Hierarchical TBX6-FOXC Regulatory Logic Controls Human Trunk Mesoderm Diversification

John-Poul Ng-Blichfeldt[1], Rosalind Drummond[1], David Lando[1], Samina Kausar[1], Melania Barile[1], and Anna Philpott[1,2]

[1] Cambridge Stem Cell Institute, University of Cambridge, Cambridge CB2 0AW, UK

[2] Department of Oncology, University of Cambridge, Cambridge CB2 0AH, UK

Summary

Mesoderm forms during gastrulation and diversifies to generate many embryonic tissues, yet the regulatory logic governing early human mesodermal lineage decisions remains poorly defined. We use human iPSC-derived 3D trunk-like structures (hTLS) to model lineage specification during early post-gastrulation development, a stage experimentally inaccessible in humans. Developing hTLS generate neural and mesodermal lineages and recapitulate progressive mesodermal diversification into paraxial (somitic) and intermediate (renal) identities. Using genetic perturbations and temporally controlled transcription factor activation, we identify duration-dependent TBX6 activity as a critical determinant of mesodermal diversification. Downstream, the forkhead box transcription factors FOXC1 and FOXC2 stabilise somitic identity and are required for sclerotome differentiation. These findings reveal a duration-dependent hierarchical regulatory logic in early human mesoderm, in which competence for diversification is established before downstream lineage identities are stabilised, positioning hTLS as a powerful platform to experimentally dissect human developmental programs.

Introduction

How transient developmental states are converted into stable lineage identities remains a fundamental problem in developmental biology. Embryonic cell identity first diversifies during gastrulation when epiblast cells generate the three embryonic germ layers: cells that ingress through the primitive streak form mesoderm and endoderm, while the remaining epiblast becomes ectoderm [1,2]. Mesoderm further diversifies into distinct lineages that form the primordia of a wide range of tissues, with specific mesodermal identities arising from discrete positions along the rostral-caudal axis of the streak [3]. Emerging from the prospective trunk region of the embryo, paraxial mesoderm undergoes segmentation into somites, transient epithelial structures that further subdivide into sclerotome, which forms vertebrae, and dermomyotome, which produces skeletal muscle [4]. Adjacent to paraxial mesoderm, intermediate mesoderm gives rise to the kidneys and gonads [5,6]. Despite its fundamental importance, regulatory mechanisms by which early mesoderm diversifies into distinct lineages, including paraxial and intermediate mesoderm, remain poorly understood.

Classic fate-mapping and grafting experiments in avian and mouse embryos demonstrated that cells of the primitive streak represent a transient multipotent state characterised by broad equivalence in lineage competence, with positional cues within the streak directing specific mesodermal identities [7-12]. These positional cues are established by signalling gradients that both pattern the embryonic axes and bias lineage specification: Wnt activity posteriorises streak identity and promotes mesodermal differentiation, whereas retinoic acid (RA) antagonises Wnt/FGF signalling to permit differentiation of trunk progenitors while also contributing to posterior neural patterning [13,14]. However, downstream of local signalling environments, how cell-intrinsic transcriptional programs convert transient multipotent states irreversibly into stable lineage-specific mesodermal identities remains poorly defined.

Recent comparative studies have highlighted divergence in developmental mechanisms between mouse and human mesodermal derivatives, including differences in morphogenesis and patterning of the kidney [15-18] and in the tempo and signalling

mechanisms underlying somitogenesis [19,20]. However, direct investigation of these early developmental events in humans is constrained by technical and ethical limitations [21]. Advances in single-cell transcriptomics have provided valuable insights into transcriptional states associated with early lineage diversification in human embryos, yet these stages remain experimentally inaccessible in vivo [22-24]. Recently, the development of human pluripotent stem cell–based 3D differentiation systems has begun to enable functional investigations of early human developmental programs by modelling key aspects of embryonic development in vitro [25-33].

Here, we employ recently described human induced pluripotent stem cell-derived trunk-like structures (hTLS) to investigate the molecular mechanisms governing early diversification of intermediate (renal) and paraxial (somitic) mesoderm lineages during early post-gastrulation human trunk development. By combining temporal control of transcription factor activity with genetic perturbations, we uncover a previously undefined duration-dependent hierarchical regulatory logic in which transient competence for multipotent differentiation is established and subsequently stabilised to drive lineage commitment. Together, these findings establish a hierarchical framework governing human mesodermal fate diversification and position human trunk-like structures as a tractable platform for dissecting early human developmental programs.

Results

Human trunk-like structures (hTLS) model progressive mesodermal diversification

To investigate how trunk mesodermal lineages emerge and diversify during early post-gastrulation human development, we adapted published protocols to generate three-dimensional trunk-like structures (hTLS) from human iPSCs. In this system, iPSCs are exposed to a transient pulse of high Wnt signalling, followed by reduced Wnt signalling combined with RA, prior to aggregation [26]. Under these signalling conditions, neural and mesodermal identities emerge concomitantly, recapitulating lineage diversification characteristic of posterior (post-cranial) embryonic development (Figure 1A) [26]. Differentiating aggregates embedded in Matrigel elongated and generated $SOX2^+$ neural and $SOX2^-$ non-neural tissues in parallel (Figures 1B and S1A,A').

At day 5, $SOX2^+$ structures exhibited lumens and expressed PAX6, consistent with neural tube-like identity (Figure S1B) [34]. Surrounding $SOX2^-$ tissue expressed the paraxial mesoderm marker forkhead box C2 (FOXC2), and the nephron progenitor marker PAX8, indicating the concurrent emergence of both paraxial (somitic), and intermediate (renal) mesodermal identities (Figures 1C and S1B). By day 8, hTLS contained somitic ($DACH2^+$) and renal ($PAX8^+$) epithelial structures with apical-basal polarity, indicated by the apical localisation of ZO1 (Figure 1D). Renal epithelial structures expressed CDH1 (E-Cadherin/ECAD), hepatocyte nuclear factor beta (HNF1B) and PAX8, consistent with nephron tubular identity (Figure S1C). Similar neural, somitic, and renal lineages were observed in hTLS generated from an independent iPSC line (IMR-90-4), demonstrating protocol robustness across different genetic backgrounds (Figure S1D).

To define the transcriptional programs underlying lineage emergence in hTLS, we performed single nucleus RNA-sequencing (snRNA-seq) across a time course from day 0 to day 7, including undifferentiated iPSCs as a control (Figures 1E and S1E-G). Unsupervised clustering revealed the presence of *MEOX1*$^+$ somitic mesoderm, intermediate mesoderm-derived *PAX2*$^+$ kidney progenitors and their derivatives

(*HNF1B*$^{+}$ nephron tube and *NPHS1*$^{+}$ podocytes), as well as neural tube-derived *MAP2*$^{+}$ neurons and *TFAP2B*$^{+}$ neural crest-like populations (Figures 1E' and S1G-J). *PECAM-1*$^{+}$ endothelial cells were observed by snRNA-seq and immunofluorescence (Figure S1K).

To assess the developmental correspondence of hTLS to human embryogenesis *in vivo*, we compared the hTLS snRNA-seq dataset to reference atlases spanning zygote to post-conception week 2 (PCW2/CS7) [24], and to an independent PCW3/CS10 human embryo dataset [23]. hTLS cells from days 0-4 mapped predominantly to CS7-stage embryonic populations (Figure S1L), whereas cells from days 5-7 showed strong correspondence to PCW3 neural tube, somite, intermediate mesoderm and endothelial lineages, while lacking notochord, endoderm, cardiac and erythroid populations (Figure 1F). These analyses support the use of hTLS as an experimentally tractable model of early post-gastrulation human trunk development.

To examine mesodermal diversification in greater detail, we isolated mesodermal cells in the hTLS snRNA-seq dataset and performed secondary unsupervised clustering (Figures 1G and S1M-P). Early time points were dominated by a single mesodermal population enriched for *TBX6*, *FGF10* and *PDGFRA* (Early mesoderm; Figures 1H and S1P). From day 3 onward, additional clusters emerged corresponding to nephron progenitors (*PAX8*$^{+}$ *CDH6*$^{+}$) and somite progenitors (*MEOX1*$^{+}$ *PAX3*$^{+}$), followed by diversification into nephron derivatives, alongside dorsal somite (*PAX3*$^{+}$ *DACH2*$^{+}$)-derived muscle (*TTN*$^{+}$) and mesenchymal (*FAP*$^{+}$) populations at later time points (Figures 1G,H and S1P). These data indicate that hTLS recapitulate progressive mesodermal diversification into to paraxial and intermediate mesoderm derivatives, and therefore enable detailed mechanistic investigation of mechanisms controlling their divergence.

To reconstruct developmental trajectories underlying mesodermal diversification, we performed pseudotime inference using Slingshot [35], identifying four major trajectories corresponding to nephron tube, podocytes, somite-muscle and somite-mesenchymal lineages (Figure 1J). While cluster identities were used to define trajectory topology, gene expression dynamics along pseudotime were analysed at

single-cell resolution. Using tradeSeq [36], we identified lineage-specific transcriptional programs associated with terminal states along each trajectory (Figure 1K,L). Nephron tube and podocyte lineages exhibited distinct renal epithelial (*PAX2*, *PATJ*) and glomerular (*WT1*. *MAGI2*) gene signatures, respectively, whereas somite-muscle and somite-mesenchymal lineages were marked by sarcomeric (*TTN*, *NEB*) or extracellular matrix-associated (*COL23A1*, *SULF1*) gene expression. Together, these analyses establish hTLS as a system in which trunk mesodermal lineages progressively diversify along transcriptionally distinct trajectories, providing a foundation for mechanistic interrogation of human trunk mesoderm specification and diversification.

TBX6 is required for establishment of a transient multipotent mesodermal competence state in hTLS

To identify factors regulating the initial establishment of mesodermal multipotency in hTLS, we compared the earliest emerging mesodermal and neural populations within the hTLS snRNA-seq dataset. Differential expression analysis identified 27 transcription factors enriched in early mesoderm, with T-box transcription factor *TBX6* as one of the most highly enriched (Figure 2A, A'). Although TBX6 is classically associated with somitic mesoderm specification, recent work has suggested that its expression precedes overt lineage diversification [37,38], raising the possibility that TBX6 functions more broadly to establish mesodermal competence. We therefore examined the timing and requirement for TBX6 expression during early hTLS differentiation.

We observed by immunofluorescence that TBX6 was expressed in iPSCs after 24 hours of differentiation (day 0 of the protocol; Figure S2A,B). At day 2 within aggregated hTLS, TBX6 and SOX2 marked distinct, non-overlapping cell populations, with TBX6 expression declining at day 3 and largely depleted by day 4 (Figure 2B). Time course RT-qPCR analyses revealed that induction of *TBX6* preceded markers of both somitic mesoderm (*FOXC2*, *MEOX1*, and *PAX3)* and kidney mesoderm (*OSR1*, *PAX2*, *PAX8*; Figure S2D). Thus, consistent with observations in mouse embryos *in vivo* [37], TBX6 is transiently expressed throughout the mesoderm prior to lineage diversification in hTLS, suggesting a broader role in establishing mesodermal competence.

We next investigated the role of TBX6 in mesodermal emergence. We engineered iPSCs using dead-Cas9-based CRISPR interference (Figure 2C) [39] to suppress *TBX6* expression under the control of Doxycycline (Dox) with cytoplasmic GFP as a readout for expression of dCas9 in response to Dox (TBX6-dCas9-KRAB). Addition of Dox at the onset of differentiation (day -1) resulted in ~82% knockdown of *TBX6* expression at day 0 compared to untreated cells ($p = 0.0042$; Figure 2D,E). *TBX6* knockdown was associated with decreased expression of the early mesodermal markers *TBXT* ($p = 0.011$) and *MSGN1* ($p = 0.0031$; Figure S2E). Depletion of TBX6 protein after Dox treatment was confirmed by immunofluorescence (Figure S2F). Next, TBX6-dCas9-KRAB iPSCs were differentiated to hTLS and harvested at day 5 of the protocol (Figure 2D). Control hTLS did not express GFP, indicating lack of knockdown cassette expression, and exhibited somitic ($FOXC2^+$), renal ($PAX8^+$) and neural ($SOX2^+$) tissue as observed previously (Figure 2F). By contrast, Dox-treated TBX6-dCas9-KRAB hTLS showed heterogeneous GFP expression with a reduction in both $FOXC2^+$ and $PAX8^+$ tissue and an expanded $SOX2^+$ domain. Remaining $FOXC2^+PAX8^+SOX2^-$ cells were GFP^- and thus did not respond to Dox, indicating that renal and somitic mesodermal differentiation was mutually exclusive with TBX6 depletion (Figure 2F). Quantification revealed a 78% decrease in the number of $FOXC2^+$ cells per confocal image ($p < 0.0001$; Figure 2G), and an 85% decrease in $PAX8^+$ cells ($p < 0.0001$; Figure 2H), with a corresponding 84% increase in $SOX2^+$ cells ($p < 0.0001$; Figure 2I). Similar results were obtained with an independent iPSC line with a distinct gRNA to target *TBX6* (Figure S2G). These data indicate that TBX6 is necessary for establishment of a multipotent mesodermal population in hTLS, from which both somitic and renal lineages emerge.

Acute TBX6 expression promotes mesodermal transcriptional progression without commitment to a specific lineage

TBX6 is transiently expressed at low levels in a heterogeneous manner early in hTLS differentiation (Figure S2B). This led us to ask if TBX6 is only permissive for mesodermal diversification, or whether its overexpression would drive mesodermal progression. We therefore engineered a piggyBac-insertable, Dox-inducible TBX6

overexpression system, with cytoplasmic mCherry as a fluorescent readout for successful transgene overexpression, to more uniformly elevate TBX6 expression in the earliest stages of hTLS differentiation (PB-Dox-TBX6; Figure 3A-C).

After 24 hours of differentiation (day 0 of the protocol), control cells lacked mCherry expression and formed compact colonies, whereas cells differentiated in the presence of Dox showed robust mCherry expression and adopted a more dispersed, mesenchymal-like morphology (Figure 3D). RT-qPCR confirmed strong induction of *TBX6* expression ($p<0.0001$) and a concomitant reduction in *SOX2* levels ($p = 0.0093$; Figure S3A). *TBX6* overexpression significantly upregulated markers associated with both paraxial mesoderm (*FOXC1*, *FOXC2*), and intermediate mesoderm (*OSR1*), suggesting activation of diverse mesodermal lineage programs (Figure S3A).

To define global transcriptional changes induced by acute TBX6 overexpression, we performed bulk RNA-seq after 24 hours in the presence or absence of Dox (Figures 3B,C and S3B,C). TBX6 induction resulted in widespread transcriptional rewiring, including downregulation of pluripotency genes (e.g. *NANOG*) and alternative lineage genes (e.g. the endoderm regulator *FOXA2*), and upregulation of genes associated with various differentiated mesodermal states including *FOXC1* and *OSR1* (Figure 3E). Gene ontology (GO) analysis revealed enrichment of biological processes related to differentiation of mesodermal derivatives, including muscle, bone, and mesenchyme development (Figure S3D).

Gene set enrichment analysis (GSEA) using reference human embryonic atlases demonstrated selective enrichment of mesodermal transcriptional programs following TBX6 induction, alongside depletion of epiblast and primitive streak signatures (Figures 3E and S3E-I) [23,24]. Notably, TBX6-induced programs aligned with multiple mesodermal lineages, including paraxial, presomitic, and intermediate mesoderm, while neural programs were depleted (Figure S3E). These findings were supported by module-based mapping to human embryonic reference atlases (Figure S3E-L). Together, these data indicate that acute TBX6 overexpression alone is sufficient to promote transcriptional progression toward differentiated mesodermal states but does not enforce commitment to a single mesodermal lineage.

Duration of TBX6 activity determines commitment to somitic or alternative mesodermal fates

Previous studies in mouse embryonic stem cells suggested that the duration of TBX6 expression can influence mesodermal differentiation [38]. Given that acute TBX6 induction activates transcriptional programs associated with multiple mesodermal lineages, we next asked whether altering the duration of TBX6 activity might influence mesodermal lineage commitment in hTLS. We first investigated the consequences of sustained TBX6 overexpression by differentiating PB-Dox-TBX6 cells to hTLS in the continuous presence of Dox (Figure 4A,B). As a control, iPSCs harbouring Dox-inducible mCherry only were differentiated to hTLS (PB-Dox-mCherry; Figure 4B). At day 5, hTLS generated from both PB-Dox-TBX6 and PB-Dox-mCherry cells exhibited many mCherry^{+} cells, indicating that our piggyBac-based Dox-inducible transgene expression system remained unsilenced in these cells, likely owing to the presence of multiple insulator sequences (Figure 4B-D).

At day 5, TBX6::mCherry^{+} cells predominantly expressed the somitic marker FOXC2, lacked expression of PAX8 and SOX2, and adopted a mesenchymal morphology, indicating restriction toward a somitic mesoderm fate (Figure 4C). By contrast, hTLS generated from PB-Dox-mCherry cells exhibited mCherry expression within epithelial structures across somitic, renal, and neural compartments (Figure 4D), confirming that sustained TBX6 activity specifically, and not Dox-inducible transgene expression per se, drives a specific somitic lineage bias.

Since TBX6 activity is transient both *in vivo* and during hTLS development (Figure S2A), we next asked whether restricting the duration of ectopic TBX6 activity might enable mesodermal diversification. To precisely control TBX6 protein stability while retaining a record of prior TBX6 expression, we utilised a FKBP domain fused to TBX6 to develop a conditional degradation and lineage-tracing system that we termed DegTrace (Figure 4E) [40]. In this system, Dox induces transcription of the *TBX6-FKBP-T2A-mCherry* cassette, whereas subsequent addition of dTag selectively depletes TBX6 protein by engaging the fused FKBP domain with E3 ubiquitin ligases for proteasomal degradation [40], without affecting the co-translated mCherry reporter, thus enabling

lineage tracing of cells that have transiently expressed TBX6 (Figures 4E). We validated DegTrace in iPSCs, confirming that dTag-mediated degradation selectively depleted TBX6 protein while preserving the mCherry reporter (Figure S4A,B).

We next applied DegTrace to 3D trunk differentiation. TBX6 overexpression was induced with Dox from day -1, and protein degradation was triggered by dTag addition at day 0 (after 24 h), day 1 (48 h), or day 2 (72 h) (Figure 4E). Dox alone induced the *TBX6* transgene irrespective of subsequent dTag addition (Figure 4F). Sustained TBX6 overexpression (Dox alone) robustly induced the somitic mesoderm markers *FOXC1* and *MEOX1* (Figure 4F). By contrast, restricting TBX6 activity to the first 24 h led to diminished expression of these somitic markers, whereas inducing degradation after 48 or 72 h of TBX6 activity resulted in partial marker retention (Figure 4F).

Consistent with these transient transcriptional effects, immunofluorescence analysis revealed that cells exposed to transient TBX6 activity (24 h) failed to maintain mesodermal identity and instead adopted a $SOX2^{+}PAX3^{+}$ neural epithelial fate (Figures 4G and S4D), whereas cells exposed to an intermediate duration of TBX6 activity (72 h) differentiated into mesodermal epithelial structures expressing both FOXC2 and PAX8 (Figure 4H). By contrast, sustained TBX6 overexpression promoted mesenchymal differentiation and restriction to a $FOXC2^{+}PAX8^{-}$ somitic fate (Figure 4C,G).

Together, these findings indicate that a minimum duration of TBX6 activity is needed to suppress neural identity and promote mesodermal identity. An intermediate duration enables acquisition of a multipotent epithelial trunk mesoderm state with renal competence. By contrast, prolonged TBX6 activity drives epithelial-to-mesenchymal transition, represses renal identity, and locks cells into the somitic lineage. Notably, the intermediate duration of exposure closely mirrors the endogenous dynamics of TBX6 expression observed in hTLS (Figures 2B and S2A-D).

FOXC1 and FOXC2 establish and safeguard somitic mesoderm identity

Having established that TBX6 duration governs mesodermal fate output, we next sought to identify downstream mechanisms that stabilise specific mesodermal identities.

TBX6 alone biased differentiation toward somite-like fates in our human trunk-like embryoid model, so we sought to investigate downstream transcriptional regulators of the somitic lineage. Analysis of the hTLS snRNA-seq dataset (Figure 1E) identified a somitic progenitor population enriched for canonical somitic transcription factors including *FOXC1*, *FOXC2*, *PAX3*, *TCF15* (Paraxis), *SIX1*, *MEOX1* and the Hedgehog pathway mediator *GLI2* (Figure 5A,B).

We focussed on investigating the function of FOXC1 and FOXC2 in human embryoids based on prior genetic evidence that combined loss of these factors in mice disrupts somite formation and leads to expansion of intermediate (renal) mesoderm [41]. As prior mouse studies indicated partial functional redundancy between FOXC1 and FOXC2 in early development [41-43], we engineered a dual knockdown system using dCas9-KRAB, hereafter referred to as FOXC1/2-dCas9-KRAB (Figures 5C and S5C).

To examine the role of FOXC1/2 specifically in somitic mesoderm differentiation, we adapted a published 2D somitic mesoderm differentiation protocol [44] that recapitulates progression from primitive streak through paraxial mesoderm to an early somite-like state (Figures 5D and S5A,B). Dual knockdown of *FOXC1* and *FOXC2* was induced at the primitive streak stage, and cells were harvested at the early somite stage for bulk RNA-seq (Figures 5D and S5D,E).

Dual FOXC1/2 knockdown efficiently suppressed expression of both factors and was accompanied by a robust decrease in canonical somitic markers, including *RIPPLY1*, *PAX3*, *TCF15, HAND1* and *TWIST1* (Figure 5E). Concomitantly, dual FOXC1/2 knockdown resulted in the induction of genes associated with alternative mesodermal lineages, including renal (e.g. *NPHS1, EYA2, JAG1)* and muscle differentiation programs (e.g. *MYMX, MYH11, MYBPC2*; Figures 5D and S5G). Early and paraxial mesodermal genes (*CDX2, TBX6, MSGN1*) were upregulated, indicating that dual FOXC1/2 knockdown did not prevent establishment of paraxial mesoderm, but instead prevented progression from paraxial mesoderm to early somite (Figure S5F). GO analysis confirmed depletion of patterning and skeletal development programs, with

concomitant enrichment of muscle, mesenchymal, and renal system processes (Figure S5H).

To better understand how FOXC1/2-dependent genes correspond to human embryonic programs in vivo, we performed GSEA using human embryonic reference atlases [23], which revealed that FOXC1/2 knockdown resulted in depletion of signatures associated with somitic mesoderm derivatives, and enrichment of alternative lineage programs, including intermediate mesoderm and cardiomyocyte signatures (Figure S5I). Moreover, GSEA using independent human embryonic heart [45] and kidney [46] reference atlases revealed that genes upregulated upon FOXC1/2 knockdown were enriched for cardiomyocyte (Figure S5J) and renal epithelial tube signatures (Figure S5K). Together, these results indicate that FOXC1 and FOXC2 are required to establish somite transcriptional identity and to suppress alternative mesodermal differentiation programs.

To investigate mechanisms regulating FOXC1/2-dependent gene expression, we used published ATAC-seq profiles from the equivalent early somite population generated with the same protocol [44]. Motif enrichment analysis was performed on accessible chromatin regions adjacent to genes we found to be downregulated following dual FOXC1/2 knockdown, using accessible regions associated with expressed genes that were not FOXC1/2-dependent as background. This analysis revealed enrichment of forkhead, homeobox, and basic helix-loop-helix (bHLH) motifs within FOXC1/2-dependent regulatory regions, suggesting that FOXC1 and FOXC2 act within a broader regulatory landscape of combinations of transcriptional regulators rather than as solitary lineage-determining factors (Figure 5F).

To test whether FOXC1/2 regulate somite identity in hTLS, we induced dual knockdown during hTLS differentiation (Figure 5G). While neural and renal tissues were preserved, $PAX3^{+}SOX2^{-}$ somitic tissue was markedly reduced (68% reduction; $p < 0.0001$), demonstrating a specific requirement for FOXC1/2 in establishing somitic mesoderm identity in hTLS (Figures 5H-K and S5L).

Finally, to determine whether TBX6 function relies on FOXC1/2, we generated iPSCs harbouring both a Dox-inducible TBX6 overexpression cassette and the FOXC1/2-dCas9-KRAB knockdown system (Figure S5M,N). TBX6 overexpression alone induced multiple somitic genes, including *TWIST1*, *TWIST2*, *MEOX1*, *MEOX2* and *TBX18*; however, this induction was abolished upon simultaneous FOXC1/2 knockdown, indicating that TBX6-driven somitic differentiation depends on FOXC1/2 activity (Figure S5N). By contrast, the somitic genes *TCF15* and *SIX1* were not induced by TBX6 but were downregulated upon FOXC1/2 knockdown, suggesting TBX6-independent maintenance of somitic identity by FOXC factors, and some TBX6-responsive genes, including MESP2 [38] and OSR1, were unaffected by FOXC1/2 knockdown. These data indicate parallel FOXC-independent TBX6 pathways (Figure S5N).

Together, these data support a model in which the duration of TBX6 activity enables mesodermal diversification, while FOXC transcription factors function downstream of TBX6 to constrain mesodermal plasticity and establish somitic identity.

