(kdrl:EGFP)^{s843}* [52], *Tg(fli1a:EGFP)^{y1}* [53].

Transient Injections and Heat shock:

For all experiments, 0.5-1.0nl of plasmid was injected at the 1-cell stage to deliver the following amounts of plasmid DNA:

hsp:H2B::mCherry-P2A-sox32 and *hsp:mCherryCAAX-P2A-Oct4* (20ng)

hsp:Oct4-P2A-(H2B)mCherry-sox32: (30-35ng)

hsp: (H2B)mCherry: (30ng)

mylpfa:Oct4-P2A-mCherry-sox32:(30-35ng)

mylpfa:mCherry: (30ng)

Antibody Staining: Antibody staining was performed as previously described [54] using the following antibodies and staining reagents: (see Supplement Materials and Methods). Samples were imaged on a Zeiss LSM710 running Zen 2010 (Black). Final image processing was performed using Image-J (vs. 1.48b) and/or Photoshop CS6.

Whole mount In situ hybridization: *In situ* hybridization was performed as previously described[55]. Samples were imaged on a Leica M165FC using Leica Application Suite v. 3.8 or on a Zeiss Confocal LSM710 running Zen 2010 (Black). Adobe Photoshop CS6 and/or ImageJ64 (vs.1.48b) was used for final image processing. In situ probe plasmid references or primers?

Fluorescent Whole mount In situ hybridization: *Fluorescent in situ hybridization* combined with immunofluorescence was performed according to published protocols[56-58] with minor modifications. Probe against *vmhc* was previously described (ZDB-GENE-99123-5) [59,60]. Following hybridization, embryos were incubated with both anti-digoxigenin-POD (Roche) and chick anti-GFP antibody. *In situ* probe was first detected with TSA Plus fluorescein (Perkin Elmer) followed by incubation with AlexaFluor 594 donkey anti-chick secondary antibody (Invitrogen) to detect *sox17:GFP* localization. Stained embryos were mounted in SlowFade Gold anti-fade reagent with DAPI (Molecular Probes) prior to imaging on a Zeiss Confocal LSM710 running Zen 2010 (Black).

Heat Shock: 5 cycles of heat shock were performed from 24-48 hpf. (One cycle:3 hours +heat shock at 38.5°C, 2.5 hours at 28°C)

Live Imaging:

Microscope setup: Zebrafish embryos were imaged with a custom-built light sheet microscope [61]. A laser engine (Toptica MLE, 488nm,561nm) was used as the excitation light source. The fluorescence signal was captured by a water dipping objective (OLYMPUS, UMPLFLN 20XW, 20X/0.5) placed perpendicular to the light sheet direction. Collected signal was filtered (Chroma ET BP525/50, ET LP575) prior to image acquisition.

Sample embedding: Zebrafish embryos were de-chorionated at 48 hpf and transferred into a low-melting agarose (0.6%) solution prepared with E3 medium and tricaine (Maintained at 37°C). Using a syringe and needle, samples were drawn into a cleaned FEP tube (inner diameter: 0.8mm; wall thickness: 0.4mm; Bola) and the bottom of the tube plugged with solidified 2% agarose for additional support during imaging. The plugged FEP tube was mounted on the stage assembly so that the sample was positioned vertically at the intersection of illumination and detection optical paths of the light sheet microscope. The sample chamber temperature was maintained at 28°C using a custom-built perfusion based temperature control system.

Time-lapse acquisition: Samples were moved along the detection axis while being illuminated by the excitation light sheet. Images were taken every 2 microns with a high speed sCMOS camera (Zyla 4.2 PLUS, Andor) at 100 frames per second. The sample was imaged every minute and different excitation wavelengths were imaged sequentially. A typical z-stack of around 300 images was required to cover the region of interest. The overall length of the time lapse recording was approximately 16 hours, resulting in 960 individual 3D stacks in each channel.

Fluorescent Activated Cell Sorting: To isolate single cells, fluorescence activated cell sorting (FACS) was performed on wild type embryos injected at the 1-cell stage with either *mylpfa:mCherry* or *mylpfa:Oct4-P2A-mCherry-P2A-sox32*. At 24 hpf, injected embryos were placed in 0.003% phenylthiourea (PTU) to inhibit melanocyte formation and prevent pigmentation. At 48 hpf, embryos were inspected and only healthy/developmentally normal embryos were manually dechorionated. Following collection in 1.5ml eppi tubes, pooled embryos were washed in 1xPBS, incubated in 1xPBS(+Mg/Ca;)+ 0.05 mg/ml Liberase TM (Roche) at 37°C for 60 min and triturated with a P1000 pipette. The resulting suspension was filtered with a 30µm Celltrics cell strainer (Sysmex), spun down (300g for 10min at 4°C) and resuspended in ice cold 1xPBS +0.9% FBS(Gibco). SYTOX Red (Thermo Fischer Scientific) was added at 1:1000 to exclude dead cells immediately prior to sorting for either qPCR analysis or quantification of induction efficiency.