FOXC1 and FOXC2 stabilise somitic identity and preserve competence for further diversification

Having found that FOXC1/2 are required to establish somite mesoderm identity, we next asked whether they are needed to maintain somite identity once established. To investigate this, we differentiated cells to early somite stage and then induced dual knockdown (Figure 6A). Dual FOXC1/2 knockdown alone led to a reduction in somite markers *TCF15*, *TBX18* and *MEOX2,* whereas *MEOX1* and *SIX1* were unchanged, suggesting that FOXC1/2 loss partially destabilised the early somite program (Figure 6B).

We next tested whether dual knockdown affected the ability of early somite cultures to differentiate into sclerotome, a transient embryonic structure that gives rise to the vertebral skeleton. Studies of avian and mouse development have established that sclerotome differentiation *in vivo* is triggered in the ventral somite by Hedgehog ligands

derived from the notochord and floorplate [47,48]. This process is characterised by repression of PAX3, upregulation of FOXC1/2, and induction of downstream markers including TWIST1, PAX1, PAX9, and SOX9 together with an epithelial-to-mesenchymal transition (EMT) (Figure S6A).

We adapted a 2D somite-to-sclerotome differentiation protocol in which early somite cultures were exposed to smoothened agonist (SAG) to activate Hedgehog signalling and trigger sclerotome differentiation (Figure 6A) [44]. SAG treatment resulted in efficient generation of sclerotome characterised by suppression of *PAX3* and upregulation of *FOXC1*, *FOXC2*, *TWIST1*, *PAX9*, *PAX1* and *SOX9* compared to untreated cells (Figure 6A,C,D).

Dual FOXC1/2 knockdown induced at the early somite stage, in combination with SAG treatment, strongly attenuated induction of sclerotome markers (Figure 6C,D), although induction of canonical Hedgehog pathway targets *PTCH1* and *GLI1* by SAG was unaffected by FOXC1/2 knockdown (Figure 6D). These data indicate that FOXC1/2 are not necessary for Hedgehog signalling per se, but instead support maintenance of somite identity and thereby preserve competence for sclerotome differentiation, or indeed may play a more direct role in sclerotome differentiation.

As FOXC1/2 knockdown inhibited the establishment and maintenance of somite identity, we next asked whether they are sufficient to promote somite differentiation when overexpressed. Given the functional redundancy of FOXC1 and FOXC2 *in vivo* and in hTLS, we focussed on FOXC1 and generated a Dox-inducible FOXC1 overexpression line (PB-Dox-FOXC1; Figure 6E). FOXC1 overexpression was induced at day 2 of differentiation to match the timeframe of the previous FOXC1/2 knockdown experiments in hTLS (Figures 5G-J and 6F).

RT-qPCR analysis of 2D-cultured hTLS generated with PB-Dox-FOXC1 cells revealed that FOXC1 overexpression robustly upregulated *TWIST1* (~2-fold; $p < 0.01$), a bHLH transcription factor associated with ventral somite identity, sclerotome specification and EMT. FOXC1 overexpression also upregulated the ventral somite marker *MEOX1* (~1.4-fold; $p < 0.05$), but only modestly induced PAX9 and *TWIST2* and failed to induce SOX9 (Figure 6G).

Overexpression of a FOXC1 variant lacking the forkhead DNA-binding domain (FOXC1ΔFKH) failed to upregulate *TWIST1,* even at the highest Dox concentration (1 µM) despite robust transgene induction, indicating that *TWIST1* induction is specific to transcriptionally active FOXC1 (Figure 6H).

To investigate the effects of FOXC1 overexpression in a context more closely resembling native embryonic tissue, PB-Dox-FOXC1 cells were differentiated to 3D-hTLS, and FOXC1 overexpression was induced from day 2 of hTLS differentiation (Figures 6I). Robust, heterogeneous mCherry induction was observed in elongating hTLS and maintained through days 3-5 (Figure 6J). FOXC1-overexpressing cells were largely excluded from the epithelial cysts that formed and instead occupied a sub-epithelial position, consistent with acquisition of a mesenchymal-like state (Figure 6K). By contrast, FOXC1ΔFKH-overexpressing cells were observed within epithelial cysts, indicating that FOXC1 transcriptional activity is needed to drive the mesenchymal phenotype (Figure 6N).

Cells overexpressing wild type FOXC1 (FOXC1::mCherry^{+}) within hTLS lacked expression of SOX2, PAX3, or PAX8, indicating that FOXC1 overexpression did not promote neural, dorsal somite, or renal fates (Figure 6L,M). Moreover, FOXC1-overexpressing cells robustly expressed TWIST1 (Figure 6M) but rarely co-expressed the sclerotome marker SOX9 (Figure S6H), suggesting that FOXC1 was insufficient to fully execute the sclerotome differentiation program alone.

More generally, snRNA-seq analysis of untreated hTLS did not reveal a sclerotome population (Figure S1G-J), likely owing to the absence of notochord- and floorplate-derived Hedgehog ligands in this system (Figure S6A,B). To test whether hTLS retain competence for sclerotome differentiation, we activated Hedgehog signalling with SAG from day 5 to day 8 of differentiation (Figure S6C). SAG treatment led to expansion of FOXC2^{+} tissue and depletion of PAX3^{+} dorsal somitic domains, consistent with somite ventralisation and sclerotome differentiation (Figure S6D-G).

Together, these results indicate that FOXC1 and FOXC2 are required to establish and stabilise somitic mesoderm identity and to enable somite-to-sclerotome differentiation. Moreover, while FOXC1 overexpression is sufficient for *TWIST1*

induction and EMT-like behaviour, full sclerotome differentiation requires additional cues such as Hedgehog signalling.

TWIST1 is required for Hedgehog-driven sclerotome differentiation

Because our ATAC-seq analysis suggested FOXC factors co-operate with bHLH factors to establish the somitic transcriptional landscape, we hypothesised that a bHLH factor might act as the direct effector of FOXC1/2. FOXC1 overexpression upregulated the bHLH factor TWIST1 and drove EMT in hTLS, and TWIST1 has known roles in inducing EMT [49,50], positioning TWIST1 as a plausible candidate bHLH functioning downstream of FOXC1. We edited iPSCs to harbour a Dox-inducible TWIST1 overexpression cassette and differentiated cells to hTLS (Figure S7A,B). Inducible TWIST1 overexpression led to sub-epithelial localisation of TWIST1::mCherry^{+} cells and exclusion from epithelial cysts, closely resembling the FOXC1 overexpression phenotype (Figure S7C), suggesting that TWIST1 may function downstream of FOXC1 to drive EMT in this system.

Finally, since FOXC1 and FOXC2 were necessary for sclerotome differentiation, we asked whether TWIST factors are also required. We performed dual knockdown of TWIST1 and TWIST2 during SAG-induced differentiation (Figure S7D), and found that dual knockdown strongly reduced SAG-induced expression of sclerotome markers, but did not affect induction of FOXC1 or FOXC2 by SAG (Figure S7E). Together, these data suggest that FOXC factors may drive EMT via TWIST1, and support a model in which TWIST factors act as necessary effectors of Hedgehog-driven sclerotome differentiation in human embryonic mesoderm.

Discussion

How lineage identities first emerge and diversify during early human embryogenesis remains a fundamental unresolved question, largely owing to the experimental inaccessibility of post-gastrulation human development *in vivo*. While recent single-cell atlases have mapped transcriptomic states during early human development, experimental dissection of the mechanisms governing lineage diversification during this window has remained limited. To address this knowledge gap, we use a human iPSC-derived trunk-like structure (hTLS) embryoid model, combined with genetic tools we developed that enable precise temporal control of transcription factor activity or dual gene knockdown, to uncover a hierarchical regulatory logic underlying early human mesodermal diversification. Our data support a model in which TBX6 functions as a “temporal gate” that governs mesodermal diversification, with the duration of TBX6 activity determining intermediate versus somitic mesodermal differentiation. Downstream of TBX6, FOXC1/2 act as a “lock” to establish and stabilise somitic mesoderm identity, preserving the ability of somite cells to undergo sclerotome differentiation in response to Hedgehog signalling, in part through induction of TWIST factors. Given that many developmental transcription factors are transiently expressed, our findings suggest that the duration of transcription factor activity may encode competence windows that hierarchically restrict lineage plasticity, potentially representing a general principle of lineage specification during embryogenesis.

We benchmarked our embryoid model against human embryonic *in vivo* reference atlases, validating hTLS as an experimentally tractable model that recapitulates key aspects of post-gastrulation human embryonic development. The presence of intermediate (pre-renal) and paraxial (pre-somitic) mesoderm in our hTLS model enabled investigation into mechanisms regulating emergence of these distinct mesodermal subtypes. Fate-mapping of peri-gastrulation avian and mouse embryos showed that intermediate and paraxial mesoderm arise from discrete positions along the rostral-caudal axis of the primitive streak [10]. Moreover, rostrally-located streak cells fated to become head mesenchyme and anterior somites would adopt lateral fates including intermediate and lateral plate mesoderm when grafted caudally [7], indicating

developmental equivalence of the multipotent cellular pool that generates diverse mesodermal subtypes. While morphogen gradients including BMP and FGF pattern the mediolateral axis of the mesoderm [3,51], specific transcription factors mediating identity changes had remained poorly defined. Our finding that FOXC1/2 commits cells to paraxial mesoderm fates and suppresses alternative fates including intermediate mesoderm provides a molecular basis for mesodermal fate commitment likely operating downstream of local microenvironmental cues. Our epistasis experiments, and our finding that FOXC1 and FOXC2 are induced by acute TBX6 overexpression, position FOXC1/2 functionally downstream of TBX6; however, direct analysis of TBX6 binding is needed to confirm whether this is a chromatin-level interaction or an indirect mechanism via intermediates.

Our DegTrace system enabled precise control over the duration of transcription factor activity while simultaneously tracing the fate of modified cells, allowing us to experimentally interrogate how transcription factor dynamics influence lineage outcomes. Using this system, we uncovered a duration-dependent TBX6 regulatory logic underlying mesodermal diversification. TBX6 is classically described as a critical regulator of somitic mesoderm specification. TBX6 null mice lack somitic mesoderm and instead exhibit three neural tubes [52,53], and TBX6 has been proposed to repress the SOX2 locus to permit somitic mesoderm differentiation in neuromesodermal progenitors [53]. Accordingly, recent human population genetics studies have revealed TBX6 variants to be associated with congenital vertebral malformations, reflecting the somitic origin of vertebrae [54-56]. Our finding that FOXC1/2 act downstream of TBX6 to establish and stabilise somitic mesoderm identity provides a mechanistic explanation for these defects. Consistent with this, FOXC1 mutations in humans can cause Axenfeld-Rieger syndrome, which includes skeletal abnormalities such as scoliosis in addition to ocular defects [57,58].

Our observation that TBX6 knockdown depletes both kidney and somite mesoderm is consistent with a model in which a multipotent mesodermal population expressing TBX6 gives rise to both intermediate and paraxial mesoderm during post-gastrulation development. Studies of early mouse development mapped initial TBX6

expression broadly throughout the primitive streak, becoming progressively more posteriorly restricted as development proceeds [37]. A multipotent TBX6^{+} progenitor of intermediate and somitic mesoderm is supported by lineage tracing studies of TBX6-expressing cells in mice, which showed labelling of somitic and intermediate mesoderm when traced at early stages, and labelling restricted to somites if traced later [59]. These observations are consistent with a model in which longer durations of TBX6 activity bias cells toward somitic fate, whereas shorter exposure permits alternative mesodermal outcomes. Supporting this broader role, TBX6 variants have been associated with defects in intermediate mesoderm derivatives including the kidneys, in both human genetic studies and mouse models [60-65]. Together with these studies, our findings position TBX6 as a potential regulator of intermediate mesoderm development in addition to its known role as a somitic mesoderm specifier.

The duration-dependent effects of TBX6 activity revealed by DegTrace are consistent with a recent report showing that a defined pulse of TBX6 induction enhances cardiomyocyte differentiation efficiency in pluripotent stem cell protocols [38]. These observations raise the question of how the activity duration of a single transcription factor can encode competence for divergent lineage outcomes. One possibility is that prolonged TBX6 exposure permits accumulation of downstream cofactors that modify transcriptional output, whereas transient exposure fails to stabilise these networks. Additionally, mutually repressive interactions may occur between downstream targets as has recently been suggested for *osr1* and *foxc2* in mouse [65]. Dissecting the chromatin and network dynamics underlying this temporal encoding will be an important direction for future work. The relative resistance of our piggyBac system to transgene silencing enabled inducible, sustained TBX6 activity, highlighting its utility for interrogating temporal transcription factor dynamics in differentiating systems, and offers advantages over recently described non-inducible transgene expression systems [66].

Our findings support a model in which FOXC1 and FOXC2 establish somite identity and stabilise it by suppressing alternative lineage programs including intermediate mesoderm. We observed upregulation of muscle genes upon dual

FOXC1/2 knockdown, which may indicate de-repression of a cardiomyocyte differentiation program and suggests that FOXC1/2 may suppress lateral plate mesoderm identity to enable somitic differentiation. This contrasts with prior studies of FOXC1/2 double knockout mice which exhibited loss of both somitic and cardiac lineages [43], suggesting possible divergence in the function of FOXC1 and FOXC2 between human and mice, and highlighting the importance of using human-specific models including hTLS to study early human developmental mechanisms. Notably, FOXC1 overexpression alone was not sufficient to drive somite differentiation. Our analysis of published ATAC data of early somite cultures revealed that FOXC motifs were co-enriched with bHLH and homeobox factors specifically at FOXC1/2-dependent genes compared to non-dependent genes. It is therefore likely that FOXC1 acts in a broader combinatorial regulatory framework to establish somite identity that may require cooperativity with bHLH transcription factors. We found that potential FOXC1/2 targets included bHLH factors such as TCF15 and HAND1 and homeobox factors including MEOX1, which could function as cooperative partners with FOXC1/2 in somite differentiation. Moreover, FOXC1 overexpression drove induction of the bHLH factor *TWIST1* and EMT-like behaviour. Thus, FOXC1-TWIST1 cooperativity may have a specific role in regulating the acquisition of a migratory, mesenchymal phenotype during sclerotome differentiation. However, FOXC1 expression is not sufficient to drive full sclerotome differentiation, and additional cues such as Hedgehog signalling are needed. The established role of FOXC1 in cancer-associated EMT [67,68] suggests that this developmental function in promoting mesenchymal transition may be co-opted in pathological contexts.

In all, our study reveals a novel regulatory mechanism governing human mesodermal diversification and highlights the utility of iPSC-derived embryoid models in combination with transgene overexpression and genetic perturbations for gaining insights into previously inaccessible developmental windows in early human embryogenesis.

Figure Legends

Figure 1 - Human trunk-like structures recapitulate progressive mesodermal diversification

(A) Schematic of the protocol used to generate human trunk-like structures (hTLS) from induced pluripotent stem cells (iPSCs), adapted from [26].

(B) Brightfield and corresponding green fluorescence images of hTLS generated from SOX2:mNeon iPSCs from day 1 to day 5 of the protocol, showing expansion of mNeon^{+} and mNeon^{-} domains. Scale bars = 300 µm.

(C and D) Immunofluorescence of hTLS showing (C) FOXC2 (magenta), SOX2 (yellow), PAX8 (cyan) expression at day 5, with DAPI as nuclear counterstain (grey), or (D) DACH2 (magenta), SOX2 (yellow), PAX8 (cyan), and ZO1 (grey) expression at day 8. hTLS are outlined with dotted white lines. Scale bars, 100 µm.

(E) Schematic of the hTLS single-nucleus RNA-sequencing (snRNA-seq) strategy. (E') UMAP representation of 13,618 hTLS cells coloured by time point, with major lineages annotated (see Supplemental Figure 1G,H). PS, pluripotent stem cells.

(F) Heatmap showing transcriptional similarity between clusters identified from hTLS snRNA-seq (days 5-7) and clusters from post-conception week 3 (PCW3) human embryo scRNA-seq data [23].

(G) UMAP representation of 8,088 cells of hTLS mesodermal cells identified from the snRNA-seq dataset (see Supplemental Figure S1M, N), coloured by unsupervised clustering and annotated based on marker expression.

(H) Alluvial plot showing the proportion of hTLS mesodermal cells at each time point assigned to clusters defined in (G).

(I) Expression of representative marker genes for clusters in (G) overlaid on UMAP plots.

(J) UMAP representation of hTLS mesodermal cells (as in G) with pseudotime trajectories inferred using Slingshot [35].

(K) Within-lineage standardized expression dynamics of the top 20 genes defining each terminal state along the pseudotime trajectories shown in (J), identified using tradeSeq endpoint differential expression [36]. Genes are ranked by the strength and consistency of their differential expression at lineage endpoints. Heatmaps show z-scored fitted expression along pseudotime within each lineage (red, higher; blue, lower). Bars indicate the maximum fitted expression of each gene within the corresponding lineage.

(L) Expression of representative genes identified in (K) (*MECOM, WT1, SULF1* and *TTN*) plotted along pseudotime for each lineage. Individual cells are shown as dots, with fitted expression curves shown as lines and coloured by lineage.

See also Figure S1.

Figure 2 - TBX6 defines early mesoderm and is critical for its diversification

(A) Schematic of bifurcation of caudal epiblast cells into mesodermal and neural-fated cells (coloured in UMAP representation of 13,618 hTLS cells underneath). (A') Genes differentially expressed between mesoderm and neural clusters (as in A), with p_val_adj < 0.05 & avg_log2FC > 0.58. Top 10 up- and down-regulated transcription factors (according to log2FC) are annotated.

(B) Immunofluorescence showing TBX6 (green), TBXT/Brachyury (yellow), and SOX2 (magenta) expression in hTLS at days 2, 3 and 4, with nuclei labelled with DAPI (blue). Putative neuromesodermal progenitors (NMPs) ($SOX2^+$, $TBXT^+$, $TBX6^-$) are outlined with a dotted line. Arrowhead indicates $TBX6^+$ mesodermal domain showing progressive loss of TBX6 expression. Scale bars, 100 µm.

(C) Schematic of the dCas9-KRAB CRISPR interference gene knockdown system [39].

(D) Schematic of experimental protocol to suppress *TBX6* expression in hTLS.

(E) RT-qPCR for *TBX6* in TBX6-dCas9-KRAB iPSCs differentiated in the hTLS protocol and harvested after 24 hours, either untreated (control) or treated with Dox from day -1 (1 µM). Dots represent data from 4 independent batches. Statistical significance was

assessed using two-way ANOVA on log-transformed values, with treatment as the column factor and differentiation batch as the row factor; p value corresponds to the main effect of treatment (** $p < 0.01$). Data are plotted as $2^{-\Delta\Delta Ct}$ for visualisation; bars represent mean ± SEM; dots represent n=4 independent differentiation batches per group.

(F) Immunofluorescence of GFP (green), SOX2 (magenta), FOXC2 (yellow), or PAX8 (cyan) expression in day 5 hTLS generated from TBX6-dCas9-KRAB iPSCs, either untreated (control, top) or treated with Dox from day -1 (1 µM, bottom). White brackets indicate $SOX2^-$, $FOXC2^+$, $PAX8^+$ cells confined to a GFP^- region. Scale bars, 100 µm.

(G-I) Quantification of number of nuclei per confocal image positive for (G) FOXC2, (H) PAX8, or (I) SOX2, in day 5 TBX6-dCas9-KRAB hTLS either untreated (control, white bars; n=20 hTLS across 2 independent batches) or treated with Dox from day -1 (1 µM, coloured bars; n=22 hTLS across 2 independent batches). Dots represent individual hTLS; bars represent median ± IQR; Mann-Whitney test.

C6, CHIR (6 µM); C3, CHIR (3 µM); Ri, ROCK inhibitor; RA, retinoic acid; MG, Matrigel.

See also Figure S2.

Figure 3 - TBX6 drives transcriptional progression toward advanced mesodermal fates

(A) Schematic of a piggyBac-insertable construct for doxycycline (Dox)-inducible overexpression of TBX6-FKBP-T2A-mCherry.

(B) Inducible TBX6 transgene overexpression strategy. A piggyBac donor plasmid encoding a Dox-inducible TBX6-FKBP-T2A-mCherry cassette with puromycin resistance was integrated into iPSCs using piggyBac transposase. Following selection, edited cells were differentiated to hTLS with or without Dox to induce transgene expression.

(C) Experimental design for *TBX6* overexpression during the first 24 hours of the hTLS protocol in piggyBac-edited iPSCs (PB-Dox-TBX6), followed by bulk RNA-seq.

(D) Brightfield and corresponding red fluorescence images of PB-Dox-TBX6 cells 24 hours after initiation of the hTLS protocol, either without Dox (control) or with Dox (0.1 µM). Scale bars, 300 µm.

(E) Volcano plot of differential gene expression following 24 h of TBX6 overexpression. Genes with padj < 0.05 and |log2FoldChange| > 0.58 (±1.5-fold) are highlighted (red, upregulated; blue, downregulated). Genes encoding the top 10 up- and down-regulated transcription factors, ranked by log2FoldChange, are shown in darker colours.

(F) Ridgeplots of gene set enrichment analysis (GSEA) of genes following TBX6 overexpression (Dox versus control), ranked by the DESeq2 Wald statistic [69], and tested against the top 50 marker genes for each annotated PCW2 human embryo scRNA-seq population from Tyser et al. [22]. Ridgeplots indicate the distribution of each population's marker genes along the ranked list; colour denotes FDR-adjusted p-values.

ITR, inverted terminal repeat; TRE3G, third-generation tetracycline-responsive promoter; cHS4, chicken hypersensitive site 4 insulator; pA, polyadenylation signal; PuroR, puromycin resistance gene; GSEA, gene set enrichment analysis; PCW2, post-conception week 2.

See also Figure S3.

Figure 4 - Duration of TBX6 activity determines commitment to a somitic mesoderm fate

(A) Experimental design to overexpress *TBX6* during the hTLS protocol in PB-Dox-TBX6 iPSCs.

(B) Schematics of piggyBac-insertable constructs for Dox-inducible overexpression of TBX6-FKBP-V-T2A-mCherry (top), or mCherry only (bottom).

(C, D) Immunofluorescence of hTLS made with (C) PB-Dox-TBX6 iPSCs, or (D) PB-Dox-mCherry iPSCs, following sustained transgene induction (1 µM Dox from day -1), showing expression of SOX2 (grey), PAX8 (cyan), FOXC2 (yellow), and the mCherry

transgene (magenta). Dotted white boxes indicate regions magnified in subsequent panels. Scale bars: main, 100 µm; insets, 20 µm.

(E) Schematic of the Dox-TBX6-dTag system. Dox induces TBX6-FKBP-T2A-mCherry expression, while subsequent dTag treatment selectively depletes FKBP-fused TBX6 protein. mCherry fluorescence is retained, enabling lineage tracing of cells that experienced transient TBX6 overexpression.

(E') Experimental design to test various lengths of TBX6 protein exposure in 2D- or 3D-cultured hTLS followed by RT-qPCR or IF analysis.

(F) RT-qPCR for TBX6, or somitic mesoderm markers *FOXC1* and *MEOX1* in 2D-cultured hTLS generated from PB-Dox-TBX6 iPSCs either untreated (Ctl (control), white bars), treated with Dox only from day -1 (red bars), or treated with Dox from day -1 plus dTag added either at day 0, day 1 or day 2 (blue bars). . Statistical significance was assessed using two-way ANOVA on log-transformed values, with treatment as the column factor and differentiation batch as the row factor; p value corresponds to the main effect of treatment (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$). Data are plotted as $2^{-\Delta\Delta Ct}$ for visualisation; bars represent mean ± SEM; dots represent mean of technical replicates from n=5 independent differentiation batches per group.

(G) Immunofluorescence of mCherry (magenta), FOXC2 (yellow), PAX3 (cyan) and SOX2 (blue) in representative day 5 PB-Dox-TBX6 hTLS. Top row: hTLS treated with Dox only from day -1 (7 days of TBX6 activity). Bottom row: hTLS treated with Dox from day -1 followed by dTag from day 0 (24 h TBX6 activity). Dotted white boxes indicate regions of mCherry-expressing cells shown at higher magnification; an mCherry^{+} FOXC2^{-}, PAX3^{+}, SOX2^{+} epithelial structure is outlined in the bottom row. Scale bars: main, 100 µm; insets, 20 µm.

(H) Immunofluorescence of mCherry (magenta), FOXC2 (yellow) and PAX8 (blue) expression in representative day 5 PB-Dox-TBX6 hTLS. Top row: hTLS treated with Dox from day -1 followed by dTag from day 0 (24 h of TBX6 activity). Bottom row: hTLS treated with Dox from day -1 followed by dTag from day 2 (72 h of TBX6 activity). Dotted white boxes indicate mCherry-expressing structures shown at higher magnification;

mCherry^{+} epithelial structures are outlined with dotted lines in the insets. Scale bars: main, 100 µm; insets, 20 µm.

See also Figure S4.

Figure 5 – FOXC1 and FOXC2 establish and safeguard somitic mesoderm identity

(A) UMAP of 8,088 mesodermal subset cells, with the somite progenitor cluster (as in Figure 1G) highlighted.