Single cell RNA-seq library preparation and sequencing: FACs for scRNA-Seq was performed at the SBP Medical Discovery Institute FACs core with slight modifications. Embryos were injected with *hsp:Oct4-P2A-H2B::mCherry-sox32* (30-35ng) or hsp:H2B::mCherry (30-35ng). At 50 hpf, experimental and control embryos were pooled separately and dissociated by incubating with FACSmax Cell Dissociation Solution (Genlantis) at 28°C for 15-20 minutes with gentle trituration every 5 minutes. Sorted cells were captured in FACS buffer and concentrated to recommended levels suitable for use with the 10X Chromium platform. In collaboration with the SBP Medical Disocovery Institute Genomics Core, cDNA libraries were generated with the Chromium Next GEM Single Cell 3' Reagent Kits (v. 3.1) according to manufacturer's protocol.

cDNA quantity and quality was confirmed using a Bioanalyzer(Agilent). Following library construction, samples were pooled and sequenced on a full Illumnia Novaseq SP Flowcell and sequenced 28x91, yielding an average of 1,752 (*Oct4-P2A-H2B::mCherry-sox32*) and 2,129 (*H2B::mCherry*) reads per cell. Sequencing was performed on an Illumina Novaseq 6000 in collaboration with the La Jolla Institute for Immunology Sequencing Core.

sc-RNA-Seq data analysis:

Bioinformatics analysis was performed in the Sanford Burnham Prebys Bioinformatics Core. Raw 10X single cell sequencing reads were processed and aligned against the zebrafish genome (GRCz11) using Cell Ranger software pipeline version 3.1.0 (10x Genomics). The resulting sequencing data from Oct4/H2B::mCherry/sox32 treated cells and H2B::mCherry treated cells were merged using the Cell Ranger aggr command and the merged count matrix was analyzed using Seurat v3 (Butler et al., 2018). Gene features having at least one read in more than three cells were used in the analysis. Quality control to eliminate low quality cells and doublets was performed using standard cut offs: > 200 and < 5000 gene features with at least one read, <100,000 total read counts. The resulting count matrix was then analyzed using Principal Component Analysis (PCA) resulting in 105 principal components based on a Jackstraw significance of $p<0.01$. For cell clustering, Seurat FindNeighbors and FindClusters functions were applied to the 105 principal components with a resolution of 0.5 using the default Louvain algorithm. 32 cell clusters were identified from this clustering analysis. Dimension reduction and cell clustering visualization was performed with the UMAP algorithm. Cell cluster biomarkers were examined using the Seurat FindMarkers function. To annotate and label cells from the 32 identified clusters, we compared our single cell sequencing data to the zebrafish single cell atlas[14]. Seurat FindTransferAnchors and TransferData functions were used to integrate our single cell sequencing data set with the zebrafish single cell atlas and transfer cell annotation labels. The identified 32 cell clusters were annotated based on the cell types defined by the transferred cell labels[62]. For analysis, these 32 clusters were manually merged to 9 based on cell identity labels. A two sided Fisher's exact test was performed to test the proportion of cells originating from different germ layers or different cell clusters. To overcome dropouts of low expressing endoderm genes, we utilized known endoderm factors and also extracted endoderm biomarkers from endoderm single-cell literature[15] to identify any potential endoderm cells in non-endoderm clusters from Oct4/H2B::mCherry/sox32 treated cells and H2B::mCherry treated cells. We then performed two sided Fisher's exact test on cells expressing any of the known endoderm biomarkers to estimate the difference in the proportion of endoderm cells from the two cell populations.

Quantitative PCR: For qPCR experiments, cells were sorted with a FACSAria II (BD Biosciences). Following FACS, sorted cells were homogenized with a QIAshredder (QIAGEN)and total RNA extracted using the RNeasy Mini Kit (Qiagen). cDNA was synthesized using i-Script Supermix (Bio Rad) and qPCR performed using iQ SYBR Green Supermix (Bio Rad) according to manufacturer protocols. Samples were loaded on a 384 well plate and analyzed on an ABI 7900HT (Applied Biosystems). Primers to detect zebrafish transcripts were designed and are described in Supplement Material and Methods. Relative expression levels of genes were calculated by the following formula: relative expression = 2−(Ct[gene of interest]

−Ct[housekeeping gene]). To test statistical significance, the non-parametric Mann-Whitney test was performed (2-tailed, 95% confidence interval).

Reporter Induction Efficiency: To analyze reporter induction efficiency, *sox17:GFP*, *elav3:GFP,* or *fli1a:GFP* embryos were injected with either *mylpfa:mCherry* or *mylpfa:Oct4-P2A-mCherry-P2A-sox32*. Upregualtion of fluorescent reporters was quantified by performing FACS using a BD LSRFortessa X-20 (BD Biosciences). Data was analyzed with FlowJo V-10 (FlowJo, LLC). Percent induction efficiency was determined using the following formula: (total number of GFP+ and mCherry+ Cells)/(total number of mCherry+ cells) x 100. Error was determined using the percentage of GFP+ and mCherry+ cells obtained from each respective reporter line injected with *mylpfa:mCherry*.

SU5402 treatment: 3μM SU5402 (sc-204308; Santa Cruz Biotechnology) was added to egg water with 1%DMSO and incubated 24-48 hpf.

Expression Constructs: Cloning details and sequence available upon request for the following: *Hsp:Oct4-P2A-mCherryCAAX;cryaa:dsRed, Hsp:pou5f3-P2A-mCherryCAAX;cryaa:dsRed, Hsp70:H2B::mCherry-P2A-Cas32; cmlc2:dsRed, Hsp:(mouse)FoxA3-P2A-mCherryCAAX; cryaa:dsRed, Hsp:(zebrafish) FoxA2-P2A-mCherryCAAX; cryaa:dsRed, Hsp70:H2B::mCherry-P2A-(zebrafish) Gata5; cmlc2:dsRed, Hsp70:Oct4-P2A-mCherry-P2A-sox32; cryaa:dsRed, Hsp70:Oct4-P2A-H2B::mCherry-P2A-sox32; cryaa, Hsp70:H2B::mCherry-P2A-sox32; cryaa, Hsp70:Oct4(L80A)-P2A-mCherry-P2A-sox32; cryaa:dsRed, Hsp70:mCherry; cryaa:dsRed, mylpfa:Oct4-P2A-mCherry-P2A-sox32; cryaa, mylpfa:mCherry; cryaa.* Oct4 was cloned from FUW-OSKM, a gift from Rudolf Jaenisch (Addgene plasmid # 20328 ; http://n2t.net/addgene:20328 ; RRID:Addgene_20328) [63].

Statistics: Statistics and graphs not related to scRNA-Seq analysis were generated using Prism software (vs.8)

Declarations

Data Availability: The scRNA-Seq datasets generated in the current study are available at GEO (Submission in process).

Author Contributions:

P.D.S.D. conceptualized project, designed initial experiments, and oversaw all studies. P.D.S.D., J.J.L., and C.C. designed experiments. J.J.L. generated all constructs. C.C., J.J.L., R.E.P., J.M., R.M., C.Z. and XX.I.Z. performed experiments and analyzed the data with P.D.S.D. A.G. prepared live samples for light-sheet imaging. J.Y., P.D.S.D., J.J.L analyzed scRNA-Seq data. J. He built the light-sheet microscope and performed time-lapse image acquisition and processing. J. Huisken oversaw live light-sheet imaging experiments. D.T. oversaw studies by R.E.P.

J.J.L. and P.D.S.D. prepared the figures. P.D.S.D. and J.J.L. wrote the manuscript with coauthors. All authors reviewed and contributed to editing the manuscript.

Competing Interests: The authors declare no competing interests.

Acknowledgements:

We dedicate this work to the memory of Dr. John F. Fallon whose friendship, mentorship, scientific rigor and passion have set standards to which we aspire. We thank the zebrafish research community for the numerous reagents shared and members of our laboratory for critical discussions. We thank Jesus Olvera and Cody Fine in the UCSD Human Embryonic Stem Cell Core Facility for technical assistance with flow cytometry experiments and supported by funds from the CIRM Major Facilities Grant (FA1-00607) to the Sanford Consortium for Regenerative Medicine. We thank the La Jolla Institute for Immunology Sequencing Core for sc-RNA sequencing (S10OD025052). We thank Drs. Alexandre Colas, Gregg Duester and Alessandra Sacco for their critical reading of the manuscript and apologize to authors whose work we were not able to cite due to space constraints. This work was supported by funds from the W. M. Keck Foundation (2017-01) and National Institutes of Health (NIH) R01DK098092-04 and U01DK105541 to P.D.S.D., Diabetes Research Connection (DRC) Project #08 to J.J.L, and the Larry L. Hillblom Foundation Fellowship (#2016-D-008-FEL) to J.M.

References

1 Waddington, C. H. The Strategy of the Genes. *London: Allen and Unwin* (1957).

2 de Lazaro, I. & Kostarelos, K. Engineering Cell Fate for Tissue Regeneration by In Vivo Transdifferentiation. *Stem Cell Rev* **12**, 129-139, doi:10.1007/s12015-015-9624-6 (2016).

3 Riddle, M. R. *et al.* Transorganogenesis and transdifferentiation in C. elegans are dependent on differentiated cell identity. *Dev Biol* **420**, 136-147, doi:10.1016/j.ydbio.2016.09.020 (2016).

4 Riddle, M. R., Weintraub, A., Nguyen, K. C., Hall, D. H. & Rothman, J. H. Transdifferentiation and remodeling of post-embryonic C. elegans cells by a single transcription factor. *Development* **140**, 4844-4849, doi:10.1242/dev.103010 (2013).

5 Kalacheva, N. V., Eliseikina, M. G., Frolova, L. T. & Dolmatov, I. Y. Regeneration of the digestive system in the crinoid Himerometra robustipinna occurs by transdifferentiation of neurosecretory-like cells. *PLoS One* **12**, e0182001, doi:10.1371/journal.pone.0182001 (2017).

6 Ariyachet, C. *et al.* Reprogrammed Stomach Tissue as a Renewable Source of Functional beta Cells for Blood Glucose Regulation. *Cell Stem Cell* **18**, 410-421, doi:10.1016/j.stem.2016.01.003 (2016).

7 Zorn, A. M. & Wells, J. M. Vertebrate endoderm development and organ formation. *Annu Rev Cell Dev Biol* **25**, 221-251, doi:10.1146/annurev.cellbio.042308.113344 (2009).

8 Halloran, M. C. *et al.* Laser-induced gene expression in specific cells of transgenic zebrafish. *Development* **127**, 1953-1960 (2000).

9 Sakaguchi, T., Kikuchi, Y., Kuroiwa, A., Takeda, H. & Stainier, D. Y. The yolk syncytial layer regulates myocardial migration by influencing extracellular matrix assembly in zebrafish. *Development* **133**, 4063-4072, doi:10.1242/dev.02581 (2006).

10 Dickmeis, T. *et al.* A crucial component of the endoderm formation pathway, CASANOVA, is encoded by a novel sox-related gene. *Genes Dev* **15**, 1487-1492, doi:10.1101/gad.196901 (2001).