(B) Volcano plot of genes differentially expressed in the somite progenitor cluster compared with the remaining mesodermal subset. Genes with padj < 0.05 and |log2FoldChange| > 0.58 (±1.5-fold) are highlighted (red, upregulated; blue, downregulated). The top 20 upregulated transcription factors according to log2FC are shown in darker colour. Box highlights FOXC1 and FOXC2.

(C) Inducible dual FOXC1 and FOXC2 CRISPRi strategy. Two piggyBac donor constructs each encoding a Dox-inducible dCas9-KRAB-IRES-GFP cassette and a single gRNA targeting either FOXC1 (with PuroR), or FOXC2 (with NeoR) were co-integrated into iPSCs using piggyBac transposase. Dual antibiotic selection yields a single cell line harbouring both cassettes. Dox treatment induces dCas9-KRAB expression, enabling simultaneous knockdown of *FOXC1* and *FOXC2* during somitic mesoderm differentiation.

(D) Experimental design to test dual knockdown of *FOXC1* and *FOXC2* in a 2D somitic mesoderm differentiation protocol (adapted from Loh et al., [44]), followed by bulk RNA-seq.

(E) Volcano plot of differential gene expression following dual knockdown of *FOXC1* and *FOXC2* in 2D somitic mesoderm differentiation. Genes with padj < 0.05 and |log2FoldChange| > 0.58 (±1.5-fold) are highlighted (red, upregulated; blue, downregulated). Selected genes associated with somitic, kidney, and muscle identities are highlighted in darker colours. (Somitic: *RIPPLY1, SIX1, FOXC1, FOXC2, MEOX1, MEOX2, TCF15, PAX3*; kidney: *OSR1, NPHS1, NPHS2, EYA2, PAX2, EPCAM, CDH1,*

CDH6, TWIST1, JAG1, HEYL, HES5; muscle *MYMX, MYH11, MYBPC2, MYLK3, MYH7B, MYOF*).

(F) Dot plot showing enrichment of transcription factor motif families identified in regulatory regions of FOXC1/2-dependent genes versus expressed non-dependent genes (defined in (E)), using early somite ATAC-seq data [44]. Dot position indicates the number of significantly enriched motifs per family (BH-adjusted $q < 0.01$); dot size reflects total motifs representation in HOMER; colour denotes the significance ($-\log_{10} q$) of the most enriched motif in each family.

(G) Experimental design to test dual knockdown of *FOXC1* and *FOXC2* in hTLS.

(H-J) Quantification of the number of nuclei per confocal image that are (G) $SOX2^+$, (H) $PAX8^+$, or (I) $PAX3^+SOX2^-$, in FOXC1/2-dCas9-KRAB hTLS either untreated (control, white bars) or treated with Dox from day -1 (Dox, coloured bars) (n=16 hTLS across 2 independent batches per group). Dots represent individual hTLS; bars represent median ± IQR; Mann-Whitney test (*** $p < 0.001$).

(K) Immunofluorescence of hTLS generated from FOXC1/2-dCas9-KRAB iPSCs and harvested at day 5, either untreated (control, top) or treated with Dox from day -1 (1 µM, bottom), showing GFP (green), SOX2 (yellow), PAX3 (magenta) and PAX8 (grey). Dotted white lines outline $PAX3^+SOX2^+$ neural structures, white arrowheads indicate $PAX3^+SOX2^-$ somitic tissue depleted upon Dox treatment. Scale bars, 100 µm.

ITR, inverted terminal repeat; TRE3G, third-generation tetracycline-responsive promoter; cHS4, chicken hypersensitive site 4 insulator; pA, polyadenylation signal; PuroR, puromycin resistance gene; NeoR, neomycin resistance gene.

See also Figure S5.

Figure 6 - FOXC1 drives a partial sclerotome differentiation program

(A) Experimental design to test dual FOXC1 and FOXC2 knockdown in a 2D sclerotome differentiation protocol (adapted from Loh et al., [44]), followed by RT-qPCR.

(B-D) RT-qPCR for B) early somite markers, C) FOXC1 and FOXC2, and D) dermomyotome and sclerotome markers and canonical Hedgehog target genes, in FOXC1/FOXC2-dCas9-KRAB iPSCs differentiated to early somite and then either untreated (control) or treated from day 3 with doxycycline (Dox, B-D), or also with smoothened agonist (SAG), or combined Dox and SAG treatment (C,D). Two-way ANOVA on log-transformed values, with treatment as the column factor and differentiation batch as the row factor. p values correspond to the main effect of treatment (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$). Data are plotted as $2^{-\Delta\Delta Ct}$; bars represent mean ± SEM; dots represent mean of technical replicates from n=5 independent differentiation batches per group.

(E) Schematic of the FOXC1 overexpression strategy

(F) Schematic of experimental design to overexpress *FOXC1* in 2D-cultured hTLS for subsequent RT-qPCR.

(G) Schematic of FOXC1 protein showing domain architecture, together with RT-qPCR analysis of *FOXC1* transgene, *TWIST1*, *MEOX1*, *PAX9, TWIST2* and *SOX9* expression in hTLS derived from PB-Dox-FOXC1 iPSCs at day 5, either untreated (control) or treated with Dox (0.1 µM) from day 2. Statistical significance was assessed using two-way ANOVA on log-transformed values, with treatment as the column factor and differentiation batch as the row factor. The p values shown correspond to the main effect of treatment (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$). Data are plotted as $2^{-\Delta\Delta Ct}$ for visualisation; bars represent mean ± SEM; dots represent mean of technical replicates from n=4 independent differentiation batches per group. N-TAD, N-terminal transactivation domain; C-TAD, C-terminal transactivation domain.

(H) Schematic of FOXC1 [ΔFKH] protein showing domain architecture with FKH DNA-binding domain replaced with a flexible linker; plots show RT-qPCR analysis for transgene and TWIST1 in day 5 2D-hTLS generated from iPSCs harbouring Dox-inducible FOXC1 [ΔFKH] and treated with varying concentrations of Dox (0, 0.01, 0.1, 1 µM) from day 2:. N-TAD, N-terminal activation domain; FKH, forkhead domain; C-TAD,

C-terminal activation domain; grey box, putative inhibitory domain. Statistical significance was assessed using two-way ANOVA on log-transformed values (ΔCt), with treatment as the column factor and differentiation batch as the row factor. The p values shown correspond to the main effect of treatment (** $p < 0.01$, **** $p < 0.0001$). Data are plotted as $2^{-\Delta\Delta Ct}$; bars represent mean ± SEM; dots represent mean of technical replicates from n=2-3 independent differentiation batches per group.

(I) Experimental design to overexpress FOXC1 in hTLS

(J) Brightfield and corresponding red fluorescence images of hTLS generated from PB-Dox-FOXC1 iPSCs at days 3, 4, and 5. Scale bars, 300 µm.

(K) Immunofluorescence of hTLS generated from PB-Dox-FOXC1 iPSCs following FOXC1 overexpression from day 2, stained for mCherry (magenta) and F-actin (phalloidin, grey). White dotted box indicates region shown at higher magnification, in which individual epithelial cysts are outlined with dotted lines. Arrows indicate FOXC1-overexpressing cells occupying a subepithelial position. Scale bars: main, 100 µm; insets, 20 µm.

(L) Immunofluorescence of hTLS generated from PB-Dox-FOXC1 iPSCs following FOXC1 overexpression from day 2, stained for mCherry (magenta), PAX3 (cyan), and SOX2 (yellow). Arrows indicate FOXC1-overexpressing cells lacking both PAX3 and SOX2. Scale bars: main, 100 µm; insets, 20 µm.

(M) Immunofluorescence of hTLS generated from PB-Dox-FOXC1 iPSCs following FOXC1 overexpression from day 2, stained for mCherry (magenta), PAX8 (grey), and TWIST1 (green). Arrows indicate FOXC1-overexpressing cells expressing TWIST1. Scale bars: main, 100 µm; insets, 20 µm.

(N) Immunofluorescence of hTLS generated from PB-Dox-FOXC1[ΔFKH] iPSCs following overexpression from day 2, stained for mCherry (magenta), PAX3 (cyan), and F-actin (phalloidin, grey). White dotted box indicates region shown at higher magnification, in which individual epithelial cysts are outlined with dotted lines. Arrows indicate FOXC1[ΔFKH]-overexpressing cells within epithelial cysts. Scale bars: main, 100 µm; insets, 20 µm.

See also Figures S6 and S7.

Supplementary Figure 1. snRNA-seq quality control and additional immunofluorescence analyses (related to Figure 1)

(A) Immunofluorescence of day 5 hTLS generated from SOX2::mNeon reporter iPSCs showing endogenous mNeon signal (green), and staining for SOX2 (grey), FOXC2 (magenta) and PAX8 (cyan) expression. A $SOX2^+$ neural tube-like structure is outlined with a dotted line. mNeon expression is specific to $SOX2^+$ cells within the neural tube-like structure but does not label all $SOX2^+$ cells. Scale bars, 100 µm.

(A') Representative hTLS from two independent differentiation batches generated using SOX2::mNeon iPSCs, shown in brightfield and corresponding green fluorescence to illustrate intra- and inter-batch variability.

(B) Immunofluorescence of FOXC2 (red), PAX6 (white) and SOX2 (green) expression in hTLS at day 5, with nuclei counterstained with DAPI (blue). Arrows indicate $SOX2^+$ $PAX6^+$ neural tube-like structures; arrowheads indicate $FOXC2^+$ epithelial somite-like structures. Scale bars, 100 µm.

(C) Immunofluorescence of the nephron markers PAX8 (green), ECAD/CDH1 (yellow), and HNF1B (magenta) in a day 8 hTLS. hTLS is outlined with dotted white line. Scale bar, 100 µm.

(D) Immunofluorescence of PAX8 (cyan), SOX2 (yellow), and FOXC2 (magenta) in a day 5 hTLS derived from an independent iPSC line, IMR-90-4. Nuclei are counterstained with DAPI, grey. hTLS is outlined with dotted line. Scale bars, 100 µm.

(E) Violin plots showing the number of detected genes per cell at each time point.

(F) Violin plots showing the percentage of mitochondrial reads per cell at each time point.

(G) UMAP representation of 13,618 hTLS cells coloured by unsupervised clustering with annotations determined by marker gene expression.

(H) Dot plot showing expression of selected marker genes for each cluster in (G).

(I) UMAP plots of 13,618 hTLS cells as in (G), with expression of selected marker genes overlaid.

(J) Immunofluorescence of PAX3 (cyan), SOX2 (yellow), and the neural crest marker SOX9 (magenta) in hTLS at day 8. White box indicates magnified region. Dotted white line in the insets outline a $PAX3^{+}SOX2^{+}$ neural tube-like structure; arrowheads indicate delaminating $SOX9^{+}$ neural crest cells. Scale bars: main image, 100 µm; insets, 20 µm.

(K) Immunofluorescence of the neural/somite marker PAX3 (cyan), the neural marker SOX2 (yellow) and the endothelial marker PECAM-1/CD31 (magenta), in hTLS at day 5. White box indicates magnified region. Arrowheads indicate a $PECAM\text{-}1^{+}$ vessel-like structure nestled between $SOX2^{+}$ and $PAX3^{+}$ neural/somitic tissue. Scale bars: main image, 100 µm; insets, 20 µm.

(L, L', L'') Transcriptional similarity between hTLS snRNA-seq dataset and human reference embryogenesis dataset from Zhao et al., [24]. UMAP representations of human reference embryogenesis dataset coloured by (L) time point and (L') cell type using annotations from the reference. (L'') projection of hTLS snRNA-seq cells (iPSC and days 0-4) onto reference UMAP, coloured by time point, with reference cells in grey.

(M) UMAP of hTLS cells as in (G), coloured by unsupervised clustering at resolution = 0.03 to identify mesodermal cells (cluster 0), which were subsequently sub-clustered to isolate the mesodermal subset.

(N) Heatmap of top 10 marker genes defining clusters in (M). Enriched genes were first determined with Wilcoxon rank sum test with thresholds log2FoldChange >1 and padj < 1E8.

(O) UMAP of mesodermal subset (as in Figure 1G) coloured by time point.

(P) Heatmap showing top 50 genes defining each cluster of the mesodermal subset annotated as in Figure 1G, with selected marker genes presented. Enriched genes were first determined with Wilcoxon rank sum test with thresholds log2FoldChange >1 and padj < 0.005.

Supplementary Figure 2. TBX6 is expressed transiently throughout the mesoderm before mesoderm diversification (related to Figure 2)

(A) Experimental design to differentiate iPSCs in the hTLS protocol on glass-bottomed dishes without aggregation for subsequent immunofluorescence analysis.

(B) Immunofluorescence of differentiated iPSCs (as outlined in A) showing expression of TBX6 (green), TBXT/Brachyury (yellow) and SOX2 (magenta) expression, with nuclei counterstained with DAPI (blue). Scale bars, 100 µm.

(C) Strategy to differentiate iPSCs using the hTLS protocol followed by dissociation then reseeding at day 0 to generate hTLS pools maintained in 2D culture for subsequent RT-qPCR analysis.

(D) Time-course RT-qPCR analysis of selected of early mesoderm, somite, kidney and neural markers during the hTLS protocol. Dots and bars represent mean ± SEM; n = 4 wells of hTLS per time point across 2 independent differentiation batches.

(E) RT-qPCR analysis of *TBXT* and *MSGN1* expression in TBX6-dCas9-KRAB iPSCs differentiated using the hTLS protocol and harvested after 24 hours (as in Figure 2E), either untreated (control) or treated with Dox (1 µM) from day -1. Dots represent n=4 independent differentiation batches. Two-way ANOVA on log-transformed values (ΔCt), with treatment as the column factor and differentiation batch as the row factor. P values correspond to treatment effect (** $p < 0.01$). Data are plotted as $2^{-\Delta\Delta Ct}$; bars represent mean ± SEM.

(F) Immunofluorescence of TBX6-dCas9-KRAB iPSC cells differentiated as in (A) either without or with Dox treatment from day -1 to induce TBX6 knockdown, showing expression of TBX6 (magenta), and F-actin (grey) to outline cells, with GFP (green) as a readout for successful knockdown expressed heterogeneously after Dox treatment. Dotted box indicates magnified region. Arrowheads indicate expression of TBX6 in cells lacking GFP and which thus did not knockdown TBX6. Scale bars main images 100 µm; insets, 20 µm.

(G) Immunofluorescence of day 5 hTLS generated from a second independent TBX6-dCas9-KRAB iPSC line with a distinct gRNA to target TBX6, either with no treatment (control, top) or following Dox treatment from day -1 (1µM Dox, bottom), showing GFP (green), SOX2 (magenta), FOXC2 (yellow), and PAX8 (cyan). White brackets indicate $FOXC2^{+}PAX8^{+}$ cells confined to a GFP^{-} region. Scale bars, 100µm.

Supplementary Figure 3. TBX6 RNA-seq quality control and additional analysis (related to Figure 3)

(A) RT-qPCR analysis of *TBX6, SOX2,* somitic mesoderm markers *FOXC1* and *FOXC2,* and the intermediate mesoderm marker *OSR1,* in PB-Dox-TBX6 iPSCs differentiated using the hTLS protocol and harvested after 24 h, either untreated (control) or treated with Dox (0.1 µM) from day -1. Dots represent n=5 independent differentiation batches. Two-way ANOVA on log-transformed values (ΔCt), with treatment as the column factor and differentiation batch as the row factor. P values correspond to treatment effect (** $p < 0.01$, **** $p < 0.0001$). Data are plotted as $2^{-\Delta\Delta Ct}$; bars represent mean ± SEM.

(B) Principal component plots of TBX6 overexpression bulk RNA-seq data.

(C) Sample-to-sample correlation heatmap of the TBX6 overexpression bulk RNA-seq dataset.

(D) Dot plot showing gene ontology (GO) biological processes (BP) analysis, highlighting the top 15 processes upregulated following TBX6 overexpression. Dot size indicates gene count; colour denotes FDR-adjusted p-value.

(E) Ridge plots of gene set enrichment analysis (GSEA) of genes following TBX6 overexpression (Dox versus control), ranked by the DESeq2 Wald statistic [69], and tested against the top 50 marker genes for each annotated PCW3 human embryo scRNA-seq population from Zeng et al. [23]. Ridge plots indicate the distribution of each population's marker genes along the ranked list; colour denotes FDR-adjusted p-values.

(F) UMAP of scRNA-seq data from PCW2 (CS7) human embryo from Tyser et al. [22], coloured by cell type.

(G) UMAP of TBX6 expression in the Tyser et al. dataset.

(H) Projection of TBX6-upregulated gene module onto the Tyser et al. UMAP. The module was defined as the top 500 genes upregulated upon TBX6 overexpression in hTLS (padj < 0.05, log2FoldChange > 2) and filtered for overlap between the Tyser and Zeng datasets (221 genes total remaining).

(H) Violin plot showing enrichment of the TBX6-upregulated gene module across cell types in the Tyser et al. dataset, using cluster annotations as in (E).

(I) UMAP of scRNA-seq data from PCW3 (CS10) human embryo dataset from Zeng et al. [23], coloured by cell type.

(J) UMAP of TBX6 expression in Zeng et al. dataset.

(K) Projection of the TBX6-upregulated gene module (as in H) onto the Zeng et al. dataset.

(L) Violin plot showing enrichment of the TBX6-upregulated gene module across cell types in the Zeng et al. dataset, using cluster annotations as in (I).

Supplementary Figure 4. DegTrace (TBX6-FKBP-V-T2A-mCherry lineage tracing) system (related to Figure 4).

(A) Experimental design to test depletion of the TBX6-FKBP fusion protein following dTag treatment in PB-Dox-TBX6 iPSCs cultured in iPSC maintenance medium (Stemflex).

(B) Immunofluorescence of F-Actin (phalloidin, grey), mCherry (magenta) and TBX6 (green), with nuclei counterstained with DAPI (blue), in representative PB-Dox-TBX6 iPSCs cultured as in (A) and either untreated (no Dox, top row), treated with Dox for 48 h (0.1 µM, middle row), or treated with Dox for 48 h followed by dTag for 24 h (2 µM, bottom row). Scale bars, 100 µm.

(C) Experimental design to test depletion of the TBX6-FKBP fusion protein in hTLS generated from PB-Dox-TBX6 iPSCs, followed by immunofluorescence (IF) analysis.

(D) Immunofluorescence of mCherry (magenta), TBX6 (green), PAX8 (cyan), and PAX3 (grey) in representative day 5 PB-Dox-TBX6 hTLS treated as in (C), with Dox (0.1 µM) from day -1 followed by dTag (2 µM) from day 0. White boxes indicate mCherry^{+} epithelial structures shown at higher magnification; dotted lines in the insets outline the mCherry^{+} structures. The top row shows an hTLS retaining low levels of TBX6 within the mCherry^{+} structure, whereas the bottom row shows an mCherry^{+} structure with TBX6 fully depleted. Scale bars: main, 100 µm; insets, 20 µm.

Supplementary Figure 5. Dual FOXC1 and FOXC2 knockdown quality control and additional analyses (related to Figure 5)

(A) Schematic of the 2D somitic mesoderm differentiation protocol, adapted from Loh et al. [44].

(B) Time-course RT-qPCR analysis of selected marker genes during differentiation. Dots and bars represent mean ± SEM (n = 2).

(C) Schematic of piggyBac-insertable constructs for Dox-inducible dual CRISPRi-mediated knockdown of *FOXC1* and *FOXC2*.

(D) Principal component analysis of the dual FOXC1 and FOXC2 knockdown bulk RNA-seq dataset.

(E) Sample-to-sample correlation heatmap of the dual FOXC1 and FOXC2 knockdown bulk RNA-seq dataset.

(F, G) Gene-wise boxplots showing DESeq2 [69] normalized counts (log10 scale) for (F) early and paraxial mesodermal markers and (G) selected genes downregulated in somite differentiation and upregulated in kidney- or muscle-associated programs following dual FOXC1 and FOXC2 knockdown. FOXC1 and FOXC2 were among the most strongly downregulated genes (log2FC ≈ −3; FDR < 10^{-1}□□), confirming efficient repression. Dots represent biological replicates; boxplots indicate the IQR across

replicates; grey indicates control and coloured indicates Dox treatment. Asterisks denote DESeq2 adjust p-values (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

(H) Dot plot showing selected enriched gene ontology (GO) biological process (BP) terms among genes upregulated (UP) or downregulated (DOWN) following dual FOXC1/FOXC2 knockdown. Dot size indicates gene ratio; colour denotes FDR-adjusted p-value.

(I-K) Ridge plots of gene set enrichment analysis (GSEA) of genes following dual FOXC1 and FOXC2 knockdown (Dox versus control), ranked by the DESeq2 Wald statistic [69], tested against marker gene sets for (I) fetal human nephron cell types from Stewart et al. [46], (J) PCW6.5-7 human heart cell types from Asp et al. [45], and (K) PCW3-5 human embryo scRNA-seq populations from Zeng et al. [23]. Ridge plots indicate the distribution of marker genes along the ranked list; colour denotes FDR-adjusted p-values.

(L) Immunofluorescence of hTLS generated from FOXC1/FOXC2-dCas9-KRAB iPSCs and harvested at day 5, either untreated (control, top) or treated with Dox from day -1 (1 µM, bottom), showing GFP (green), SOX2 (yellow), FOXC2 (magenta) and PAX8 (cyan), illustrating depletion of FOXC2 upon dual FOXC1 and FOXC2 knockdown. Scale bars, 100 µm.

(M) Schematic of experimental design to test epistatic relationship between TBX6 and FOXC1/2 in 2D-cultured hTLS.

(N) RT-qPCR analysis of TBX6, FOXC1, FOXC2, TCF15/Paraxis, SIX1, MESP2, and OSR1 in 2D-cultured hTLS derived from PB-Dox-TBX6 or Dox-TBX6-dCas9-FOXC1/2 iPSCs at day 5, either untreated (control) or treated with Dox (0.1 µM) from day -1. Dots represent mean of technical replicates from $n = 2$ independent differentiation batches per condition. Two-way ANOVA on log-transformed values (ΔCt), with treatment as the column factor and differentiation batch as the row factor. P values correspond to main treatment effect (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$). Data are plotted as $2^{-\Delta\Delta Ct}$ for visualisation, normalised to the untreated control condition within each iPSC line; bars represent mean ± SEM.

Supplementary Figure 6. hTLS retain sclerotome differentiation potential (related to Figure 6)

(A) Schematic of sclerotome differentiation *in vivo*. Hedgehog ligands secreted by the notochord signal to the ventral domain of epithelial somites to activate sclerotome-associated genes including FOXC1, FOXC2, TWIST1, PAX1, PAX9, and SOX9, to downregulate PAX3, and to initiate epithelial-to-mesenchymal transition (EMT), while dorsal somites remain PAX3^{+}.

(B) UMAP plots of 13,618 hTLS snRNA-seq cells (as in Figure 1E) showing minimal expression of Hedgehog ligands (*SHH*, *IHH*, *DHH*), and expression of downstream Hedgehog pathway mediators (*GLI1*, *GLI2, GLI3*).

(C) Experimental design to test the effect of Hedgehog pathway activation using smoothened agonist (SAG; 500 nM) in hTLS from day 5 to day 8 of the protocol.

(D) Immunofluorescence of SOX2 (yellow), FOXC1 (grey), and PAX9 (cyan) expression in day 8 hTLS cultured as in (C), either untreated (control, top row) or treated with SAG from day 5 (500 nM, bottom row). Scale bars, 100 µm.

(E-G) Quantification of the number of nuclei per confocal image positive for (E) SOX2, (F) FOXC2, or (G) PAX3 (whilst also SOX2^{-}) in day 8 hTLS either untreated (control, white bars) or treated with SAG from day 5 (500 nM, coloured bars). Dots represent individual hTLS (n = 12 hTLS across 2 independent differentiation batches); bars represent median ± IQR; Mann-Whitney test (** $p < 0.01$, *** $p < 0.001$).

(H) Immunofluorescence of mCherry (magenta), FOXC2 (grey), and SOX9 (green) expression in day 5 hTLS generated from PB-Dox-FOXC1 iPSCs treated with Dox from day 2 (0.1 µM). White box indicates magnified region; white arrows indicate mCherry^{+} FOXC2^{+} cells expressing SOX9; arrowheads indicate a rare mCherry^{-} FOXC2^{+} SOX9^{+} cell. Scale bars: main, 100 µm; insets, 20 µm.

(I) Schematic of the Dox-inducible FOXC2-overexpression strategy and corresponding immunofluorescence of SOX2 (blue), mCherry (magenta), PAX8 (grey), and PAX3

(green) expression in a day 5 hTLS generated from PB-Dox-FOXC2 iPSCs and treated with Dox from day 2 (0.1 µM). White box indicates magnified region; the dotted lines outline an mCherry^{+}, PAX8^{-}, PAX3^{-} cell cluster positioned between PAX8^{+} and PAX3^{+} tissue. Scale bars: main, 100 µm; insets, 20 µm.

Supplementary Figure 7. Dual TWIST1 and TWIST2 knockdown in sclerotome differentiation (related to Figure 7)

(A) Schematic of a piggyBac-insertable construct for Dox-inducible overexpression of TWIST1.