11 Kikuchi, Y. *et al.* casanova encodes a novel Sox-related protein necessary and sufficient for early endoderm formation in zebrafish. *Genes Dev* **15**, 1493-1505, doi:10.1101/gad.892301 (2001).

12 Reim, G., Mizoguchi, T., Stainier, D. Y., Kikuchi, Y. & Brand, M. The POU domain protein spg (pou2/Oct4) is essential for endoderm formation in cooperation with the HMG domain protein casanova. *Dev Cell* **6**, 91-101 (2004).

13 Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* **36**, 411-420, doi:10.1038/nbt.4096 (2018).

14 Farnsworth, D. R., Saunders, L. M. & Miller, A. C. A single-cell transcriptome atlas for zebrafish development. *Dev Biol* **459**, 100-108, doi:10.1016/j.ydbio.2019.11.008 (2020).

15 Nowotschin, S. *et al.* The emergent landscape of the mouse gut endoderm at single-cell resolution. *Nature* **569**, 361-367, doi:10.1038/s41586-019-1127-1 (2019).

16 Mazzotti, A. L. & Coletti, D. The Need for a Consensus on the Locution "Central Nuclei" in Striated Muscle Myopathies. *Front Physiol* **7**, 577, doi:10.3389/fphys.2016.00577 (2016).

17 Schmid, B. *et al.* High-speed panoramic light-sheet microscopy reveals global endodermal cell dynamics. *Nature communications* **4**, 2207, doi:10.1038/ncomms3207 (2013).

18 Alexander, J. & Stainier, D. Y. A molecular pathway leading to endoderm formation in zebrafish. *Curr Biol* **9**, 1147-1157, doi:10.1016/S0960-9822(00)80016-0 (1999).

19 Xu, Y. *et al.* Fast skeletal muscle-specific expression of a zebrafish myosin light chain 2 gene and characterization of its promoter by direct injection into skeletal muscle. *DNA Cell Biol* **18**, 85-95, doi:10.1089/104454999315655 (1999).

20 Qian, H. *et al.* Reversing a model of Parkinson's disease with in situ converted nigral neurons. *Nature* **582**, 550-556, doi:10.1038/s41586-020-2388-4 (2020).

21 Axen, K. V. & Pi-Sunyer, F. X. Long-term reversal of streptozotocin-induced diabetes in rats by intramuscular islet implantation. *Transplantation* **31**, 439-441 (1981).

22 Capito, C. *et al.* Mouse muscle as an ectopic permissive site for human pancreatic development. *Diabetes* **62**, 3479-3487, doi:10.2337/db13-0554 (2013).

23 Christoffersson, G. *et al.* Clinical and experimental pancreatic islet transplantation to striated muscle: establishment of a vascular system similar to that in native islets. *Diabetes* **59**, 2569-2578, doi:10.2337/db10-0205 (2010).

24 Rafael, E. *et al.* Intramuscular autotransplantation of pancreatic islets in a 7-year-old child: a 2-year follow-up. *American journal of transplantation : official journal of the American Society of Transplantation and the American Society of Transplant Surgeons* **8**, 458-462, doi:10.1111/j.1600-6143.2007.02060.x (2008).

25 Manfroid, I. *et al.* Reciprocal endoderm-mesoderm interactions mediated by fgf24 and fgf10 govern pancreas development. *Development* **134**, 4011-4021 (2007).

26 Mfopou, J. K., Chen, B., Mateizel, I., Sermon, K. & Bouwens, L. Noggin, retinoids, and fibroblast growth factor regulate hepatic or pancreatic fate of human embryonic stem cells. *Gastroenterology* **138**, 2233-2245, 2245 e2231-2214, doi:S0016-5085(10)00329-X [pii]

10.1053/j.gastro.2010.02.056 (2010).

27 Sawada, A. *et al.* Fgf/MAPK signalling is a crucial positional cue in somite boundary formation. *Development* **128**, 4873-4880 (2001).

28 Buitrago-Delgado, E., Nordin, K., Rao, A., Geary, L. & LaBonne, C. NEURODEVELOPMENT. Shared regulatory programs suggest retention of blastula-stage potential in neural crest cells. *Science* **348**, 1332-1335, doi:10.1126/science.aaa3655 (2015).

29 Christen, B., Robles, V., Raya, M., Paramonov, I. & Izpisua Belmonte, J. C. Regeneration and reprogramming compared. *BMC Biol* **8**, 5, doi:10.1186/1741-7007-8-5 (2010).

30 Kagias, K., Ahier, A., Fischer, N. & Jarriault, S. Members of the NODE (Nanog and Oct4-associated deacetylase) complex and SOX-2 promote the initiation of a natural cellular reprogramming event in vivo. *Proc Natl Acad Sci U S A* **109**, 6596-6601, doi:10.1073/pnas.1117031109 (2012).

31 Rossello, R. A. *et al.* Mammalian genes induce partially reprogrammed pluripotent stem cells in non-mammalian vertebrate and invertebrate species. *Elife* **2**, e00036, doi:10.7554/eLife.00036 (2013).

32 Esch, D. *et al.* A unique Oct4 interface is crucial for reprogramming to pluripotency. *Nat Cell Biol* **15**, 295-301, doi:10.1038/ncb2680 (2013).

33 Blau, H. M. *et al.* Plasticity of the differentiated state. *Science* **230**, 758-766, doi:10.1126/science.2414846 (1985).

34 Ladewig, J., Koch, P. & Brustle, O. Leveling Waddington: the emergence of direct programming and the loss of cell fate hierarchies. *Nat Rev Mol Cell Biol* **14**, 225-236 (2013).