(B) Experimental design to test TWIST1 overexpression in hTLS.

(C) Immunofluorescence of hTLS generated from PB-Dox-TWIST1 iPSCs showing mCherry (magenta), FOXC2 (cyan), F-actin (phalloidin, yellow) and SOX2 (blue) expression at day 5 following Dox treatment from day 2. The white box indicates magnified region. Dotted white lines in magnified panels indicate FOXC2^{+} epithelial structures, and white arrows indicate mCherry^{+} TWIST1-overexpressing cells occupying a subepithelial position. Scale bars: main, 100 µm; insets, 20 µm.

(D) Experimental design to test dual TWIST1 and TWIST2 knockdown in a 2D sclerotome differentiation protocol (adapted from Loh et al., [44]), followed by RT-qPCR.

(E) RT-qPCR analysis of FOXC1, FOXC2, TWIST1, TWIST2, PAX9, PAX1 and PTCH1 expression in TWIST1- and TWIST2-dCas9-KRAB iPSCs differentiated to sclerotome and either untreated (control), or treated from day 3 with Dox, SAG, or combined Dox and SAG treatment. Dots represent mean of technical replicates from n = 3 independent differentiation batches per condition. Two-way ANOVA on log-transformed values (ΔCt), with treatment as the column factor and differentiation batch as the row factor. P values correspond to main treatment effect (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$). Data are plotted as $2^{-\Delta\Delta Ct}$ for visualisation, normalised to the untreated control condition within each iPSC line; bars represent mean ± SEM.

1063

METHODS

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Anna Philpott ap113@cam.ac.uk.

Materials availability

All unique/stable reagents and cell lines generated in this study are available from the lead contact with a completed materials transfer agreement.

Data and code availability

Single-cell and bulk RNA-seq data have been deposited at GEO and are publicly available as of the date of publication. Accession numbers are listed in the key resources table. Microscopy data reported in this paper will be shared by the lead contact upon request. All original code has been deposited at Github and is publicly available as of the date of publication. The DOI is listed in the key resources table. Any additional information required to reanalyse the data reported in this work is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Cell Lines

Human iPSCs were cultured as previously described with modifications [70]. Human iPSCs used in this study were kindly gifted by Austin Smith [71]. iPS(IMR90)-4 cells were obtained from WiCell. All iPSCs were maintained in StemFlex (ThermoFisher, A3349401) containing Pen/Strep on Geltrex (Gibco, A14133-01)-coated plates. Cells were passaged every 3-4 days using PBS with 0.5µM EDTA and re-seeded in StemFlex containing Pen/Strep and 10 µM ROCK inhibitor Y27623 (ROCKi; Tocris, 1254). Transgenic iPSC lines (PB-Dox-TBX6, PB-Dox-FOXC1, PB-Dox-FOXC1ΔFKH, PB-Dox-TWIST1, TBX6-dCas9-KRAB, FOXC1/2-dCas9-KRAB, and TWIST1/2-dCas9-KRAB), generated as described below, were maintained in StemFlex containing

Pen/Strep, 0.5 µg/mL puromycin and/or 200 µg/mL G418 as appropriate. All cell cultures were incubated at 37°C with 7% CO_2 and 5% O_2.

Ethical Statement about use of human embryoid models

The human trunk-like structure system used in this study is a human pluripotent stem cell–derived model that although reproduces features of early post-gastrulation human developmental stages, does not generate extraembryonic tissues, and is therefore classified as a *non-integrated embryo model* under the terminology established in the 2021 International Society for Stem Cell Research (ISSCR) guidelines [72]. In addition, the model lacks multiple embryonic lineages including, but not limited to, all anterior neural tissues, lateral mesoderm, notochord, endoderm, and non-neural ectoderm. No human embryos were used in this research. This research was conducted in accordance with the 2021 ISSCR Guidelines.

METHOD DETAILS

Plasmid constructs

Target gene knockdown in iPSCs was performed as previously described with modifications [70]. pPB-Ins-U6p-sgRNAentry-EF1Ap-TetOn3G-IRES-Neo (Addgene plasmid # 183411; RRID:Addgene 183411) and pPB-Ins-TRE3Gp-KRAB-dCas9-ecDHFR-IRES-GFP-EF1Ap-Puro (Addgene plasmid # 183410; RRID:Addgene 183410) were kind gifts from Azim Surani. The SOX2:mNeon iPSC line was generated with the following plasmids: pCAS9-mCherry-Frame +0 [73] (gift from Veit Hornung, Addgene plasmid # 66939 ; http://n2t.net/addgene:66939 ; RRID:Addgene_66939), pCRISPaint-mNeon-PuroR [M1G] [74] (gift from Michael Hoelzel, Addgene plasmid # 171800 ; http://n2t.net/addgene:171800 ; RRID:Addgene_171800), and pSPgRNA [75] (gift from Charles Gersbach, Addgene plasmid # 47108 ; http://n2t.net/addgene:47108 ; RRID:Addgene_47108) containing gRNA targeting the human SOX2 locus (TGCCCCTCTCACACATGTGA). All gRNA oligos used were ordered using the custom DNA oligo synthesis service from Merck and are listed in the ‘‘Oligonucleotides’’ section of the key resources table. Specific gRNA sequences for dCas9-KRAB mediated gene

repression were designed by identifying putative transcription start sites (TSSs) for target genes according to FANTOM CAGE data publicly available on ZENBU [76], then the Wellcome Sanger Institute Genome Editing (WGE) CRISPR Finder [77] was used to identify CRISPR sequences 50-100bp downstream of a putative TSS for optimal gene knockdown [78]. Two gRNA sequences were tested per TSS. For generation of gRNA expression vectors against TBX6, complementary gRNA oligos were annealed and ligated into pPB-Ins-U6p-sgRNAentry-EF1Ap-TetOn3G- IRES-Neo (Addgene plasmid, 183411) by Esp3I digestion (NEB, R0734) followed by ligation with T4 DNA ligase (New England Biolabs, M0202S). Piggybac-based Dox-inducible transgene overexpression plasmids were created by first PCR amplifying (with Cloneamp Takara, 639298) the Dox-response element from pPB-Ins-TRE3Gp-KRAB-dCas9-ecDHFR-IRES-GFP-EF1Ap-Puro (Addgene plasmid # 183410) and the gRNA-rtTA expression cassette from pPB-Ins-U6p-sgRNAentry-EF1Ap-TetOn3G-IRES-Neo (Addgene plasmid, 183411), and cloning into VB240515-1632gka backbone (Vectorbuilder) by digesting with EcoRV (NEB, R3195) and PacI (NEB, R0547) and inserting using In Fusion Snap Assembly master mix (Takara, 638948). This yielded Dox-inducible dCas9-KRAB vectors that were used for knockdown of FOXC1/2 and TWIST1/2 factors; versions with Puro and Neo resistance were generated to combine 2 distinct gRNA-expressing plasmids into the same cells. For Dox-inducible expression of transgenes of interest, the dCas9-KRAB-U6-gRNA cassette was excised with NotI (NEB, R3189) and AsiSI (NEB, R0630) and transgenes were cloned in by In Fusion. FOXC1 and TWIST1 transgenes were synthesised by Twist Biosciences. mCherry was synthesised by Vectorbuilder (VB240515-1632gka). TBX6 was PCR amplified from TFORF1979 (Addgene 142303, kind gift from Feng Zhang). FKBP-V was PCR amplified from pBS_GGSG_Fkbp_2xHA_BB2 (Addgene 235503, kind gift from Brian Hendrich). PCR amplified fragments were gel extracted using Monarch gel extraction kit (NEB, T1120). Stellar competent E. coli (Takara, 636763) were used for transformation. Successful transformants were purified using QIAprep Spin Miniprep Kit (Qiagen, 27104) or QIAprep Plasmid Plus Midiprep Kit (Qiagen, 12943), then validated with Nanopore sequencing (Oxford Nanopore Technologies) by Plasmidsaurus.

Quantitative reverse transcriptase PCR (RT-qPCR)

To isolate RNA for RT-qPCR analysis, cells were washed twice with PBS and then lysed in 350 µl RLT lysis buffer prior to RNA extraction using the RNeasy mini kit (Qiagen, 74106) with on-column DNase digestion (Qiagen, 79254) and reverse transcribed to cDNA using the QuantiTect Reverse Transcription kit (Qiagen, 205314). Quantitative PCRs were performed using PowerUp SYBR Green Master Mix (Applied Biosystems, A25742) with forward and reverse primers diluted to final concentration of 1 µM per reaction and run either on the Quantstudio 5 384-well qPCR machine (Applied Biosystems) or the StepOne Plus with 96-well block (Applied Biosystems). Statistical analyses were performed on ΔCt values (Ct_target − Ct_reference). Data were converted to $2^{-\Delta\Delta Ct}$ values for visualisation [79]. Details of statistical tests used are described in the figure legends. RT-qPCR data were analysed using Prism (Graphpad, v11.0.0).

Transgenic cell lines

TBX6-dCas9-KRAB, FOX1/2-dCas9-KRAB, TWIST1/2-dCas9-KRAB, PB-Dox-FOXC1, PB-Dox-TBX6, PB-Dox-FOXC1, and PB-Dox-TWIST1, transgenic lines were established as previously described with modifications [70]. 6.5×10^5 iPSCs were seeded into Geltrex-coated 6-well plates in StemFlex with ROCKi, and the following day were washed with PBS and incubated in 1ml Optimem media (Gibco, 31985070) with Revitacell supplement (Fisher, 15317447). For each well of a 6-well plate, cells were transfected with Super piggyBac Transposase expression vector (Stratech, PB210PA-1-SBI) (500 ng), together with the donor plasmid(s) to a total of 2 µg plasmid DNA mixed in 3.5 µl Lipofectamine Stem transfection reagent (Life Technologies, STEM00003) in 400 µl Opti-MEM with Revitacell. DNA-lipofectamine mixtures were incubated at room temperature for 20 minutes and then added dropwise to cells, which were then incubated at 37°C for 4 hours. Media were then replaced with 3 ml StemFlex with Revitacell. The following day, media were refreshed with StemFlex with Revitacell. 48-72 hours later, media were refreshed with fresh StemFlex with Revitacell containing antibiotics 0.5 µg/mL puromycin and/or 200 µg/mL G418 for selection. Media were refreshed with selection media every 2-3 days thereafter. A week later, surviving

colonies were pooled, and knockdown efficiency was assessed by addition of 1µM Dox (a concentration we found in previous studies to be optimal [70]) followed by RT-qPCR with primers specific to the target gene.

To make the SOX2:mNeon CRISPR-edited iPSC line, were transfected as described above, using the following plasmids: pCAS9-mCherry-Frame +0 [73], pCRISPaint-mNeon-PuroR [M1G] [74], and pSPgRNA [75] containing gRNA targeting the human SOX2 locus (TGCCCCTCTCACACATGTGA). 72 hours after transfection, cells were selected with 0.5 µg/mL puromycin, and successful colonies were isolated, with mNeon expression confirmed by fluorescence microscopy.

Human trunk-like structure generation

Human trunk-like structures (hTLS) were generated using a previously described protocol with modifications [26]. Human iPSCs were grown to 70% confluency, washed with PBS containing 0.5µM EDTA and dissociated with Accutase (Day -2), counted using Countess II (Life Technologies, AMQAX1000), then seeded evenly onto Geltrex-coated 6-well plates at 400,000 cells per well in 2ml Nutristem hPSC XF medium (Sartorius, 05-100-1A) with ROCKi. The following day (Day -1), cells were incubated in Nutristem with 6µM CHIR (ABCR, AB253776). The next day (Day 0), cells were washed with PBS with EDTA, then dissociated with Accutase, centrifuged at 300g and resuspended in Essential 6 media (E6; Thermo, A1516401) containing 3µM CHIR and 500nM retinoic acid (RA), and counted. For 3D hTLS, 400 cells were seeded in 50µl of E6 with 3µM CHIR and 500nM RA per well of a round-bottom low-adhesion 96 well plate (Corning, CLS7007); the plate was then centrifuged at 200g for 2 minutes and incubated at 37°C to promote aggregation. By the following day (Day 1) cell aggregates were visible by light microscopy at 4x magnification, at which point 150 µl E6 was added to each well gently to avoid disrupting the aggregates. On Day 2, 150 µl media was carefully removed from each well, with care taken to avoid aspirating the aggregates, and 150 µl E6 containing 10% Matrigel (Corning, 356231) and 10 nM RA was gently added to each well, with Dox added as indicated. Plates were incubated at 37°C until Day 5, by which point the aggregates had elongated into trunk-like structures. To

harvest for immunofluorescence analysis, structures were gently aspirated using a glass Pasteur pipette and fixed in 3% PFA for 30 minutes as described below. Alternatively, if the experiment was continued, 50 µl E6 was added to each well together with any indicated treatment, and hTLS were cultured further until harvest. For 2D hTLS, on Day 0 after dissociation and counting, $4.8x10^4$ cells were seeded into Geltrex-coated 24-well plates in 500 µl of E6 with 3µM CHIR and 500nM RA per well. The following day (Day 1) 375 µl medium was removed from each well and replaced with 375 µl E6 only. On Day 2, 375 µl medium was removed and replaced with 375 µl E6 containing 10% Matrigel and 10 nM RA. Plates were incubated at 37^oC, then on Day 5 wells were washed twice with PBS then lysed for RNA extraction. The 2D hTLS protocol mirrors the 3D protocol in terms of treatment conditions with the exception that instead of growing as individual 3D structures, the 2D hTLS adhere to the 24 well plate and interconnect. Both protocols show similar marker gene expression dynamics, with the emergence of neural ($SOX2^+$) cells, somitic mesoderm ($FOXC2^+$) and renal mesoderm ($PAX8^+$) identities.

Immunofluorescence analysis

To harvest for immunofluorescence (IF) analysis, hTLS were transferred using a glass Pasteur pipette into a 30 ml sterilin, were centrifuged at 200g for 1 minute, washed once with PBS, then fixed in 3% PFA for 30 minutes at room temperature with gentle rotation, and then washed in PBS (3x10 min) and stored in PBS with 0.01% sodium azide at 4^oC. Prior to antibody staining, samples were incubated in blocking buffer overnight at 4^oC (PBS containing 0.25% Triton-X, 5% BSA and 0.01% sodium azide). Samples were incubated in primary antibodies diluted in wash buffer (PBS containing 0.1% Triton-X, 0.1% BSA and 0.01% sodium azide) at the indicated concentrations overnight at 4^oC. Next, samples were washed in wash buffer (3x15 min at room temperature) and incubated in secondary antibodies, with DAPI (abcam, ab228549) where indicated, diluted in wash buffer at the indicated concentrations overnight at 4^oC. The following day, samples were washed once in wash buffer and mounted on glass slides in Vectorshield (Vector Labs, H1000) in 120mm spacers (Invitrogen, S24735) with coverslips overlaid.

Imaging and image analysis

Brightfield images were taken using an EVOS M5000 microscope (Life Technologies, AMF5000SV). Confocal images were taken using a Zeiss LSM 710 or 880. Image files were analysed using Zeiss Zen (3.12) or Fiji (Image J). All immunofluorescence images presented are representative confocal slices from at least 2 independent hTLS differentiation batches. To quantify number of cells expressing a given marker, single confocal images were analysed in CellProfiler using a custom analysis pipeline [80]. Briefly, positively stained nuclei were identified using a thresholding and declumping module (IdentifyPrimaryObject module with Otsu two-class adaptive thresholding). Smoothing factors were determined empirically. Objects (nuclei) were filtered by size, and the number of objects within that channel was measured.

2D somitic mesoderm differentiation

Human iPSCs were differentiated to somitic mesoderm using a previously described protocol with modifications [44]. Human iPSCs were grown to 70% confluency, washed with PBS containing 0.5µM EDTA and dissociated with Accutase (Day -1), counted using Countess II (Life Technologies, AMQAX1000), then seeded evenly onto Geltrex-coated 24-well plates at 50,000 cells per well in 500 µl Nutristem hPSC XF medium (Sartorius, 05-100-1A) with ROCKi. The following day (Day 0), cells were incubated in Nutristem with 6µM CHIR (ABCR, AB253776) and 100 nM PIK90 (Tocris, 7902). The next day (Day 1, primitive streak stage) media were refreshed with E6 containing 3 µM CHIR, 20 ng/ml FGF2, 1 µM A-83-01 (Tocris 2939) and 250 nM LDN 193189 (Tocris 6053). On day 2 (paraxial mesoderm stage), media were refresh with E6 containing 1 µM A-83-01, 250 nM LDN 193189, 1 µM XAV 939 (Tocris 3748) and 500 nM PD0325. On day 3 cells were early somite stage, at which point cells were harvested for RNA extraction of bulk RNA-seq. Alternatively, for sclerotome differentiation cells media were refreshed with E6 containing 500 nM SAG (Tocris 4366) and cells were cultured for a further 72 hours. For dCas9-KRAB knockdown experiments, Dox was added either at day 1 or day 3 as indicated.

Antibodies

Primary antibodies used for protein detection, with their corresponding dilutions for immunofluorescence (IF) were as follows: rabbit anti-PAX8 [EPR13511] (Abcam, ab189249, 1:200), sheep anti-FOXC2 (R&D #AF6989, 1:200), rat anti-SOX2 (Invitrogen #14-9811-82, 1:200), goat anti-DACH2 (R&D #AF5230, 1:100), rabbit anti-ZO1 (Cell Signaling #13663, 1:200), rabbit anti-PAX6 (abcam #195045, 1:200), goat anti-TCF-2/ HNF-1 beta (R&D Systems, AF3330, 1:200), mouse anti-E-Cadherin (BD Bioscience, 610181, 1:200), mouse anti-PAX3 (R&D #MAB2457, 1:100), mouse anti-PAX3 (DSHB #PAX3-c, 1:100), rabbit anti-SOX9 (Sigma #AB5535, 1:200), goat anti-PECAM-1 (R&D #AF3628, 1:200), goat anti-TBX6 (R&D #AF4744, 1:200), rabbit anti-TBXT (Cell Signaling #81694, 1:200), Phalloidin-647 (Thermo #A22287, 1:400), RFP-568 booster (Proteintech #rb2AF568-50, 1:100), and mouse anti-Twist1 (abcam #ab50887, 1:100). Alexa Fluor 405-, 488-, 568- and 647-conjugated secondary antibodies (Invitrogen) were used for detection of primary antibodies.

Bulk RNA-seq

For bulk RNA-sequencing experiments, at least ~100K cells per condition were collected at the indicated time points by washing once with PBS with EDTA, dissociation with Accutase, centrifugation at 300g for 3 minutes, and then lysis in 50 µl DNA/RNA Shield (Zymo R1100-50). RNA-Seq was performed by Plasmidsaurus using Illumina Sequencing Technology with the in-house custom analysis and annotation pipeline, and the resulting counts matrices were used for downstream analysis.

Bulk RNA-seq analysis

Counts matrices were imported into R Studio (version 2024.09.1+394) running R version 4.4.2, and differential gene expression analysis between control and Dox-treated conditions was analysed by DESeq2 [69]. The significance threshold for differential expression was set as log2FoldChange > 0.58 (corresponding to fold change

> 1.5) and adjusted p value < 0.05. Transformed, variance-stabilised expression values were used for principal component analysis and sample-to-sample correlation heatmaps. Plots for individual genes used normalised counts. Gene ontology (GO) analysis was performed with the enrichGO function from clusterProfiler [81].

Single-nucleus RNA-seq library preparation

For single-nucleus RNA-sequencing experiments, cell pellets or hTLS containing 0.5-1M cells harvested at the indicated time points were snap frozen. Pellets were thawed and resuspended in 1ml of ice-cold nuclei permeabilization buffer (10mM Tris-HCL pH7.4, 10mM $NaCl_2$, 3 mM $MgCl_2$, 0.2% IGEPAL CA300, 0.1mg/ml BSA, 1x protease inhibitor cocktail (Roche cat.11873580001), 0.5U/µl ea Superase RNAse Inhibitor and RNAaseOUT). Cells were then permeabilized for 5 min and centrifuged at 600g for 5 min at 4˚C. Cells were lightly fixed by resuspending in 475µl fixation buffer (1xPBS, 0.2% formaldehyde, 0.5U/µl ea Superase RNAse Inhibitor and RNAaseOUT) and incubated on ice for 5 min. Fixation was stopped by adding 2µl 1M Tris-HCl pH7.4 containing 0.1mg/ml BSA. Nuclei were pelleted at 600g for 5 min at 4˚C and resuspended in 1xPBS containing 1x protease inhibitor cocktail and 0.5U/µl ea Superase RNAse Inhibitor and RNAaseOUT.

For in situ reverse transcription (RT) reaction 50,000 cells for each sample were aliquoted in 0.2 ml PCR tubes containing 2.5 µM sample specific barcoded primers as previously described [82]. To each tube RT mix (1x Maxima H buffer, 0.5 mM dNTPs, 0.1mg/ml BSA, 400U Maxima Reverse H Minus Reverse Transcriptase (Thermo Scientific) and 0.5 U/ul ea Superase RNAse Inhibitor and RNAaseOUT) was added to a final volume of 20 µl. The reverse transcription was performed in a thermocycler with the following program: Step 1: 50°C × 10 min; Step 2 (3 cycles): 8°C × 12 s, 15°C × 45 s, 20°C × 45 s, 30°C × 30 s, 42°C × 2 min, 50°C × 5 min; Step 3: 50°C × 10 min. After the reaction, the nuclei were transferred and pooled on ice into a 1.5-ml tube prewashed with 5% BSA in 1x PBS. Then Triton-X-100 was added to a final percent of 0.075 and nuclei were pelleted at 800g for 5 min at 4˚C.

For ligation-based combinatorial barcoding the pooled nuclei were resuspended in 1 ml of 1x NEBuffer3.1 (NEB) and transferred to 2,910 µl of ligation buffer (1x T4 ligase buffer (NEB), 0.35mg/ml BSA, 25 mM $NaCl_2$, 40,000U T4 ligase (NEB), and 0.25 U/µl ea Superase RNAse Inhibitor and RNAaseOUT). Using a multichannel pipette, 40 µl of the ligation mix was distributed to the 96 wells of the first R02 barcode plate [82]. The plate was sealed and placed on a Thermomixer (Eppendorf) and mixed for 30 sec at 37˚C at 1200rpm. Then incubated for a further 30 min at 37˚C at 300rpm. After incubation the plate was removed and 10 µl of R02 blocking solution (26 µM R02 blocking oligo 5' ATCCACGTGCTTGAGAGGCCAGAGCATTCG in 1x T4 ligase buffer) was added to each well. The plate was re-sealed and mixed again for 30 sec at 1200rpm before incubating for an additional 30 min at 37˚C. The nuclei from each well were then pooled into 1.5 ml tubes and pelleted at 800g for 5 min at 4˚C. The second round of ligation was then carried out using R03 biotin tagged barcode plate [82] similarly to the first round, except that after the 30 min ligation reaction, blocking solution (26 µM quencher oligo 5' GTGGCCGATGTTTCGGTGCGAACTCAGACC in 0.25 M EDTA) was added to quench the reaction. After 2 rounds of ligation-based barcoding, the pooled nuclei were resuspended in 1x PBS with 0.5U/µl ea Superase RNAse Inhibitor and RNAaseOUT, counted and aliquoted into 10 µl sublibraries containing 3,000 nuclei.

For each sublibrary nuclei were lysed and cross-links reversed by adding 10 µl of lysis buffer (10 mM Tris-pH 7.5, 1M $NaCl_2$, 0.5 mM EDTA, 2% SDS) and incubating at 65˚C for 5 hrs. Biotin barcoded cDNA molecules were purified by adding 80 µl of dynabead solution (10ul MyOne Streptavidin C1 dynabeads (Invitrogen) in 10 mM Tris-pH 7.5, 1M $NaCl_2$, 0.5 mM EDTA) and incubating for 1 hr at 24˚C at 1000rpm. Bound cDNAs were washed twice with 200 µl BW buffer (10 mM Tris-pH 7.5, 1M $NaCl_2$, 0.5 mM EDTA) followed by 2 additional washes with 200 µl Tris-T solution (10 mM Tris-HCL pH7.5, 0.1% Tween 20). For incorporating a 5' adaptor onto the streptavidin bound cDNA a 50 µl template switching mix (1x Maxima H buffer, 4% Ficoll, 1mM dNTPs, 2.5 µM TS oligo (AAGCAGTGGTATCAACGCAGAGTGAATrGrG+G where rG is RNA base), 500U Maxima H minus RT, 15U Superase RNAse Inhibitor and 20U RNAaseOUT) was added to each sublibrary. The bead solution was then incubated at 25˚C for 30 minutes and

then at 42°C for 90 minutes with mixing at 1000rpm. After the incubation beads were washed twice with 200 µl Tris-T solution.