35 Dong, P. D., Chu, J. & Panganiban, G. Coexpression of the homeobox genes Distal-less and homothorax determines Drosophila antennal identity. *Development* **127**, 209-216 (2000).

36 Halder, G., Callaerts, P. & Gehring, W. J. Induction of ectopic eyes by targeted expression of the eyeless gene in Drosophila. *Science* **267**, 1788-1792 (1995).

37 Blau, H. M. & Baltimore, D. Differentiation requires continuous regulation. *J Cell Biol* **112**, 781-783 (1991).

38 Jorstad, N. L. *et al.* Stimulation of functional neuronal regeneration from Muller glia in adult mice. *Nature* **548**, 103-107, doi:10.1038/nature23283 (2017).

39 Cantarelli, E. & Piemonti, L. Alternative transplantation sites for pancreatic islet grafts. *Current diabetes reports* **11**, 364-374, doi:10.1007/s11892-011-0216-9 (2011).

40 Heinrich, C., Spagnoli, F. M. & Berninger, B. In vivo reprogramming for tissue repair. *Nat Cell Biol* **17**, 204-211, doi:10.1038/ncb3108 (2015).

41 Smith, D. K. & Zhang, C. L. Regeneration through reprogramming adult cell identity in vivo. *Am J Pathol* **185**, 2619-2628, doi:10.1016/j.ajpath.2015.02.025 (2015).

42 Ohnishi, K. *et al.* Premature termination of reprogramming in vivo leads to cancer development through altered epigenetic regulation. *Cell* **156**, 663-677, doi:10.1016/j.cell.2014.01.005 (2014).

43 Heslop, J. A. *et al.* Concise review: workshop review: understanding and assessing the risks of stem cell-based therapies. *Stem cells translational medicine* **4**, 389-400, doi:10.5966/sctm.2014-0110 (2015).

44 Johannesson, B. *et al.* Comparable frequencies of coding mutations and loss of imprinting in human pluripotent cells derived by nuclear transfer and defined factors. *Cell Stem Cell* **15**, 634-642, doi:10.1016/j.stem.2014.10.002 (2014).

45 Kelaini, S., Cochrane, A. & Margariti, A. Direct reprogramming of adult cells: avoiding the pluripotent state. *Stem Cells Cloning* **7**, 19-29, doi:10.2147/SCCAA.S38006 (2014).

46 Narva, E. *et al.* High-resolution DNA analysis of human embryonic stem cell lines reveals culture-induced copy number changes and loss of heterozygosity. *Nat Biotechnol* **28**, 371-377, doi:10.1038/nbt.1615 (2010).

47 Pera, M. F. Stem cells: The dark side of induced pluripotency. *Nature* **471**, 46-47, doi:471046a [pii]

10.1038/471046a (2011).

48 Mizoguchi, T., Verkade, H., Heath, J. K., Kuroiwa, A. & Kikuchi, Y. Sdf1/Cxcr4 signaling controls the dorsal migration of endodermal cells during zebrafish gastrulation. *Development* **135**, 2521-2529, doi:dev.020107 [pii]

10.1242/dev.020107 (2008).

49 Field, H. A., Ober, E. A., Roeser, T. & Stainier, D. Y. Formation of the digestive system in zebrafish. I. Liver morphogenesis. *Dev Biol* **253**, 279-290 (2003).

50 Godinho, L. *et al.* Targeting of amacrine cell neurites to appropriate synaptic laminae in the developing zebrafish retina. *Development* **132**, 5069-5079 (2005).

51 Park, H. C. *et al.* Analysis of upstream elements in the HuC promoter leads to the establishment of transgenic zebrafish with fluorescent neurons. *Dev Biol* **227**, 279-293, doi:10.1006/dbio.2000.9898 (2000).

52 Jin, S. W., Beis, D., Mitchell, T., Chen, J. N. & Stainier, D. Y. Cellular and molecular analyses of vascular tube and lumen formation in zebrafish. *Development* **132**, 5199-5209, doi:10.1242/dev.02087 (2005).

53 Roman, B. L. *et al.* Disruption of acvrl1 increases endothelial cell number in zebrafish cranial vessels. *Development* **129**, 3009-3019 (2002).

54 Lancman, J. J. *et al.* Specification of hepatopancreas progenitors in zebrafish by hnf1ba and wnt2bb. *Development* **140**, 2669-2679, doi:10.1242/dev.090993 (2013).

55 Clements, W. K. & Kimelman, D. LZIC regulates neuronal survival during zebrafish development. *Dev Biol* **283**, 322-334, doi:10.1016/j.ydbio.2005.04.026 (2005).

56 Brend, T. & Holley, S. A. Zebrafish whole mount high-resolution double fluorescent in situ hybridization. *J Vis Exp*, doi:10.3791/1229 (2009).

57 Schoenebeck, J. J., Keegan, B. R. & Yelon, D. Vessel and blood specification override cardiac potential in anterior mesoderm. *Dev Cell* **13**, 254-267, doi:10.1016/j.devcel.2007.05.012 (2007).

58 Zhou, Y. *et al.* Latent TGF-beta binding protein 3 identifies a second heart field in zebrafish. *Nature* **474**, 645-648, doi:10.1038/nature10094 (2011).

59 Yelon, D. *et al.* The bHLH transcription factor hand2 plays parallel roles in zebrafish heart and pectoral fin development. *Development* **127**, 2573-2582 (2000).

60 Zeng, X. X. & Yelon, D. Cadm4 restricts the production of cardiac outflow tract progenitor cells. *Cell Rep* **7**, 951-960, doi:10.1016/j.celrep.2014.04.013 (2014).

61 Huisken, J. & Stainier, D. Y. Even fluorescence excitation by multidirectional selective plane illumination microscopy (mSPIM). *Opt Lett* **32**, 2608-2610, doi:10.1364/ol.32.002608 (2007).

62 Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902 e1821, doi:10.1016/j.cell.2019.05.031 (2019).

63 Carey, B. W. *et al.* Reprogramming of murine and human somatic cells using a single polycistronic vector. *Proc Natl Acad Sci U S A* **106**, 157-162, doi:10.1073/pnas.0811426106 (2009).

Figures

Figure 1: In vivo identification of factors that induce expression of early endoderm factors.

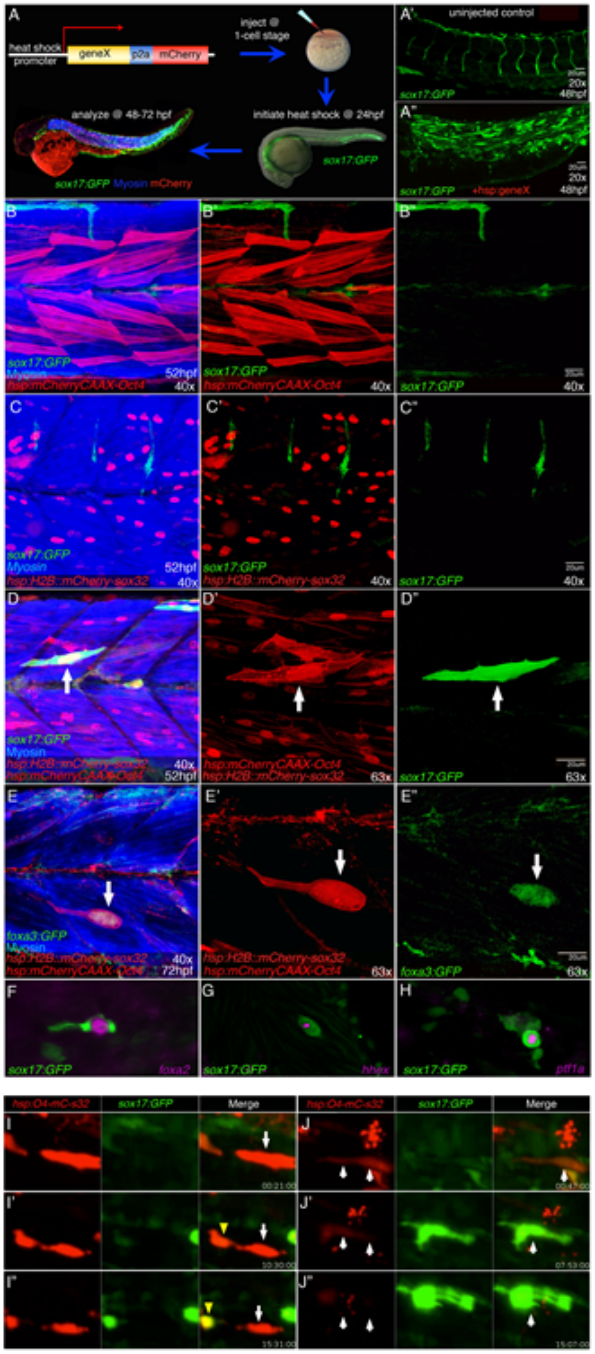

Figure 1

In vivo identification of factors that induce expression of early endoderm factors.