Washed beads were resuspended in 50 µl of PCR reaction mix (1x Q5 reaction buffer, 0.1 mM dNTPs, 0.5 uM ea Rev (AAGCAGTGGTATCAACGCAGAGT) and For (CAGACGTGTGCTCTTCCGATCT) primer, 1U Q5 High-Fidelity DNA polymerase (NEB)) and amplified with the following program: Step 1: 95°C × 1 min; Step 2 (6 cycles): 98°C × 30 s, 70°C × 15 s, 72°C × 3 min; Step 3: 72°C × 1 min. PCR reactions were purified using a 0.85X ratio of SPRI beads (Beckman). To determine the optimal PCR cycle to give 1/3rd of a saturated signal we repeated the PCR using 1/10th of the purified PCR reaction and EvaGreen (Biotium) in a qPCR machine. Using the optimal cycle we PCR amplified the remaining cDNA and purified the reactions using 0.85X ratio of SPRI beads. To normalize the amplified cDNA fragment lengths for sequencing we trimmed the libraries using Tn5 transposase (Diagenode) loaded with (5'-[Phos] CTGTCTCTTATACACATCT, 5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG) to a size range of 200-1000bp. Each trimmed sublibrary was then PCR amplified using a unique forward index P7Trueseq barcode oligo (5' CAAGCAGAAGACGGCATACGAGATXXXXXXGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC) and reverse oligo (AATGATACGGCGACCACCGAGATCTACACTAGATCGCTCGTCGGCAGCGTCAG). Purified libraries were than sequenced on a NovaSeq 6000 using 100bp paired end reads and a read depth of 25-30,000 reads per cell.

Single-nucleus RNA-seq data preprocessing

The Paired-Tag pipeline as previously described [83] (https://github.com/cxzhu/Paired-Tag) was used to process raw sequencing reads. Briefly this set of scripts takes raw reads and extracts cellular barcodes from read2 and then uses read1 sequence to map reads to human genome reference hg38. Finally demultiplexed and mapped data from different sublibraries are merged to generate the final cell barcode gene count matrix.

Single-nucleus RNA-seq analysis

Single nucleus transcriptomes were analysed from hTLS at days 2-7 across 2 independent differentiation batches, with additional d0 and 7 samples from independent batches, and SOX2:mNeon iPSC cells, included. Filtered feature barcode matrices of snRNA-seq data were converted into Seurat objects using the Seurat R package (version 5.3.1) [84], merged into a single object, and quality control filtered for cells containing > 500 genes and < 10% mitochondrial reads. Gene expression counts were normalised using SCTransform [85], followed by PCA. Visualisation and clustering were performed using the functions RunUMAP and FindNeighbours, followed by FindClusters using resolution as stated in the figure legends. The mesodermal cluster was subset by first clustering the broader dataset at 0.03 resolution (as in Figure S1M), and re-clustered using the top 41 principal components at a resolution of 0.5. Marker genes for each cluster were identified with Wilcoxon rank sum test using the FindMarkers function with no cutoffs if represented as volcano plot (which were instead set within the plot), or with log2FoldChange > 1 and padj cutoff set at 0.005 or 1E8 for heatmaps. Cell identities were manually annotated by expression of canonical marker genes.

Pseudotime analysis

To identify distinct mesodermal lineages arising during hTLS development, the mesoderm subset was clustered with resolution 0.5, and Slingshot [35] trajectories were inferred using cluster 0 as the root and clusters 2, 1, 6 and 4 as terminal states. tradeSeq [36] was used to model gene expression dynamics along each trajectory and for statistical analyses. To identify genes defining terminal cell states, the tradeSeq endpoint differential expression test (diffEndTest) [36] was used, which tests for differential expression between lineage endpoints. For each lineage, genes were ranked using a combined endpoint significance score defined as the sum of -log10-transformed p-values from the three pairwise endpoint contrasts involving that lineage. The top 20 ranked genes for each lineage were selected for visualisation. For heatmap visualisation (Figure 1K), expression dynamics along pseudotime were obtained from the fitted tradeSeq GAM as predicted expression values and transformed using log1p. Gene expression trajectories were standardised by computing z-scores relative to the

gene's mean fitted expression across all pseudotime lineages. Heatmap values thus represent global standardised expression, with positive values (red) indicating expression above the gene's global mean and negative values (blue) indicating expression below the mean. To reduce the visual influence of genes with low absolute expression levels, standardised values were scaled by an expression-dependent weight derived from the fitted log1p expression. Specifically, z-scores were multiplied by an expression-dependent logistic weighting function derived from the fitted log1p expression, thereby downweighting pseudotime regions with low fitted expression and therefore less likely to reflect biologically meaningful expression. Weighted z-scores were clipped to ±2.5 to limit the influence of extreme values.

Motif enrichment

Motif enrichment analysis was performed using HOMER (v4.11) [86] on ATAC-seq peaks from early somit populations generated using the same differentiation protocol as described by Loh et al. [11]. Public ATAC-seq peaks corresponding to early somite cells (GEO accession GSM2257295) were used as the accessible chromatin reference set. Peaks located within ±50 kb of transcription start sites of genes significantly downregulated following FOXC1/2 knockdown were identified using BEDTools. These regions were tested for transcription factor motif enrichment relative to a background set consisting of ATAC-seq peaks located within ±50 kb of transcription start sites of genes expressed but not significantly altered following FOXC1/2 knockdown. Motif enrichment was calculated using HOMER (findMotifsGenome.pl) with sequences extracted from the hg19 human genome assembly. Significantly enriched motifs were ranked by hypergeometric p-value and FDR. Enriched motifs were interpreted as candidate transcriptiocnal regulators acting at FOXC-dependent regulatory elements during early somite specification.

Mapping scRNA-seq to reference human embryogenesis datasets

We projected our data (query dataset) on two reference datasets: Zhao et al. for the earlier time points (day 0-4) [24], and Zeng et al. for the later ones (days 5-7) [23]. For the Zhao et al. reference, we used the projection pipeline provided by the authors themselves for projecting any user's query dataset onto their reference (R function: projecting_query_dataset.R). We modified only one parameter of the standard pipeline: para.list$cor.cutoff <- 0.2 instead of 0.5. We projected our datasets one sample at the time. The output of this pipeline matches each cell of the target to its closest location in the reference's umap, which allows to represent the reference landscape as published by the authors with the query dataset projected on top. For the Zeng et al dataset, we built our own projection pipeline, which consists of two parts: integration and projection. For the integration, we used the scVI method from the scVI-tools (10.1038/s41592-018-0229-2) to correct the raw counts of the reference and target datasets. Correcting the counts was necessary due to the different chemistries of the reference and the query datasets. Once corrected the counts, we computed 50 pcs, that were then used for projecting one sample of the query datasets at the time, following these steps: we computed the distance matrix among the pcs of the reference and those of the query dataset; for each cell in a query sample, we gave a score 1 to its 15 nearest neighbours (NN) of the reference based on the distance matrix, and 0 otherwise; we computed the adjacency (or similarity) score of each cell of the reference to each cell of the query dataset as the average of the scores 1 (cell being a NN) or 0 (cell not being a NN). As a result, we can visualize the originally published umap of the reference with our own data projected on top in terms of similarity scores.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analyses of bulk RNA-seq data were performed with DESeq2. Analyses of snRNA-seq data were performed with Seurat. Marker genes were analysed with Wilcoxon rank sum test. Thresholds for volcano plots were set as |log2FoldChange| > 0.58 (±1.5-fold) and padj < 0.05, and for heatmaps were set as log2FoldChange >1 (>2-fold) and padj < 1E8 (Fig S1N) and padj < 0.005 (Fig S1P). Statistical analyses of confocal microscopy quantifications and RT-qPCR data were performed within Prism.

For image quantification, data were analysed with Mann-Whitney test and visualised in plots with dots representing individual hTLS and bars represent median ± IQR. For RT-qPCR analyses, individual data points represent means of up to 4 technical replicates from a single independent differentiation batch, which were used as ‘n’s for statistical analyses. Biological replicate data were assumed have a normal distribution, while to account for batch effects, data were analysed using two-way ANOVA with batch as row factor and treatment as column factor. RT-qPCR data are visualised in plots with bars representing mean ± SEM. Asterisks in plots represent p values as follows: * ($p < 0.05$), ** ($p < 0.01$), *** ($p < 0.001$), **** ($p < 0.0001$). All the details of quantifications and statistical analyses are fully described in the main text, figure legends, and method details section.

References

1. Ghimire, S., Mantziou, V., Moris, N., and Martinez Arias, A. (2021). Human gastrulation: The embryo and its models. Dev Biol *474*, 100–108. 10.1016/j.ydbio.2021.01.006.
2. Sheng, G., Martinez Arias, A., and Sutherland, A. (2021). The primitive streak and cellular principles of building an amniote body through gastrulation. Science *374*, abg1727. 10.1126/science.abg1727.
3. Ferretti, E., and Hadjantonakis, A.K. (2019). Mesoderm specification and diversification: from single cells to emergent tissues. Curr Opin Cell Biol *61*, 110–116. 10.1016/j.ceb.2019.07.012.
4. Miao, Y., and Pourquie, O. (2024). Cellular and molecular control of vertebrate somitogenesis. Nat Rev Mol Cell Biol *25*, 517–533. 10.1038/s41580-024-00709-z.
5. Costantini, F., and Kopan, R. (2010). Patterning a complex organ: branching morphogenesis and nephron segmentation in kidney development. Dev Cell *18*, 698–712. 10.1016/j.devcel.2010.04.008.
6. McMahon, A.P. (2016). Development of the Mammalian Kidney. Curr Top Dev Biol *117*, 31–64. 10.1016/bs.ctdb.2015.10.010.
7. Garcia-Martinez, V., and Schoenwolf, G.C. (1992). Positional control of mesoderm movement and fate during avian gastrulation and neurulation. Dev Dyn *193*, 249–256. 10.1002/aja.1001930305.
8. Kinder, S.J., Tsang, T.E., Quinlan, G.A., Hadjantonakis, A.K., Nagy, A., and Tam, P.P. (1999). The orderly allocation of mesodermal cells to the extraembryonic structures and the anteroposterior axis during gastrulation of the mouse embryo. Development *126*, 4691–4701. 10.1242/dev.126.21.4691.
9. Parameswaran, M., and Tam, P.P. (1995). Regionalisation of cell fate and morphogenetic movement of the mesoderm during mouse gastrulation. Dev Genet *17*, 16–28. 10.1002/dvg.1020170104.
10. Tam, P.P., and Beddington, R.S. (1987). The formation of mesodermal tissues in the mouse embryo during gastrulation and early organogenesis. Development *99*, 109–126. 10.1242/dev.99.1.109.

11. Tam, P.P., Parameswaran, M., Kinder, S.J., and Weinberger, R.P. (1997). The allocation of epiblast cells to the embryonic heart and other mesodermal lineages: the role of ingression and tissue movement during gastrulation. Development *124*, 1631–1642. 10.1242/dev.124.9.1631.
12. Wilson, V., and Beddington, R.S. (1996). Cell fate and morphogenetic movement in the late mouse primitive streak. Mech Dev *55*, 79–89. 10.1016/0925-4773(95)00493-9.
13. Duester, G. (2008). Retinoic acid synthesis and signaling during early organogenesis. Cell *134*, 921–931. 10.1016/j.cell.2008.09.002.
14. Yamaguchi, T.P. (2001). Heads or tails: Wnts and anterior-posterior patterning. Curr Biol *11*, R713–724. 10.1016/s0960-9822(01)00417-1.
15. Kim, S., Koppitch, K., Parvez, R.K., Guo, J., Achieng, M., Schnell, J., Lindstrom, N.O., and McMahon, A.P. (2024). Comparative single-cell analyses identify shared and divergent features of human and mouse kidney development. Dev Cell *59*, 2912–2930 e2917. 10.1016/j.devcel.2024.07.013.
16. Lindstrom, N.O., Tran, T., Guo, J., Rutledge, E., Parvez, R.K., Thornton, M.E., Grubbs, B., McMahon, J.A., and McMahon, A.P. (2018). Conserved and Divergent Molecular and Anatomic Features of Human and Mouse Nephron Patterning. J Am Soc Nephrol *29*, 825–840. 10.1681/ASN.2017091036.
17. Lindstrom, N.O., McMahon, J.A., Guo, J., Tran, T., Guo, Q., Rutledge, E., Parvez, R.K., Saribekyan, G., Schuler, R.E., Liao, C., et al. (2018). Conserved and Divergent Features of Human and Mouse Kidney Organogenesis. J Am Soc Nephrol *29*, 785–805. 10.1681/ASN.2017080887.
18. Lindstrom, N.O., Guo, J., Kim, A.D., Tran, T., Guo, Q., De Sena Brandine, G., Ransick, A., Parvez, R.K., Thornton, M.E., Baskin, L., et al. (2018). Conserved and Divergent Features of Mesenchymal Progenitor Cell Types within the Cortical Nephrogenic Niche of the Human and Mouse Kidney. J Am Soc Nephrol *29*, 806–824. 10.1681/ASN.2017080890.
19. Matsuda, M., Hayashi, H., Garcia-Ojalvo, J., Yoshioka-Kobayashi, K., Kageyama, R., Yamanaka, Y., Ikeya, M., Toguchida, J., Alev, C., and Ebisuya, M. (2020).

Species-specific segmentation clock periods are due to differential biochemical reaction speeds. Science *369*, 1450–1455. 10.1126/science.aba7668.

20. Xi, H., Fujiwara, W., Gonzalez, K., Jan, M., Liebscher, S., Van Handel, B., Schenke-Layland, K., and Pyle, A.D. (2017). In Vivo Human Somitogenesis Guides Somite Development from hPSCs. Cell Rep *18*, 1573–1585. 10.1016/j.celrep.2017.01.040.

21. Rivron, N.C., Martinez Arias, A., Pera, M.F., Moris, N., and M'Hamdi H, I. (2023). An ethical framework for human embryology with embryo models. Cell *186*, 3548–3557. 10.1016/j.cell.2023.07.028.

22. Tyser, R.C.V., Mahammadov, E., Nakanoh, S., Vallier, L., Scialdone, A., and Srinivas, S. (2021). Single-cell transcriptomic characterization of a gastrulating human embryo. Nature *600*, 285–289. 10.1038/s41586-021-04158-y.

23. Zeng, B., Liu, Z., Lu, Y., Zhong, S., Qin, S., Huang, L., Zeng, Y., Li, Z., Dong, H., Shi, Y., et al. (2023). The single-cell and spatial transcriptional landscape of human gastrulation and early brain development. Cell Stem Cell *30*, 851–866 e857. 10.1016/j.stem.2023.04.016.

24. Zhao, C., Plaza Reyes, A., Schell, J.P., Weltner, J., Ortega, N.M., Zheng, Y., Bjorklund, A.K., Baque-Vidal, L., Sokka, J., Trokovic, R., et al. (2025). A comprehensive human embryo reference tool using single-cell RNA-sequencing data. Nat Methods *22*, 193–206. 10.1038/s41592-024-02493-2.

25. Gribaudo, S., Robert, R., van Sambeek, B., Mirdass, C., Lyubimova, A., Bouhali, K., Ferent, J., Morin, X., van Oudenaarden, A., and Nedelec, S. (2024). Self-organizing models of human trunk organogenesis recapitulate spinal cord and spine co-morphogenesis. Nat Biotechnol *42*, 1243–1253. 10.1038/s41587-023-01956-9.

26. Hamazaki, N., Yang, W., Kubo, C.A., Qiu, C., Martin, B.K., Garge, R.K., Regalado, S.G., Nichols, E.K., Pendyala, S., Bradley, N., et al. (2024). Retinoic acid induces human gastruloids with posterior embryo-like structures. Nat Cell Biol *26*, 1790–1803. 10.1038/s41556-024-01487-8.

27. Shahbazi, M.N., and Pasque, V. (2024). Early human development and stem cell-based human embryo models. Cell Stem Cell *31*, 1398–1418. 10.1016/j.stem.2024.09.002.

28. Makwana, K., Tilley, L., Chakravarty, P., Thompson, J., Baillie-Benson, P., Rodriguez-Polo, I., and Moris, N. (2025). Modelling co-development between the somites and neural tube in human trunk-like structures. Nat Cell Biol *27*, 2049–2062. 10.1038/s41556-025-01813-8.

29. Wu, H., and Wang, H. (2026). Progress in stem cell-based embryo models and their applications in developmental biology and biomedicine. Nat Rev Mol Cell Biol *27*, 178–193. 10.1038/s41580-025-00942-0.

30. Yamanaka, Y., Hamidi, S., Yoshioka-Kobayashi, K., Munira, S., Sunadome, K., Zhang, Y., Kurokawa, Y., Ericsson, R., Mieda, A., Thompson, J.L., et al. (2023). Reconstituting human somitogenesis in vitro. Nature *614*, 509–520. 10.1038/s41586-022-05649-2.

31. Miao, Y., Djeffal, Y., De Simone, A., Zhu, K., Lee, J.G., Lu, Z., Silberfeld, A., Rao, J., Tarazona, O.A., Mongera, A., et al. (2023). Reconstruction and deconstruction of human somitogenesis in vitro. Nature *614*, 500–508. 10.1038/s41586-022-05655-4.

32. Sanaki-Matsumiya, M., Matsuda, M., Gritti, N., Nakaki, F., Sharpe, J., Trivedi, V., and Ebisuya, M. (2022). Periodic formation of epithelial somites from human pluripotent stem cells. Nat Commun *13*, 2325. 10.1038/s41467-022-29967-1.

33. Moris, N., Anlas, K., van den Brink, S.C., Alemany, A., Schroder, J., Ghimire, S., Balayo, T., van Oudenaarden, A., and Martinez Arias, A. (2020). An in vitro model of early anteroposterior organization during human development. Nature *582*, 410–415. 10.1038/s41586-020-2383-9.

34. Zhang, X., Huang, C.T., Chen, J., Pankratz, M.T., Xi, J., Li, J., Yang, Y., Lavaute, T.M., Li, X.J., Ayala, M., et al. (2010). Pax6 is a human neuroectoderm cell fate determinant. Cell Stem Cell *7*, 90–100. 10.1016/j.stem.2010.04.017.

35. Street, K., Risso, D., Fletcher, R.B., Das, D., Ngai, J., Yosef, N., Purdom, E., and Dudoit, S. (2018). Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. BMC Genomics *19*, 477. 10.1186/s12864-018-4772-0.

36. Van den Berge, K., Roux de Bezieux, H., Street, K., Saelens, W., Cannoodt, R., Saeys, Y., Dudoit, S., and Clement, L. (2020). Trajectory-based differential expression analysis for single-cell sequencing data. Nat Commun *11*, 1201. 10.1038/s41467-020-14766-3.

37. Chapman, D.L., Agulnik, I., Hancock, S., Silver, L.M., and Papaioannou, V.E. (1996). Tbx6, a mouse T-Box gene implicated in paraxial mesoderm formation at gastrulation. Dev Biol *180*, 534–542. 10.1006/dbio.1996.0326.

38. Sadahiro, T., Isomi, M., Muraoka, N., Kojima, H., Haginiwa, S., Kurotsu, S., Tamura, F., Tani, H., Tohyama, S., Fujita, J., et al. (2018). Tbx6 Induces Nascent Mesoderm from Pluripotent Stem Cells and Temporally Controls Cardiac versus Somite Lineage Diversification. Cell Stem Cell *23*, 382–395 e385. 10.1016/j.stem.2018.07.001.

39. Tang, W.W.C., Castillo-Venzor, A., Gruhn, W.H., Kobayashi, T., Penfold, C.A., Morgan, M.D., Sun, D., Irie, N., and Surani, M.A. (2022). Sequential enhancer state remodelling defines human germline competence and specification. Nat Cell Biol *24*, 448–460. 10.1038/s41556-022-00878-z.

40. Nabet, B., Roberts, J.M., Buckley, D.L., Paulk, J., Dastjerdi, S., Yang, A., Leggett, A.L., Erb, M.A., Lawlor, M.A., Souza, A., et al. (2018). The dTAG system for immediate and target-specific protein degradation. Nat Chem Biol *14*, 431–441. 10.1038/s41589-018-0021-8.

41. Wilm, B., James, R.G., Schultheiss, T.M., and Hogan, B.L. (2004). The forkhead genes, Foxc1 and Foxc2, regulate paraxial versus intermediate mesoderm cell fate. Dev Biol *271*, 176–189. 10.1016/j.ydbio.2004.03.034.

42. Kume, T., Deng, K., and Hogan, B.L. (2000). Murine forkhead/winged helix genes Foxc1 (Mf1) and Foxc2 (Mfh1) are required for the early organogenesis of the kidney and urinary tract. Development *127*, 1387–1395. 10.1242/dev.127.7.1387.

43. Kume, T., Jiang, H., Topczewska, J.M., and Hogan, B.L. (2001). The murine winged helix transcription factors, Foxc1 and Foxc2, are both required for cardiovascular development and somitogenesis. Genes Dev *15*, 2470–2482. 10.1101/gad.907301.

44. Loh, K.M., Chen, A., Koh, P.W., Deng, T.Z., Sinha, R., Tsai, J.M., Barkal, A.A., Shen, K.Y., Jain, R., Morganti, R.M., et al. (2016). Mapping the Pairwise Choices Leading from Pluripotency to Human Bone, Heart, and Other Mesoderm Cell Types. Cell *166*, 451–467. 10.1016/j.cell.2016.06.011.

45. Asp, M., Giacomello, S., Larsson, L., Wu, C., Furth, D., Qian, X., Wardell, E., Custodio, J., Reimegard, J., Salmen, F., et al. (2019). A Spatiotemporal Organ-Wide Gene Expression and Cell Atlas of the Developing Human Heart. Cell *179*, 1647–1660 e1619. 10.1016/j.cell.2019.11.025.

46. Stewart, B.J., Ferdinand, J.R., Young, M.D., Mitchell, T.J., Loudon, K.W., Riding, A.M., Richoz, N., Frazer, G.L., Staniforth, J.U.L., Vieira Braga, F.A., et al. (2019). Spatiotemporal immune zonation of the human kidney. Science *365*, 1461–1466. 10.1126/science.aat5031.

47. Fan, C.M., Porter, J.A., Chiang, C., Chang, D.T., Beachy, P.A., and Tessier-Lavigne, M. (1995). Long-range sclerotome induction by sonic hedgehog: direct role of the amino-terminal cleavage product and modulation by the cyclic AMP signaling pathway. Cell *81*, 457–465. 10.1016/0092-8674(95)90398-4.

48. Johnson, R.L., Laufer, E., Riddle, R.D., and Tabin, C. (1994). Ectopic expression of Sonic hedgehog alters dorsal-ventral patterning of somites. Cell *79*, 1165–1173. 10.1016/0092-8674(94)90008-6.

49. Yang, J., Mani, S.A., Donaher, J.L., Ramaswamy, S., Itzykson, R.A., Come, C., Savagner, P., Gitelman, I., Richardson, A., and Weinberg, R.A. (2004). Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis. Cell *117*, 927–939. 10.1016/j.cell.2004.06.006.

50. Thiery, J.P., Acloque, H., Huang, R.Y., and Nieto, M.A. (2009). Epithelial-mesenchymal transitions in development and disease. Cell *139*, 871–890. 10.1016/j.cell.2009.11.007.

51. Row, R.H., Pegg, A., Kinney, B.A., Farr, G.H., 3rd, Maves, L., Lowell, S., Wilson, V., and Martin, B.L. (2018). BMP and FGF signaling interact to pattern mesoderm by controlling basic helix-loop-helix transcription factor activity. Elife *7*. 10.7554/eLife.31018.

52. Chapman, D.L., and Papaioannou, V.E. (1998). Three neural tubes in mouse embryos with mutations in the T-box gene Tbx6. Nature *391*, 695–697. 10.1038/35624.
53. Takemoto, T., Uchikawa, M., Yoshida, M., Bell, D.M., Lovell-Badge, R., Papaioannou, V.E., and Kondoh, H. (2011). Tbx6-dependent Sox2 regulation determines neural or mesodermal fate in axial stem cells. Nature *470*, 394–398. 10.1038/nature09729.
54. Liu, J., Wu, N., Deciphering Disorders Involving, S., study, C.O., Yang, N., Takeda, K., Chen, W., Li, W., Du, R., Liu, S., et al. (2019). TBX6-associated congenital scoliosis (TACS) as a clinically distinguishable subtype of congenital scoliosis: further evidence supporting the compound inheritance and TBX6 gene dosage model. Genet Med *21*, 1548–1558. 10.1038/s41436-018-0377-x.
55. Yang, N., Wu, N., Zhang, L., Zhao, Y., Liu, J., Liang, X., Ren, X., Li, W., Chen, W., Dong, S., et al. (2019). TBX6 compound inheritance leads to congenital vertebral malformations in humans and mice. Hum Mol Genet *28*, 539–547. 10.1093/hmg/ddy358.
56. Wu, N., Ming, X., Xiao, J., Wu, Z., Chen, X., Shinawi, M., Shen, Y., Yu, G., Liu, J., Xie, H., et al. (2015). TBX6 null variants and a common hypomorphic allele in congenital scoliosis. N Engl J Med *372*, 341–350. 10.1056/NEJMoa1406829.
57. Garza Flores, A., Nordgren, I., Pettersson, M., Dias-Santagata, D., Nilsson, D., Hammarsjo, A., Lindstrand, A., Batkovskyte, D., Wiggs, J., Walton, D.S., et al. (2023). Case report: Extending the spectrum of clinical and molecular findings in FOXC1 haploinsufficiency syndrome. Front Genet *14*, 1174046. 10.3389/fgene.2023.1174046.
58. Reis, L.M., Maheshwari, M., Capasso, J., Atilla, H., Dudakova, L., Thompson, S., Zitano, L., Lay-Son, G., Lowry, R.B., Black, J., et al. (2023). Axenfeld-Rieger syndrome: more than meets the eye. J Med Genet *60*, 368–379. 10.1136/jmg-2022-108646.
59. Concepcion, D., Washkowitz, A.J., DeSantis, A., Ogea, P., Yang, J.I., Douglas, N.C., and Papaioannou, V.E. (2017). Cell lineage of timed cohorts of Tbx6-

expressing cells in wild-type and Tbx6 mutant embryos. Biol Open *6*, 1065–1073. 10.1242/bio.026203.