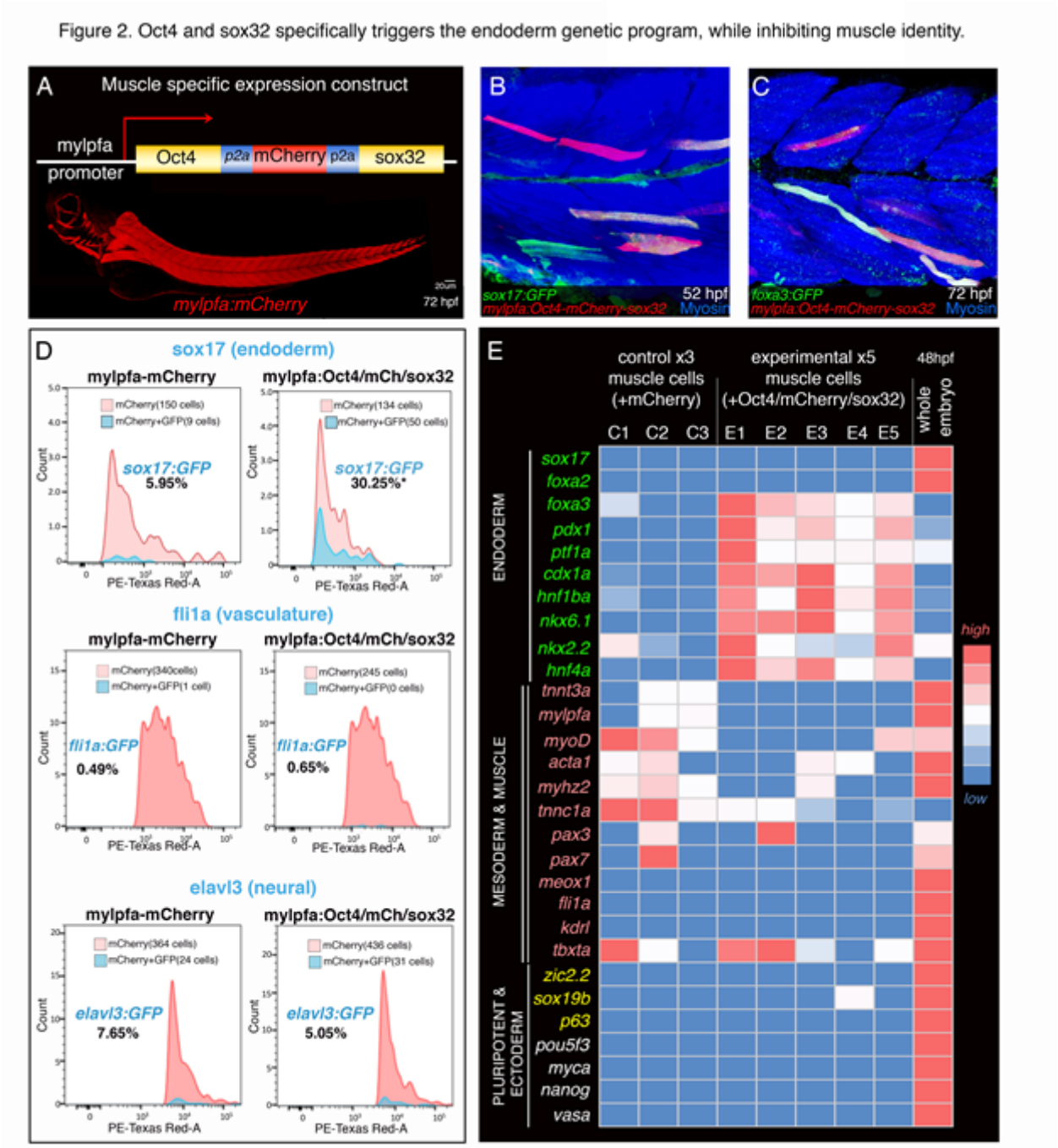

Figure 2

Oct14 and sox32 specifically triggers the endoderm genetic program, while inhibiting muscle identity

Figure 3 ptf1a expression in endoderm-induced muscle cells requires FGF signaling.

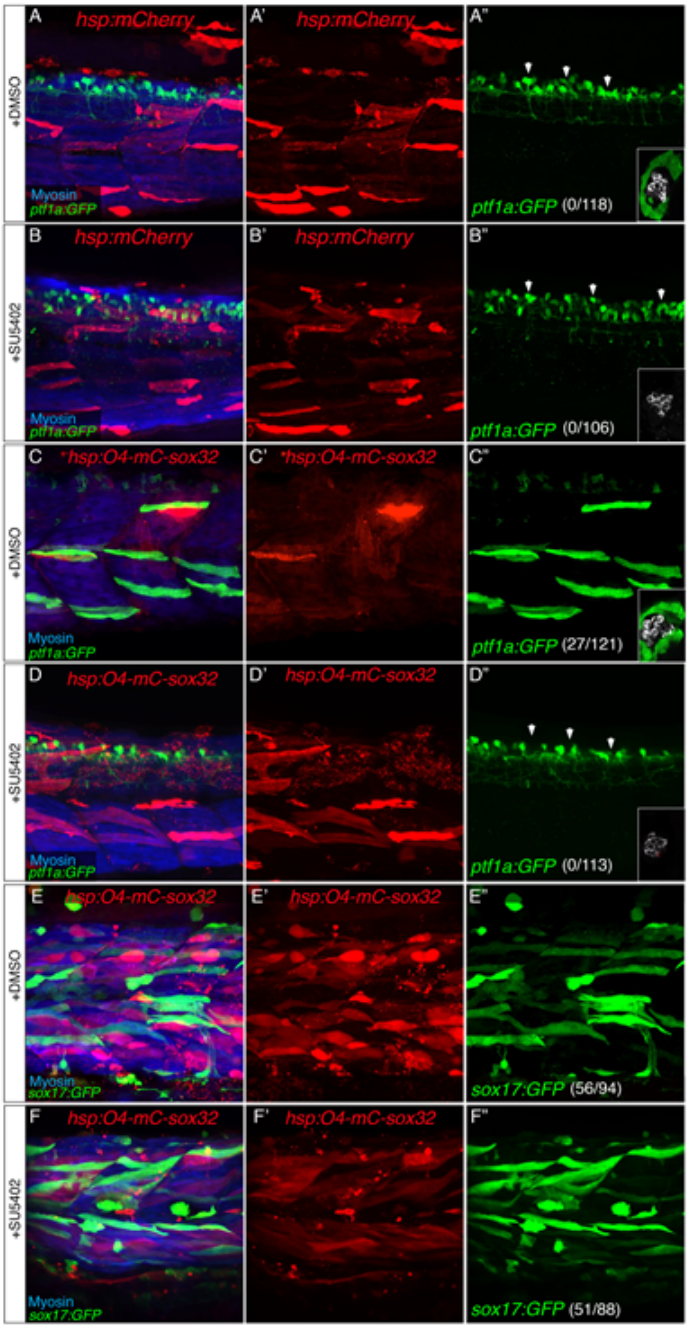

Figure 3

ptf1a expression in endoderm-induced muscle cells requires FGF signaling

Figure 4 Pluripotency defective Oct4 can induce endoderm genes in muscles.