60. Sandbacka, M., Laivuori, H., Freitas, E., Halttunen, M., Jokimaa, V., Morin-Papunen, L., Rosenberg, C., and Aittomaki, K. (2013). TBX6, LHX1 and copy number variations in the complex genetics of Mullerian aplasia. Orphanet J Rare Dis *8*, 125. 10.1186/1750-1172-8-125.

61. Dong, S., Wang, C., Li, X., Shen, Q., Fu, X., Wu, M., Song, C., Yang, N., Wu, Y., Wang, H., et al. (2019). Noncoding rare variants of TBX6 in congenital anomalies of the kidney and urinary tract. Mol Genet Genomics *294*, 493–500. 10.1007/s00438-018-1522-6.

62. Ma, C., Chen, N., Jolly, A., Zhao, S., Coban-Akdemir, Z., Tian, W., Kang, J., Ye, Y., Wang, Y., Koch, A., et al. (2022). Functional characteristics of a broad spectrum of TBX6 variants in Mayer-Rokitansky-Kuster-Hauser syndrome. Genet Med *24*, 2262–2273. 10.1016/j.gim.2022.08.012.

63. Li, G., Strong, A., Wang, H., Kim, J.S., Watson, D., Zhao, S., Vaccaro, C., Hartung, E., Hakonarson, H., Zhang, T.J., et al. (2022). TBX6 as a cause of a combined skeletal-kidney dysplasia syndrome. Am J Med Genet A *188*, 3469–3481. 10.1002/ajmg.a.62972.

64. Chapman, D.L. (2019). Impaired intermediate formation in mouse embryos expressing reduced levels of Tbx6. Genesis *57*, e23270. 10.1002/dvg.23270.

65. Hayashi, S., Suzuki, H., and Takemoto, T. (2021). The nephric mesenchyme lineage of intermediate mesoderm is derived from Tbx6-expressing derivatives of neuro-mesodermal progenitors via BMP-dependent Osr1 function. Dev Biol *478*, 155–162. 10.1016/j.ydbio.2021.07.006.

66. Uenaka, T., Napole, A., Saha, A.D., Sun, D., Singavarapu, A., Calzada, E., Chen, J., Erlebach, L., McQuade, A., Ramos, D.M., et al. (2026). Prevention of transgene silencing during human pluripotent stem cell differentiation. Cell Stem Cell. 10.1016/j.stem.2026.01.007.

67. Huang, L., Huang, Z., Fan, Y., He, L., Ye, M., Shi, K., Ji, B., Huang, J., Wang, Y., and Li, Q. (2017). FOXC1 promotes proliferation and epithelial-mesenchymal

transition in cervical carcinoma through the PI3K-AKT signal pathway. Am J Transl Res *9*, 1297–1306.

68. Ou-Yang, L., Xiao, S.J., Liu, P., Yi, S.J., Zhang, X.L., Ou-Yang, S., Tan, S.K., and Lei, X. (2015). Forkhead box C1 induces epithelial–mesenchymal transition and is a potential therapeutic target in nasopharyngeal carcinoma. Mol Med Rep *12*, 8003–8009. 10.3892/mmr.2015.4427.

69. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550. 10.1186/s13059-014-0550-8.

70. Ng-Blichfeldt, J.P., Stewart, B.J., Clatworthy, M.R., Williams, J.M., and Roper, K. (2024). Identification of a core transcriptional program driving the human renal mesenchymal-to-epithelial transition. Dev Cell *59*, 595–612 e598. 10.1016/j.devcel.2024.01.011.

71. Takashima, Y., Guo, G., Loos, R., Nichols, J., Ficz, G., Krueger, F., Oxley, D., Santos, F., Clarke, J., Mansfield, W., et al. (2014). Resetting transcription factor control circuitry toward ground-state pluripotency in human. Cell *158*, 1254–1269. 10.1016/j.cell.2014.08.029.

72. Lovell-Badge, R., Anthony, E., Barker, R.A., Bubela, T., Brivanlou, A.H., Carpenter, M., Charo, R.A., Clark, A., Clayton, E., Cong, Y., et al. (2021). ISSCR Guidelines for Stem Cell Research and Clinical Translation: The 2021 update. Stem Cell Reports *16*, 1398–1408. 10.1016/j.stemcr.2021.05.012.

73. Schmid-Burgk, J.L., Honing, K., Ebert, T.S., and Hornung, V. (2016). CRISPaint allows modular base-specific gene tagging using a ligase-4-dependent mechanism. Nat Commun *7*, 12338. 10.1038/ncomms12338.

74. Effern, M., Glodde, N., Braun, M., Liebing, J., Boll, H.N., Yong, M., Bawden, E., Hinze, D., van den Boorn-Konijnenberg, D., Daoud, M., et al. (2020). Adoptive T Cell Therapy Targeting Different Gene Products Reveals Diverse and Context-Dependent Immune Evasion in Melanoma. Immunity *53*, 564–580 e569. 10.1016/j.immuni.2020.07.007.

75. Perez-Pinera, P., Kocak, D.D., Vockley, C.M., Adler, A.F., Kabadi, A.M., Polstein, L.R., Thakore, P.I., Glass, K.A., Ousterout, D.G., Leong, K.W., et al. (2013). RNA-

guided gene activation by CRISPR-Cas9-based transcription factors. Nat Methods *10*, 973–976. 10.1038/nmeth.2600.

76. Severin, J., Lizio, M., Harshbarger, J., Kawaji, H., Daub, C.O., Hayashizaki, Y., Consortium, F., Bertin, N., and Forrest, A.R. (2014). Interactive visualization and analysis of large-scale sequencing datasets using ZENBU. Nat Biotechnol *32*, 217–219. 10.1038/nbt.2840.

77. Hodgkins, A., Farne, A., Perera, S., Grego, T., Parry-Smith, D.J., Skarnes, W.C., and Iyer, V. (2015). WGE: a CRISPR database for genome engineering. Bioinformatics *31*, 3078–3080. 10.1093/bioinformatics/btv308.

78. Gilbert, L.A., Horlbeck, M.A., Adamson, B., Villalta, J.E., Chen, Y., Whitehead, E.H., Guimaraes, C., Panning, B., Ploegh, H.L., Bassik, M.C., et al. (2014). Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. Cell *159*, 647–661. 10.1016/j.cell.2014.09.029.

79. Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods *25*, 402–408. 10.1006/meth.2001.1262.

80. Stirling, D.R., Swain-Bowden, M.J., Lucas, A.M., Carpenter, A.E., Cimini, B.A., and Goodman, A. (2021). CellProfiler 4: improvements in speed, utility and usability. BMC Bioinformatics *22*, 433. 10.1186/s12859-021-04344-9.

81. Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., et al. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation (Camb) *2*, 100141. 10.1016/j.xinn.2021.100141.

82. Rosenberg, A.B., Roco, C.M., Muscat, R.A., Kuchina, A., Sample, P., Yao, Z., Graybuck, L.T., Peeler, D.J., Mukherjee, S., Chen, W., et al. (2018). Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. Science *360*, 176–182. 10.1126/science.aam8999.

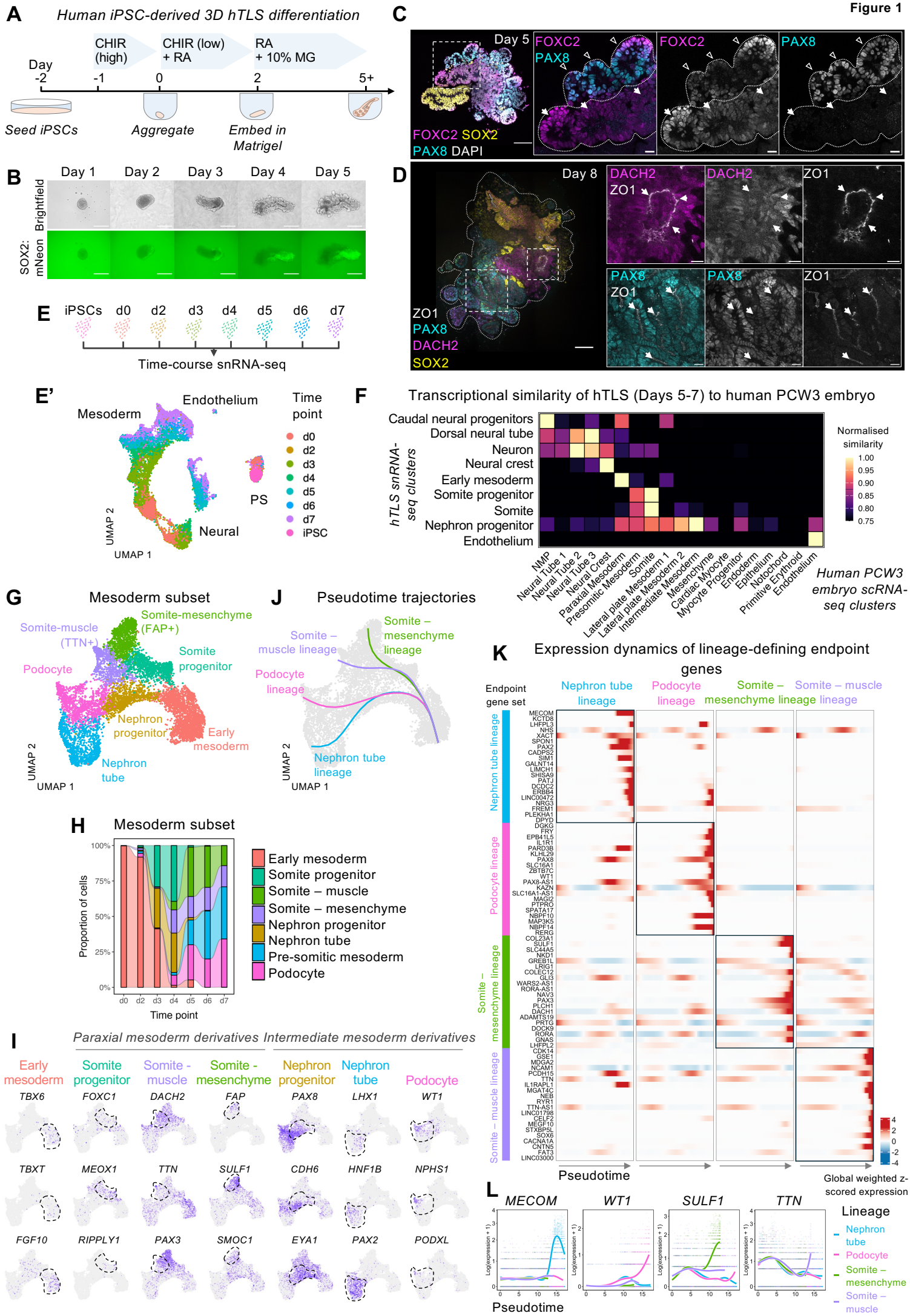
83. Zhu, C., Zhang, Y., Li, Y.E., Lucero, J., Behrens, M.M., and Ren, B. (2021). Joint profiling of histone modifications and transcriptome in single cells from mouse brain. Nat Methods *18*, 283–292. 10.1038/s41592-021-01060-3.

84. Hao, Y., Stuart, T., Kowalski, M.H., Choudhary, S., Hoffman, P., Hartman, A., Srivastava, A., Molla, G., Madad, S., Fernandez-Granda, C., and Satija, R. (2024). Dictionary learning for integrative, multimodal and scalable single-cell analysis. Nat Biotechnol *42*, 293–304. 10.1038/s41587-023-01767-y.

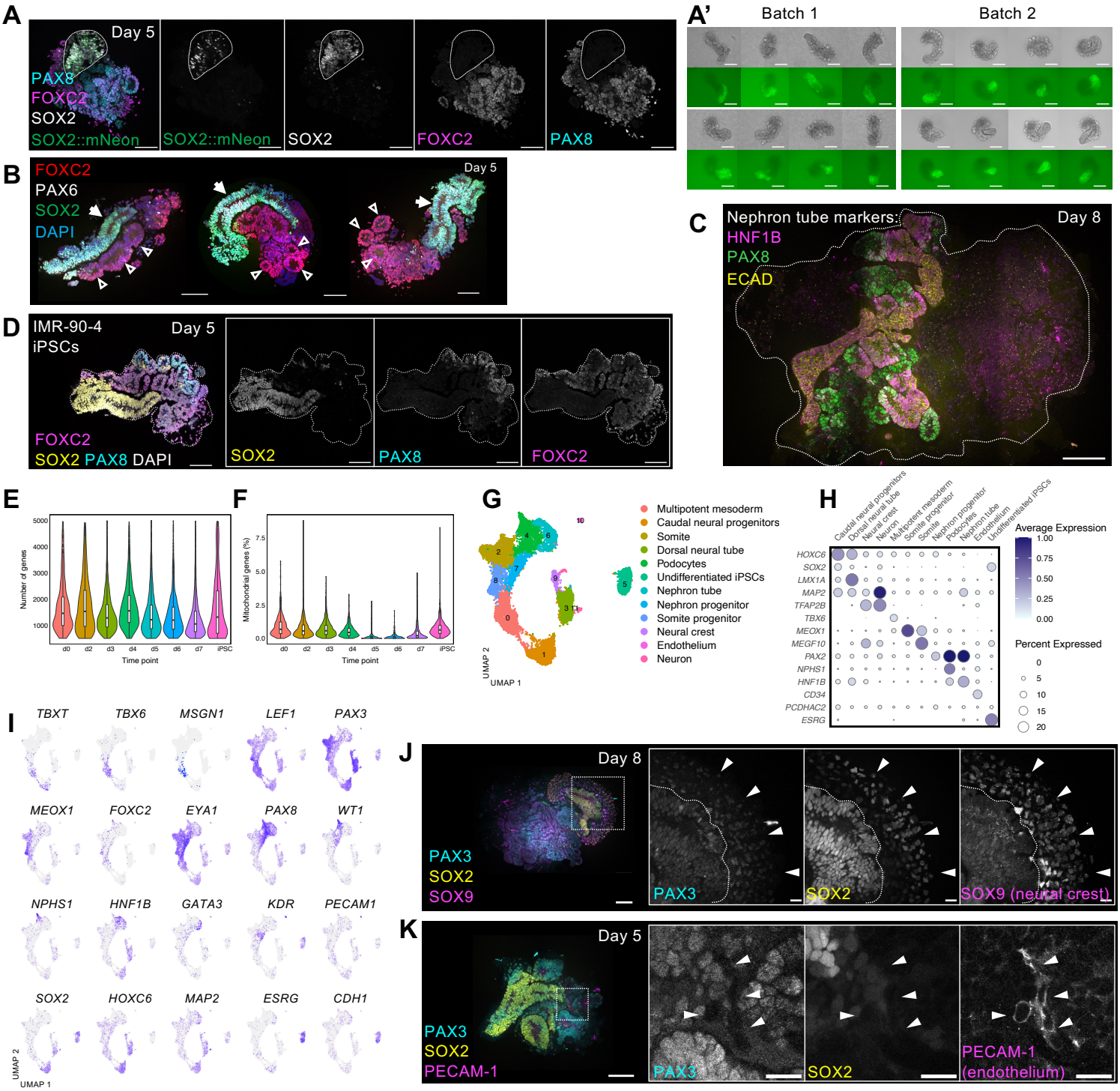
85. Hafemeister, C., and Satija, R. (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biol *20*, 296. 10.1186/s13059-019-1874-1.

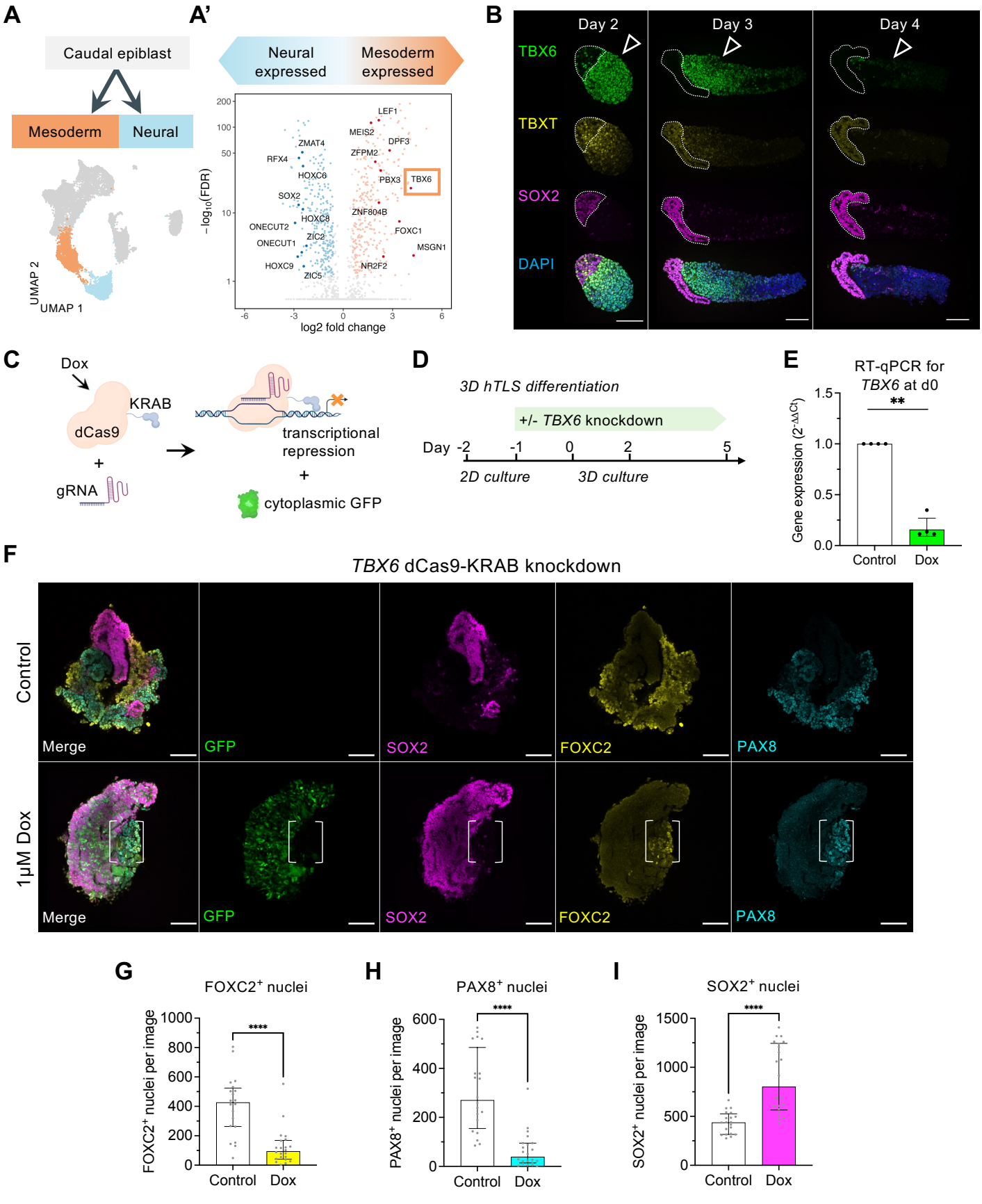
86. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell *38*, 576–589. 10.1016/j.molcel.2010.05.004.

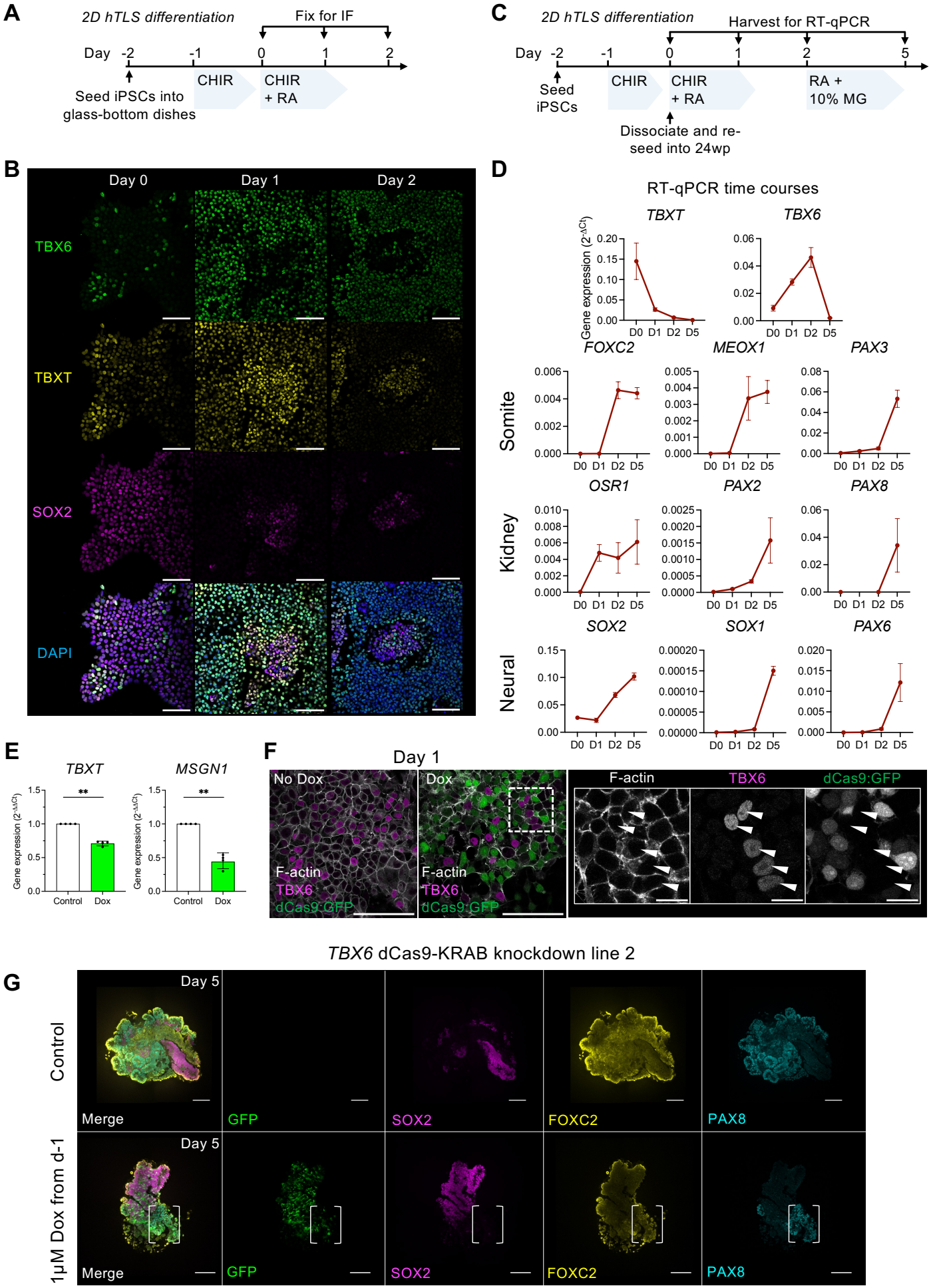
Figure 1

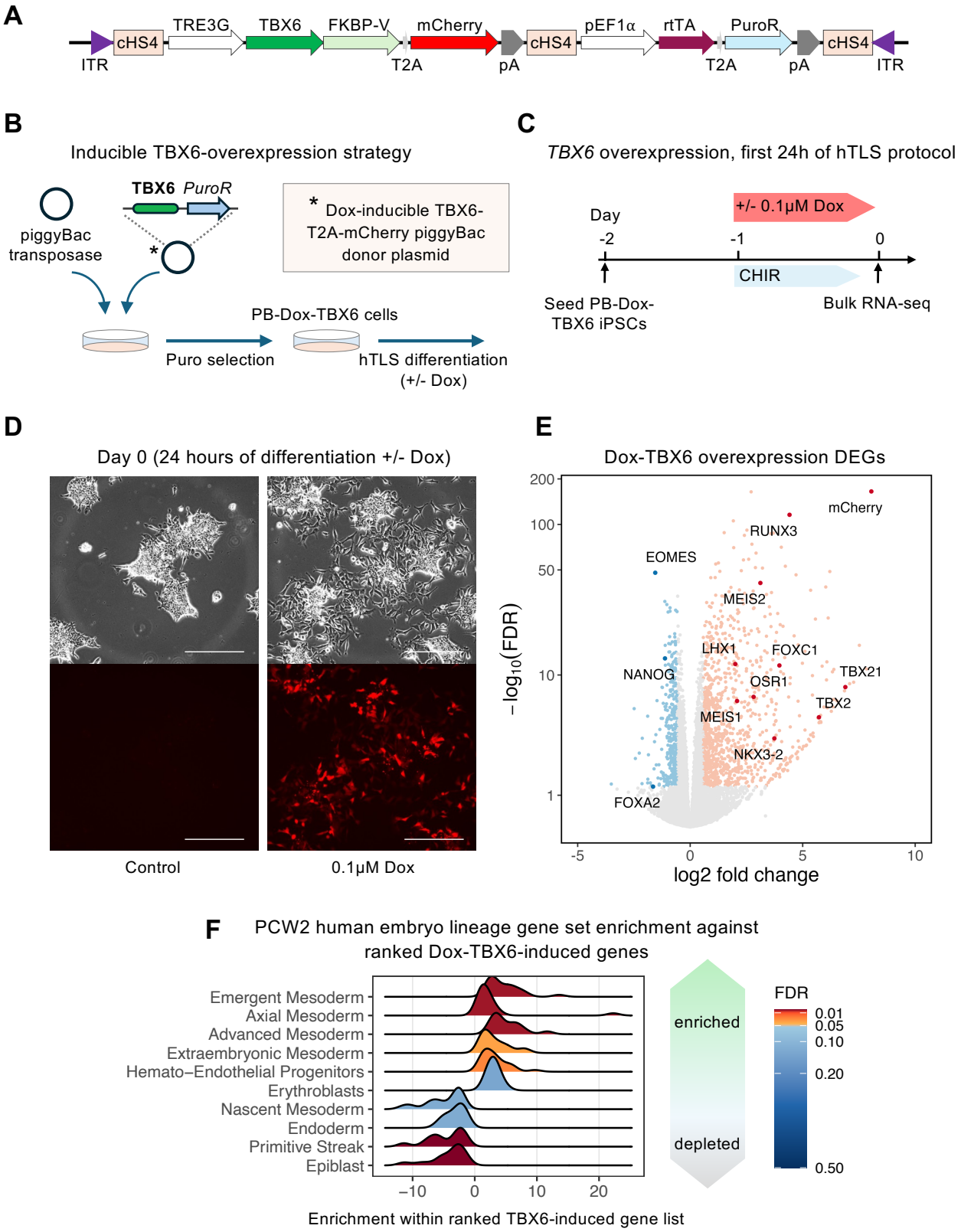

A
Human iPSC-derived 3D hTLS differentiation
CHIR (high)
CHIR (low) + RA
RA + 10% MG
Day -2
-1
0
2
5+
Seed iPSCs
Aggregate
Embed in Matrigel
B
Day 1
Day 2
Day 3
Day 4
Day 5
Brightfield
SOX2: mNeon
C
Day 5
FOXC2 SOX2
PAX8 DAPI
FOXC2
PAX8
D
Day 8
ZO1
PAX8
DACH2
SOX2
DACH2
ZO1
PAX8
E
iPSCs
d0
d2
d3
d4
d5
d6
d7
Time-course snRNA-seq
E'
Mesoderm
Endothelium
PS
Neural
UMAP 1
UMAP 2
Time point
d0
d2
d3
d4
d5
d6
d7
iPSC
F
Transcriptional similarity of hTLS (Days 5-7) to human PCW3 embryo
hTLS snRNA-seq clusters
Caudal neural progenitors
Dorsal neural tube
Neuron
Neural crest
Early mesoderm
Somite progenitor
Somite
Nephron progenitor
Endothelium
NMP
Neural Tube 1
Neural Tube 2
Neural Tube 3
Neural Crest
Paraxial Mesoderm
Presomitic Mesoderm
Somite
Lateral plate Mesoderm 1
Lateral plate Mesoderm 2
Intermediate Mesoderm
Mesenchyme
Cardiac Myocyte
Myocyte Progenitor
Endoderm
Epithelium
Notochord
Primitive Erythroid
Endothelium
Human PCW3 embryo scRNA-seq clusters
Normalised similarity
1.00
0.95
0.90
0.85
0.80
0.75
G
Mesoderm subset
Somite-muscle (TTN+)
Somite-mesenchyme (FAP+)
Somite progenitor
Podocyte
Nephron progenitor
Early mesoderm
Nephron tube
UMAP 1
UMAP 2
J
Pseudotime trajectories
Somite – muscle lineage
Somite – mesenchyme lineage
Podocyte lineage
Nephron tube lineage
UMAP 1
UMAP 2
H
Mesoderm subset
Proportion of cells
100%
75%
50%
25%
0%
d0
d2
d3
d4
d5
d6
d7
Time point
Early mesoderm
Somite progenitor
Somite – muscle
Somite – mesenchyme
Nephron progenitor
Nephron tube
Pre-somitic mesoderm
Podocyte
K
Expression dynamics of lineage-defining endpoint genes
Endpoint gene set
Nephron tube lineage
Podocyte lineage
Somite – mesenchyme lineage
Somite – muscle lineage
Pseudotime
Global weighted z-scored expression
I
Paraxial mesoderm derivatives
Intermediate mesoderm derivatives
Early mesoderm
Somite progenitor
Somite - muscle
Somite - mesenchyme
Nephron progenitor
Nephron tube
Podocyte
TBX6
FOXC1
DACH2
FAP
PAX8
LHX1
WT1
TBXT
MEOX1
TTN
SULF1
CDH6
HNF1B
NPHS1
FGF10
RIPPLY1
PAX3
SMOC1
EYA1
PAX2
PODXL
L
MECOM
WT1
SULF1
TTN
Log(expression + 1)
Pseudotime
Lineage
Nephron tube
Podocyte
Somite – mesenchyme
Somite – muscle

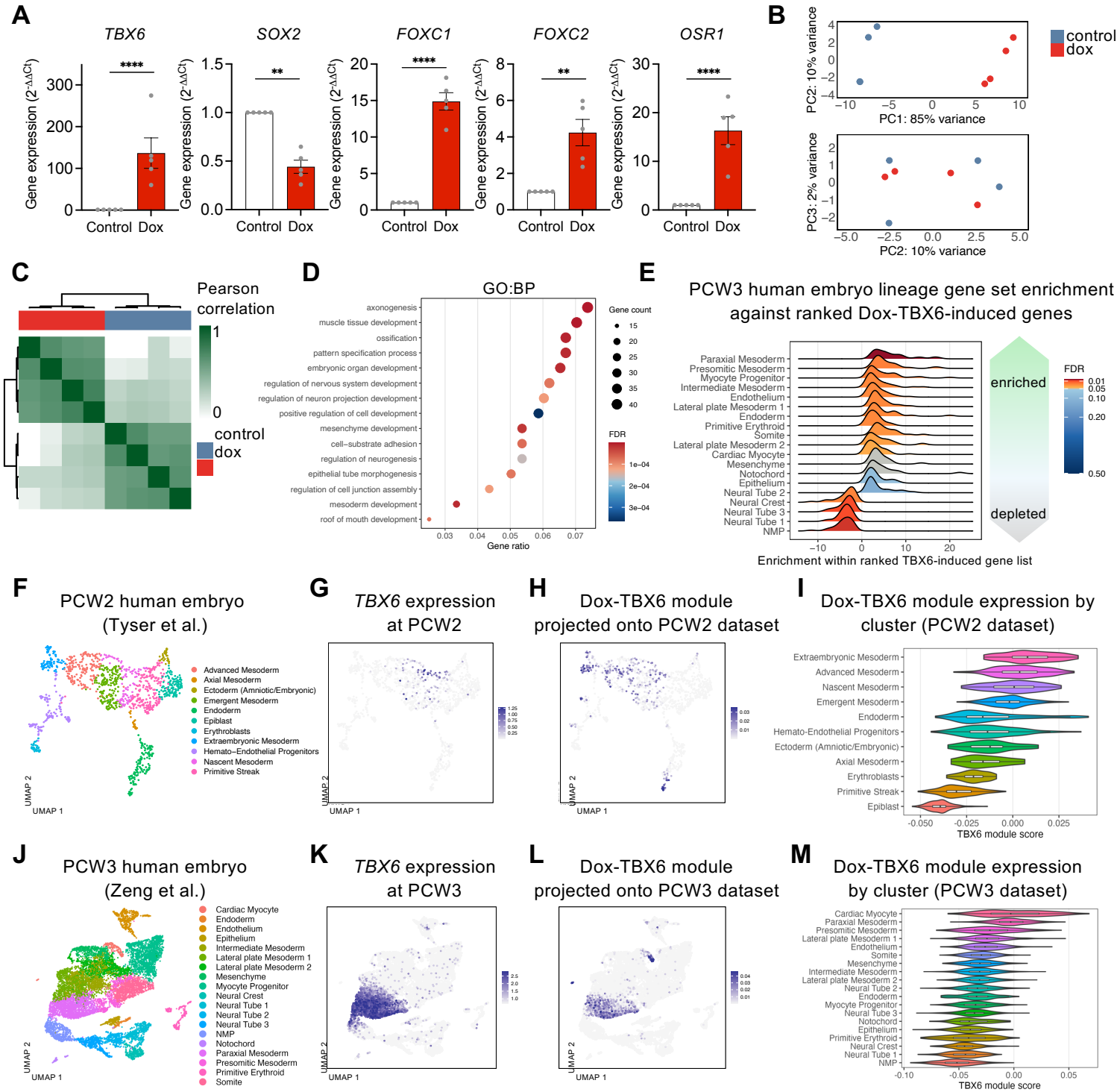
Supplementary Figure 1

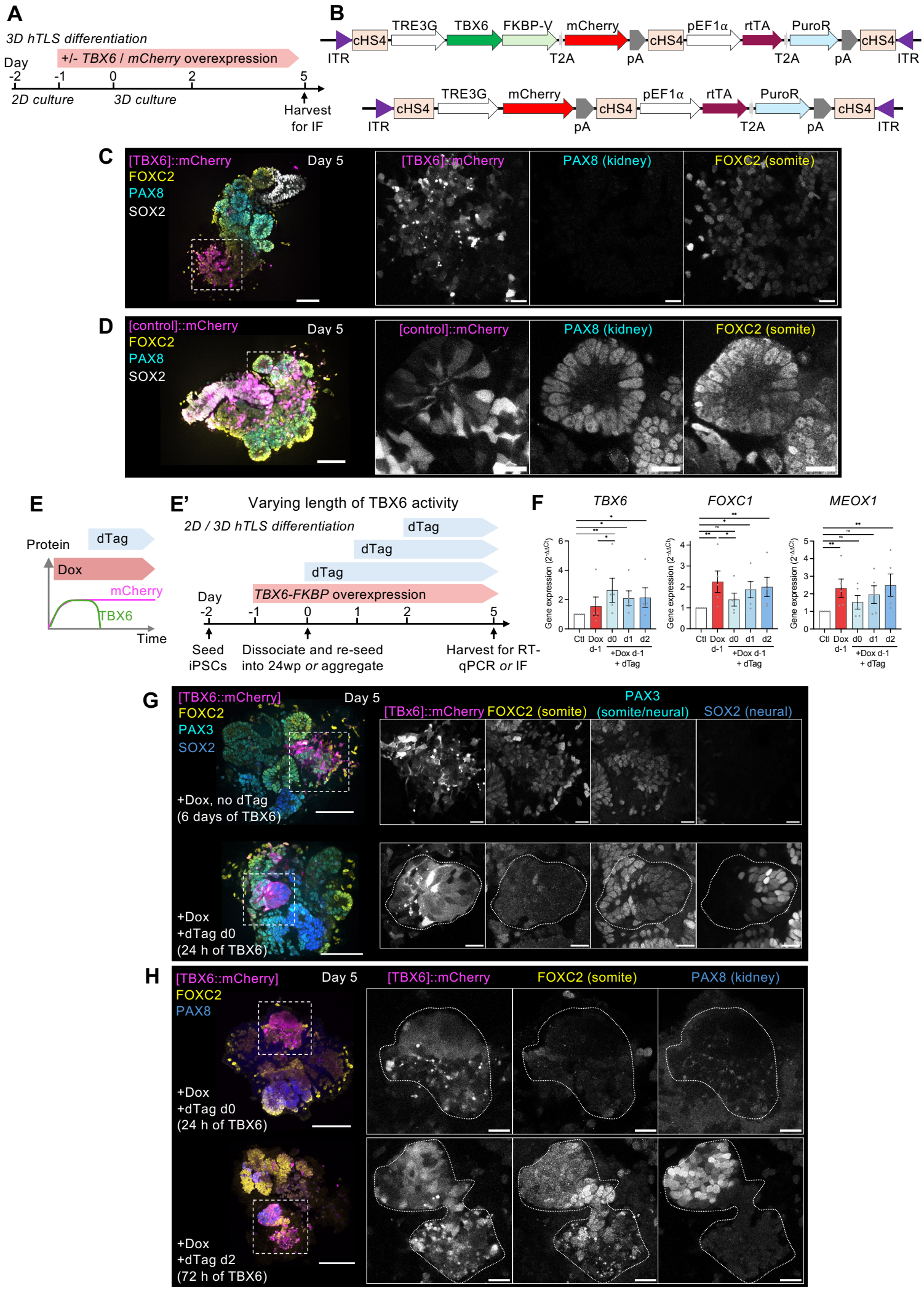


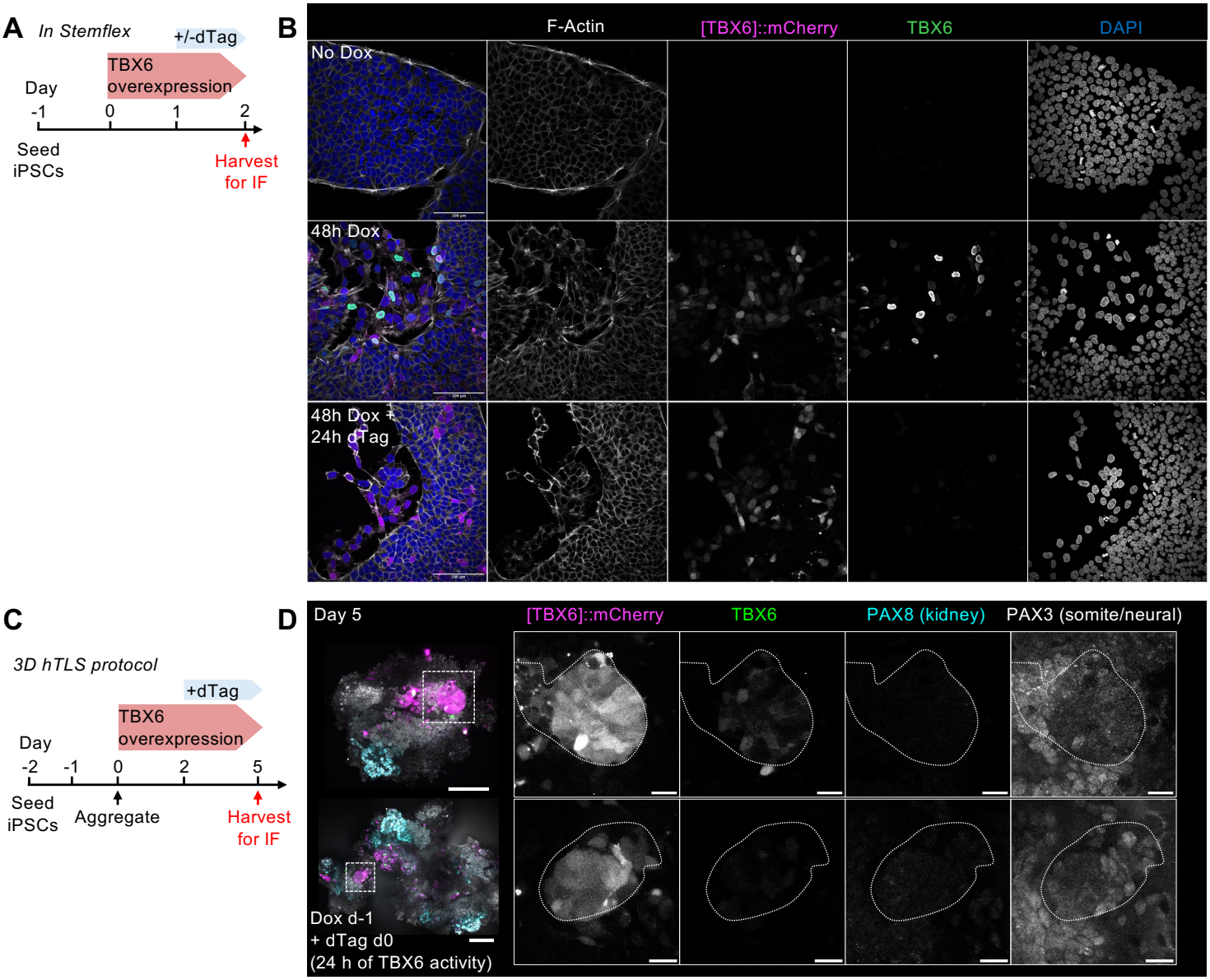

A
Caudal epiblast
Mesoderm
Neural
UMAP 2
UMAP 1
A'
Neural expressed
Mesoderm expressed
−log10(FDR)
log2 fold change
LEF1
MEIS2
DPF3
ZFPM2
ZMAT4
RFX4
HOXC6
PBX3
TBX6
SOX2
ZNF804B
HOXC8
ONECUT2
ZIC2
FOXC1
ONECUT1
MSGN1
HOXC9
NR2F2
ZIC5
B
Day 2
Day 3
Day 4
TBX6
TBXT
SOX2
DAPI
C
Dox
KRAB
dCas9
+
gRNA
transcriptional repression
+
cytoplasmic GFP
D
3D hTLS differentiation
+/- TBX6 knockdown
Day -2
-1
0
2
5
2D culture
3D culture
E
RT-qPCR for TBX6 at d0
Gene expression (2^-ΔΔCt)
Control
Dox
F
TBX6 dCas9-KRAB knockdown
Control
1μM Dox
Merge
GFP
SOX2
FOXC2
PAX8
G
FOXC2+ nuclei
FOXC2+ nuclei per image
H
PAX8+ nuclei
PAX8+ nuclei per image
I
SOX2+ nuclei
SOX2+ nuclei per image

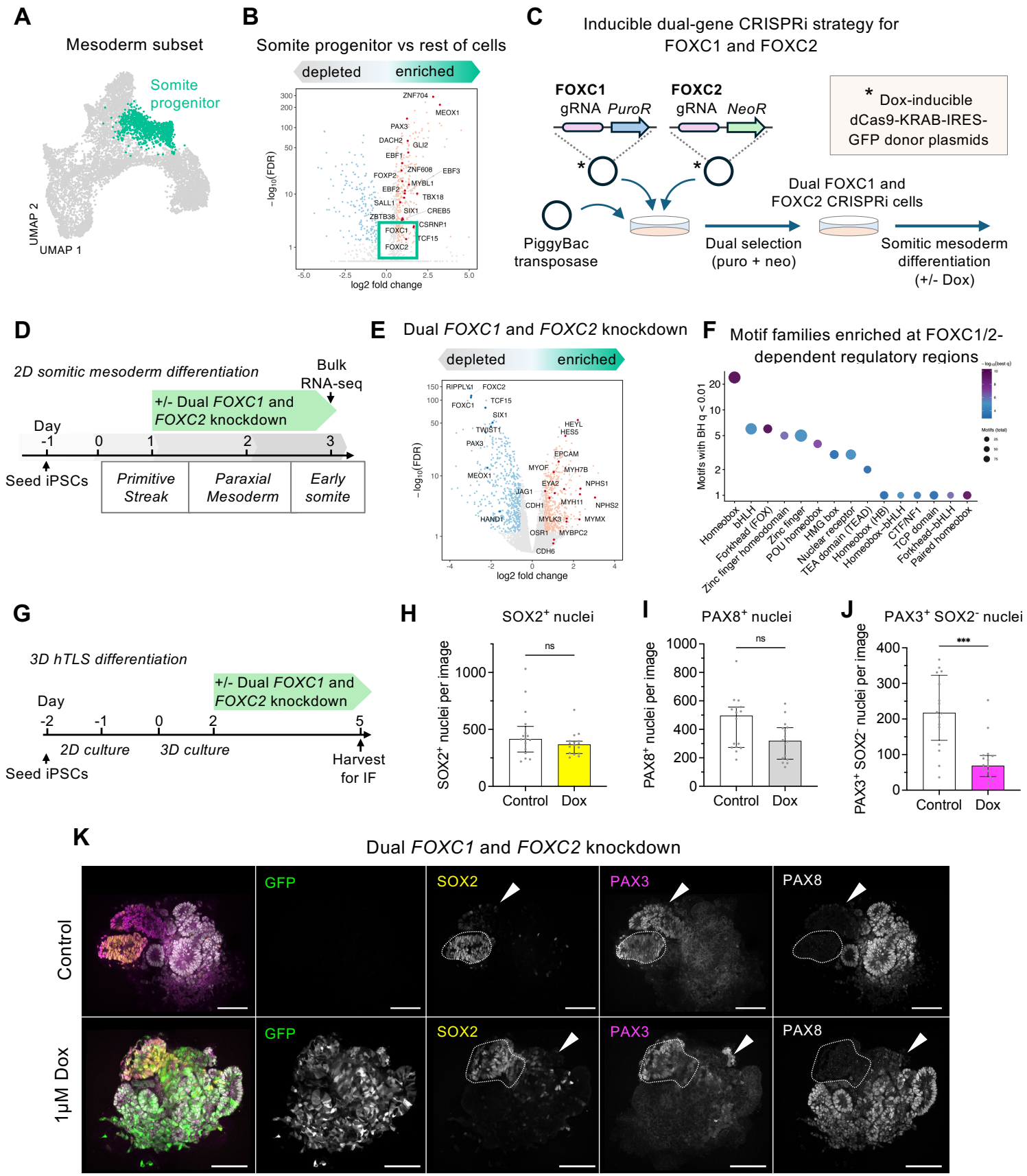

A
2D hTLS differentiation
Fix for IF
Day -2 -1 0 1 2
Seed iPSCs into glass-bottom dishes
CHIR
CHIR + RA
C
2D hTLS differentiation
Harvest for RT-qPCR
Day -2 -1 0 1 2 5
Seed iPSCs
CHIR
CHIR + RA
RA + 10% MG
Dissociate and re-seed into 24wp
B
Day 0
Day 1
Day 2
TBX6
TBXT
SOX2
DAPI
D
RT-qPCR time courses
TBXT
TBX6
FOXC2
MEOX1
PAX3
Somite
OSR1
PAX2
PAX8
Kidney
SOX2
SOX1
PAX6
Neural
Gene expression (2^-ΔΔCt)
D0 D1 D2 D5
E
TBXT
MSGN1
Control
Dox
F
Day 1
No Dox
Dox
F-actin
TBX6
dCas9:GFP
G
TBX6 dCas9-KRAB knockdown line 2
Control
1μM Dox from d-1
Day 5
Merge
GFP
SOX2
FOXC2
PAX8