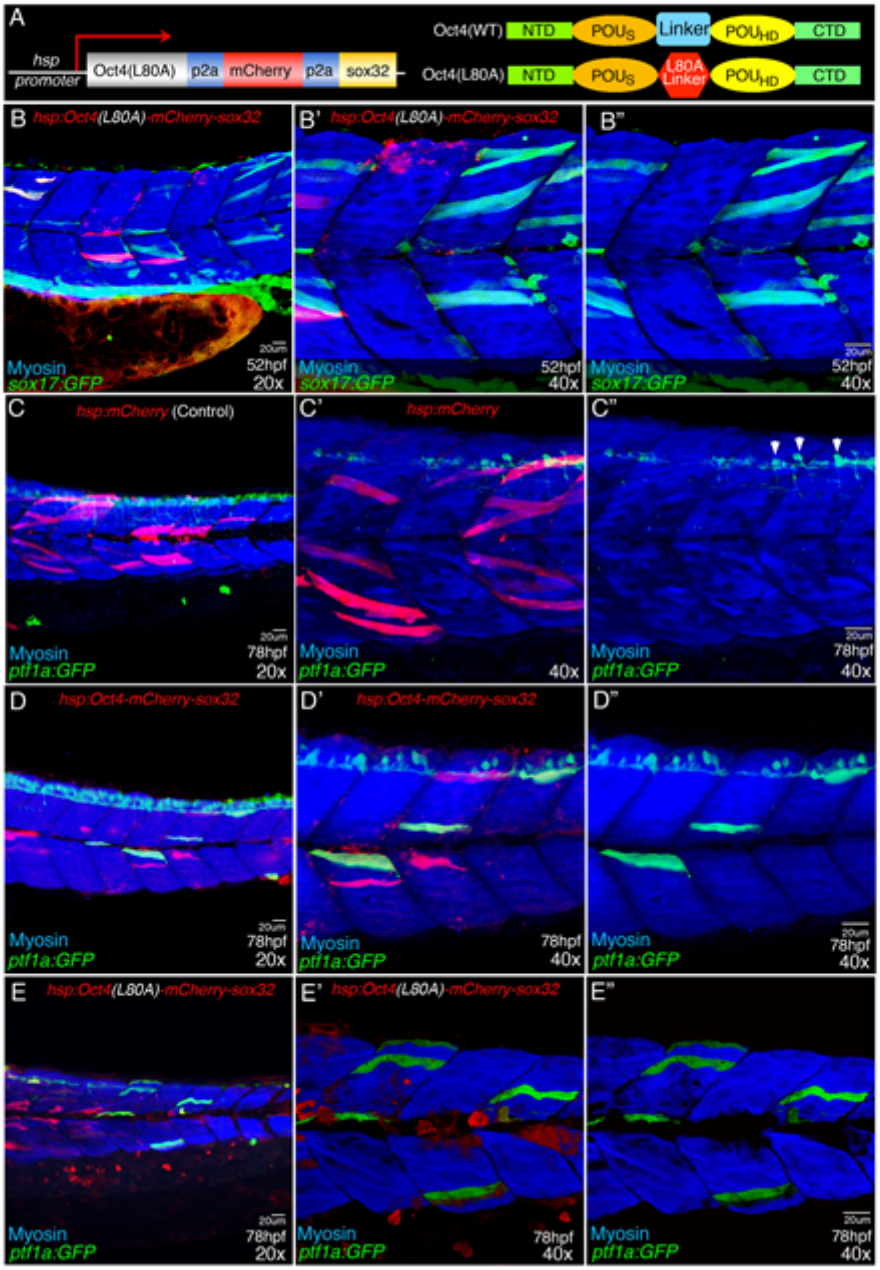

Figure 4

Pluripotency defective Oct4 can induce endoderm genes in muscles.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementFigure1X4SpecificityoffactorskdrlfliHuC.jpg

- SupplementalFigure1Legend.docx
- SupplementFigure2FinalX5CellCountsSkinNerveMusclecells.jpg
- SupplementalFigure2Legend.docx
- SupplementFigure3scRNAseq.jpg
- SupplementFigure3Legend.docx
- SupplementFigure4NuceliAggregation.jpg
- SupplementFigure4Legend.docx
- SupplementFigure5X4PtfTimecourseandmovieextensions.jpg
- SupplementFigure5Legend.docx
- SupplementFigure6FInalX7Lossofmuscletranscriptsandproteinsjpg
- SupplementFigure6Legend.docx
- SupplementFigure7Legend.docx
- SupplementFigure7FinalX4insituandQPCRofEndodermGenescopy.jpg
- SupplementMovieX1.mp4
- SupplementMovie1Legend.docx
- SupplementMovieX2.mp4
- SupplementMovie2Legend.docx
- SupplementMovieX3.mp4
- SupplementMovie3Legend.docx
- SupplementMovieX4.mp4
- SupplementMovie4Legend.docx
- SupplementaryFigureGaitingstrategyscRNASeqMcherryH2BControl.pdf
- SupplementaryFigureGaitingstrategyscRNASeqMcherryH2BOS.pdf
- SupplementaryFigureGatingstrategyforFACsanalysisqPCR.pdf
- SupplementaryTableEfficiecnyGraphsDataandstatistics.pdf
- SupplementaryTableCombinedqPCRdata.pdf
- AntibodydetailsforMaterialsandMethods.pdf
- qPCRPrimersusedinmanuscript.pdf