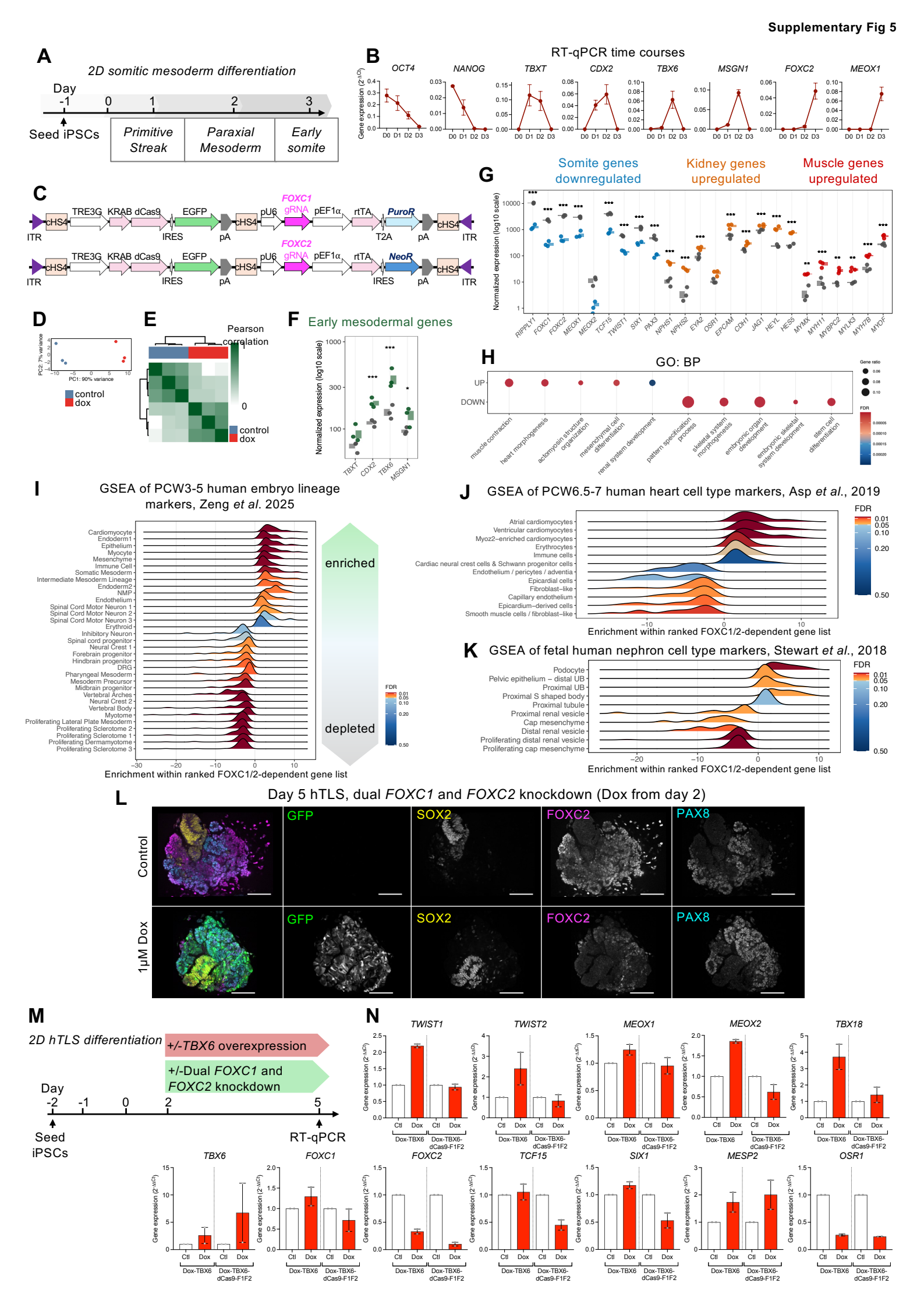

A
ITR
cHS4
TRE3G
TBX6
FKBP-V
mCherry
T2A
pA
cHS4
pEF1α
rtTA
PuroR
T2A
pA
cHS4
ITR
B
Inducible TBX6-overexpression strategy
TBX6 PuroR
piggyBac transposase
* Dox-inducible TBX6-T2A-mCherry piggyBac donor plasmid
PB-Dox-TBX6 cells
Puro selection
hTLS differentiation (+/- Dox)
C
TBX6 overexpression, first 24h of hTLS protocol
Day
+/- 0.1μM Dox
-2
-1
0
CHIR
Seed PB-Dox-TBX6 iPSCs
Bulk RNA-seq
D
Day 0 (24 hours of differentiation +/- Dox)
Control
0.1μM Dox
E
Dox-TBX6 overexpression DEGs
mCherry
RUNX3
EOMES
MEIS2
LHX1
FOXC1
NANOG
OSR1
TBX21
TBX2
MEIS1
NKX3-2
FOXA2
−log10(FDR)
log2 fold change
F
PCW2 human embryo lineage gene set enrichment against ranked Dox-TBX6-induced genes
Emergent Mesoderm
Axial Mesoderm
Advanced Mesoderm
Extraembryonic Mesoderm
Hemato–Endothelial Progenitors
Erythroblasts
Nascent Mesoderm
Endoderm
Primitive Streak
Epiblast
enriched
depleted
FDR
0.01
0.05
0.10
0.20
0.50
Enrichment within ranked TBX6-induced gene list

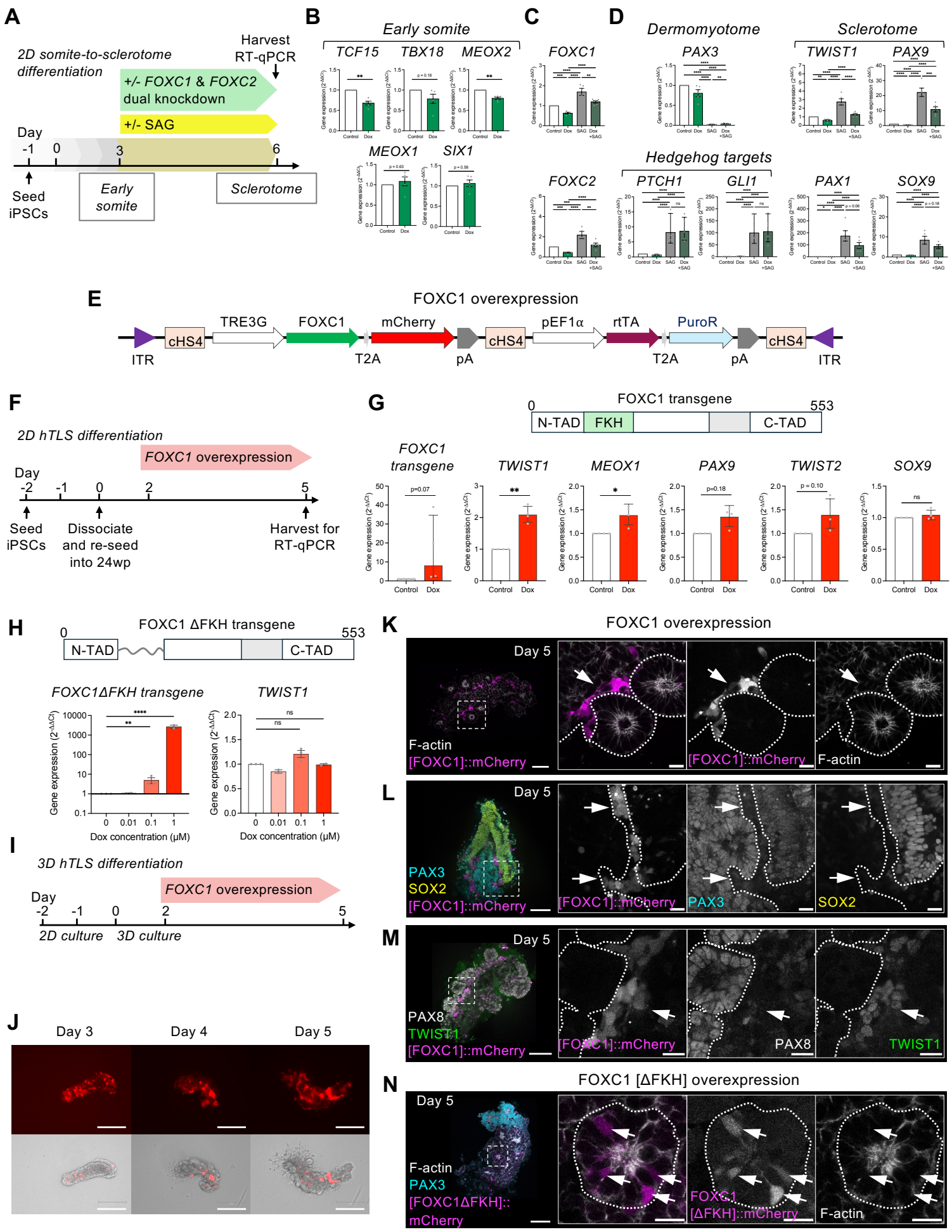

A
TBX6
SOX2
FOXC1
FOXC2
OSR1
Gene expression (2^-ΔΔCt)
Control
Dox
B
PC2: 10% variance
PC1: 85% variance
PC3: 2% variance
control
dox
C
Pearson correlation
control
dox
D
GO:BP
axonogenesis
muscle tissue development
ossification
pattern specification process
embryonic organ development
regulation of nervous system development
regulation of neuron projection development
positive regulation of cell development
mesenchyme development
cell−substrate adhesion
regulation of neurogenesis
epithelial tube morphogenesis
regulation of cell junction assembly
mesoderm development
roof of mouth development
Gene ratio
Gene count
FDR
E
PCW3 human embryo lineage gene set enrichment against ranked Dox-TBX6-induced genes
Paraxial Mesoderm
Presomitic Mesoderm
Myocyte Progenitor
Intermediate Mesoderm
Endothelium
Lateral plate Mesoderm 1
Endoderm
Primitive Erythroid
Somite
Lateral plate Mesoderm 2
Cardiac Myocyte
Mesenchyme
Notochord
Epithelium
Neural Tube 2
Neural Crest
Neural Tube 3
Neural Tube 1
NMP
enriched
depleted
FDR
Enrichment within ranked TBX6-induced gene list
F
PCW2 human embryo (Tyser et al.)
Advanced Mesoderm
Axial Mesoderm
Ectoderm (Amniotic/Embryonic)
Emergent Mesoderm
Endoderm
Epiblast
Erythroblasts
Extraembryonic Mesoderm
Hemato−Endothelial Progenitors
Nascent Mesoderm
Primitive Streak
UMAP 1
UMAP 2
G
TBX6 expression at PCW2
H
Dox-TBX6 module projected onto PCW2 dataset
I
Dox-TBX6 module expression by cluster (PCW2 dataset)
Extraembryonic Mesoderm
Advanced Mesoderm
Nascent Mesoderm
Emergent Mesoderm
Endoderm
Hemato-Endothelial Progenitors
Ectoderm (Amniotic/Embryonic)
Axial Mesoderm
Erythroblasts
Primitive Streak
Epiblast
TBX6 module score
J
PCW3 human embryo (Zeng et al.)
Cardiac Myocyte
Endoderm
Endothelium
Epithelium
Intermediate Mesoderm
Lateral plate Mesoderm 1
Lateral plate Mesoderm 2
Mesenchyme
Myocyte Progenitor
Neural Crest
Neural Tube 1
Neural Tube 2
Neural Tube 3
NMP
Notochord
Paraxial Mesoderm
Presomitic Mesoderm
Primitive Erythroid
Somite
K
TBX6 expression at PCW3
L
Dox-TBX6 module projected onto PCW3 dataset
M
Dox-TBX6 module expression by cluster (PCW3 dataset)
TBX6 module score

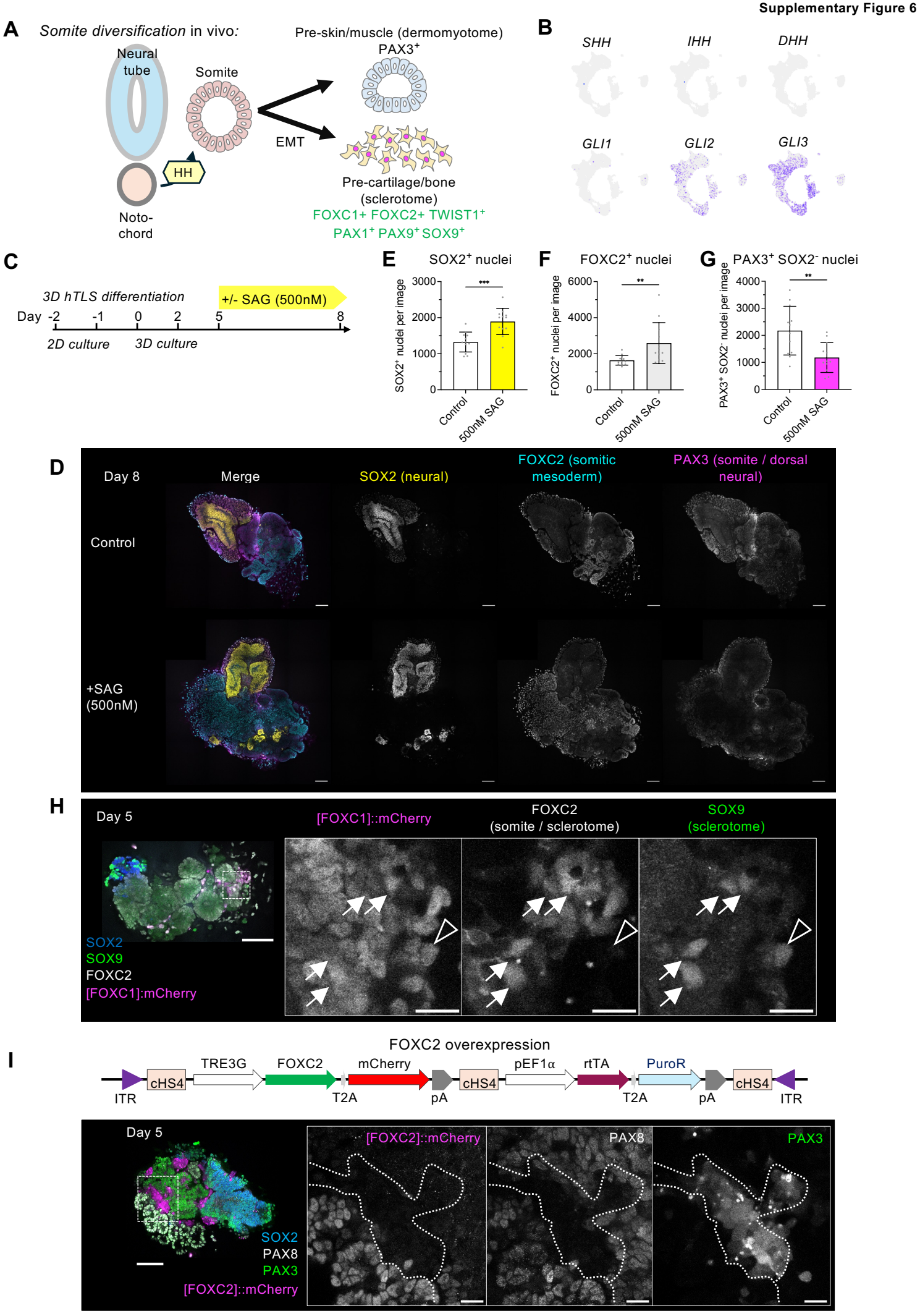

A
3D hTLS differentiation
+/- TBX6 / mCherry overexpression
Day
-2
-1
0
2
5
2D culture
3D culture
Harvest for IF
B
cHS4
TRE3G
TBX6
FKBP-V
mCherry
pEF1α
rtTA
PuroR
ITR
T2A
pA
C
[TBX6]::mCherry
FOXC2
PAX8
SOX2
Day 5
PAX8 (kidney)
FOXC2 (somite)
D
[control]::mCherry
E
Protein
dTag
Dox
mCherry
TBX6
Time
E'
Varying length of TBX6 activity
2D / 3D hTLS differentiation
TBX6-FKBP overexpression
Seed iPSCs
Dissociate and re-seed into 24wp or aggregate
Harvest for RT-qPCR or IF
F
TBX6
FOXC1
MEOX1
Gene expression (2-ΔΔCt)
Ctl
Dox d-1
d0
d1
d2
+Dox d-1 + dTag
G
[TBx6]::mCherry
PAX3 (somite/neural)
SOX2 (neural)
+Dox, no dTag (6 days of TBX6)
+Dox +dTag d0 (24 h of TBX6)
H
+Dox +dTag d2 (72 h of TBX6)

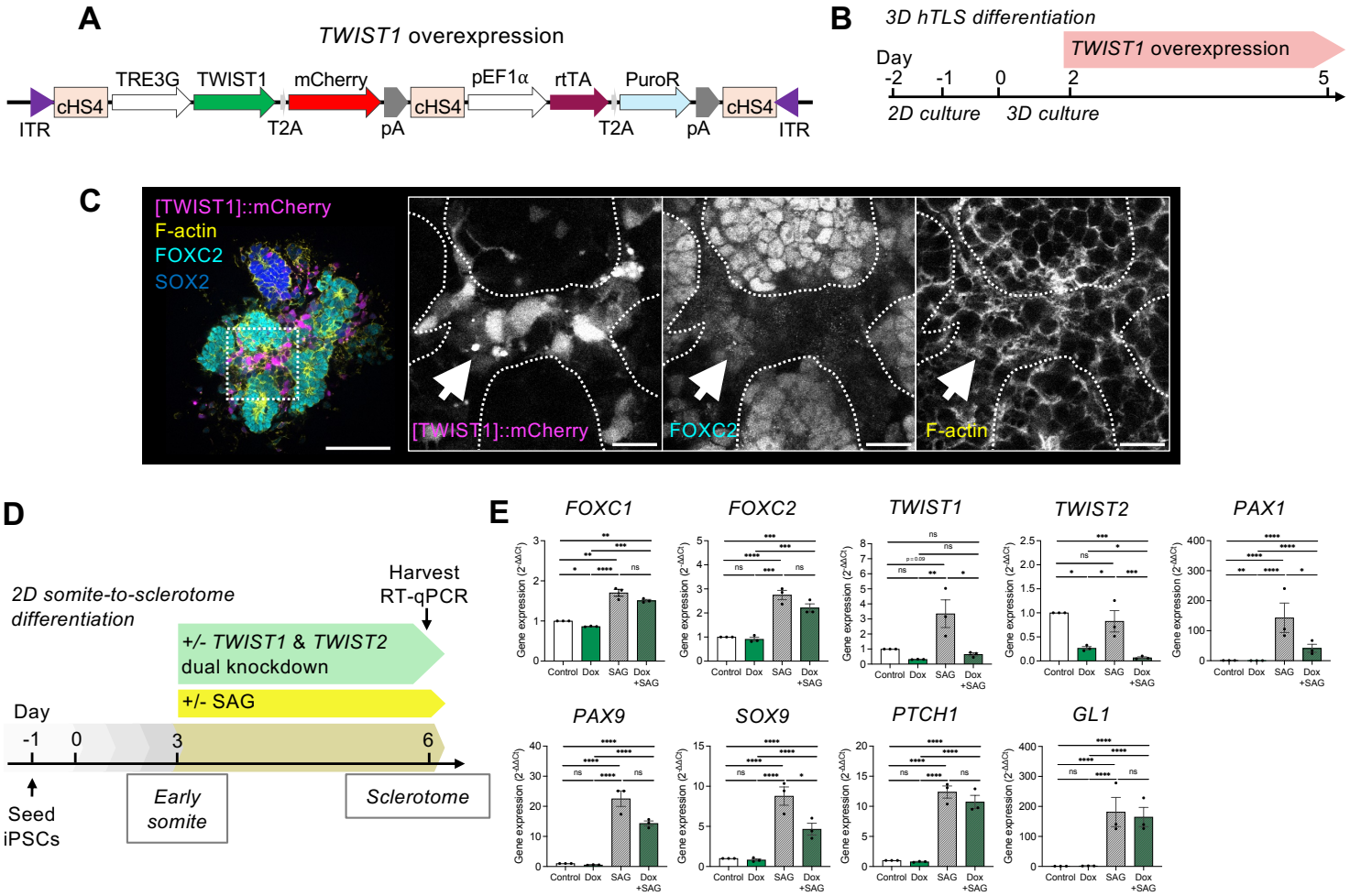

A
In Stemflex
+/-dTag
TBX6 overexpression
Day
-1
0
1
2
Seed iPSCs
Harvest for IF
B
F-Actin
[TBX6]::mCherry
TBX6
DAPI
No Dox
48h Dox
48h Dox + 24h dTag
C
3D hTLS protocol
+dTag
TBX6 overexpression
Day
-2
-1
0
2
5
Seed iPSCs
Aggregate
Harvest for IF
D
Day 5
[TBX6]::mCherry
TBX6
PAX8 (kidney)
PAX3 (somite/neural)
Dox d-1
+ dTag d0
(24 h of TBX6 activity)


A
Mesoderm subset
Somite progenitor
UMAP 2
UMAP 1
B
Somite progenitor vs rest of cells
depleted
enriched
ZNF704
MEOX1
PAX3
DACH2
GLI2
EBF1
ZNF608
EBF3
FOXP2
MYBL1
EBF2
TBX18
SALL1
SIX1
CREB5
ZBTB38
CSRNP1
FOXC1
TCF15
FOXC2
−log10(FDR)
log2 fold change
C
Inducible dual-gene CRISPRi strategy for FOXC1 and FOXC2
FOXC1
gRNA PuroR
FOXC2
gRNA NeoR
* Dox-inducible dCas9-KRAB-IRES-GFP donor plasmids
PiggyBac transposase
Dual selection (puro + neo)
Dual FOXC1 and FOXC2 CRISPRi cells
Somitic mesoderm differentiation (+/- Dox)
D
2D somitic mesoderm differentiation
Bulk RNA-seq
+/- Dual FOXC1 and FOXC2 knockdown
Day
-1
0
1
2
3
Seed iPSCs
Primitive Streak
Paraxial Mesoderm
Early somite
E
Dual FOXC1 and FOXC2 knockdown
depleted
enriched
RIPPLY1
FOXC2
FOXC1
TCF15
SIX1
TWIST1
PAX3
HEYL
HES5
EPCAM
MYOF
MYH7B
MEOX1
EYA2
NPHS1
JAG1
CDH1
MYH11
NPHS2
HAND1
MYLK3
MYMX
OSR1
MYBPC2
CDH6
−log10(FDR)
log2 fold change
F
Motif families enriched at FOXC1/2-dependent regulatory regions
Motifs with BH q < 0.01
Homeobox
bHLH
Forkhead (FOX)
Zinc finger homeodomain
Zinc finger
POU homeobox
HMG box
Nuclear receptor
TEA domain (TEAD)
Homeobox (HB)
Homeobox-bHLH
CTF/NF1
TCP domain
Forkhead-bHLH
Paired homeobox
G
3D hTLS differentiation
+/- Dual FOXC1 and FOXC2 knockdown
Day
-2
-1
0
2
5
2D culture
3D culture
Seed iPSCs
Harvest for IF
H
SOX2+ nuclei
SOX2+ nuclei per image
ns
Control
Dox
I
PAX8+ nuclei
PAX8+ nuclei per image
ns
Control
Dox
J
PAX3+ SOX2- nuclei
PAX3+ SOX2- nuclei per image
***
Control
Dox
K
Dual FOXC1 and FOXC2 knockdown
Control
1μM Dox
GFP
SOX2
PAX3
PAX8


A
2D somitic mesoderm differentiation
Day
-1
0
1
2
3
Seed iPSCs
Primitive Streak
Paraxial Mesoderm
Early somite
B
RT-qPCR time courses
OCT4
NANOG
TBXT
CDX2
TBX6
MSGN1
FOXC2
MEOX1
Gene expression (2^-ΔCt)
D0 D1 D2 D3
C
ITR
TRE3G
KRAB dCas9
EGFP
IRES
pA
cHS4
pU6
FOXC1 gRNA
FOXC2 gRNA
pEF1α
rtTA
PuroR
NeoR
T2A
D
PC2: 7% variance
PC1: 90% variance
control
dox
E
Pearson correlation
0
F
Early mesodermal genes
Normalized expression (log10 scale)
TBX1
CDX2
TBX6
MSGN1
G
Somite genes downregulated
Kidney genes upregulated
Muscle genes upregulated
Normalized expression (log10 scale)
RIPPLY1
FOXC1
FOXC2
MEOX1
MEOX2
TCF15
TWIST1
SIX1
PAX3
NPHS1
NPHS2
EYA2
OSR1
EPCAM
CDH1
JAG1
HEYL
HES5
MYMX
MYH11
MYBPC2
MYLK3
MYH7B
MYOF
H
GO: BP
UP
DOWN
muscle contraction
heart morphogenesis
actomyosin structure organization
mesenchymal cell differentiation
renal system development
pattern specification process
skeletal system morphogenesis
embryonic organ development
embryonic skeletal system development
stem cell differentiation
Gene ratio
FDR
I
GSEA of PCW3-5 human embryo lineage markers, Zeng et al. 2025
Cardiomyocyte
Endoderm1
Epithelium
Myocyte
Mesenchyme
Immune Cell
Somatic Mesoderm
Intermediate Mesoderm Lineage
Endoderm2
NMP
Endothelium
Spinal Cord Motor Neuron 1
Spinal Cord Motor Neuron 2
Spinal Cord Motor Neuron 3
Erythroid
Inhibitory Neuron
Spinal cord progenitor
Neural Crest 1
Forebrain progenitor
Hindbrain progenitor
DRG
Pharyngeal Mesoderm
Mesoderm Precursor
Midbrain progenitor
Vertebral Arches
Neural Crest 2
Vertebral Body
Myotome
Proliferating Lateral Plate Mesoderm
Proliferating Sclerotome 2
Proliferating Sclerotome 1
Proliferating Dermamyotome
Proliferating Sclerotome 3
enriched
depleted
FDR
Enrichment within ranked FOXC1/2-dependent gene list
J
GSEA of PCW6.5-7 human heart cell type markers, Asp et al., 2019
Atrial cardiomyocytes
Ventricular cardiomyocytes
Myoz2–enriched cardiomyocytes
Erythrocytes
Immune cells
Cardiac neural crest cells & Schwann progenitor cells
Endothelium / pericytes / adventia
Epicardial cells
Fibroblast–like
Capillary endothelium
Epicardium–derived cells
Smooth muscle cells / fibroblast–like
FDR
Enrichment within ranked FOXC1/2-dependent gene list
K
GSEA of fetal human nephron cell type markers, Stewart et al., 2018
Podocyte
Pelvic epithelium – distal UB
Proximal UB
Proximal S shaped body
Proximal tubule
Proximal renal vesicle
Cap mesenchyme
Distal renal vesicle
Proliferating distal renal vesicle
Proliferating cap mesenchyme
FDR
Enrichment within ranked FOXC1/2-dependent gene list
L
Day 5 hTLS, dual FOXC1 and FOXC2 knockdown (Dox from day 2)
Control
1µM Dox
GFP
SOX2
FOXC2
PAX8
M
2D hTLS differentiation
+/-TBX6 overexpression
+/-Dual FOXC1 and FOXC2 knockdown
Day
-2
-1
0
2
5
Seed iPSCs
RT-qPCR
N
TWIST1
TWIST2
MEOX1
MEOX2
TBX18
TBX6
FOXC1
FOXC2
TCF15
SIX1
MESP2
OSR1
Gene expression (2^-ΔCt)
Ctl
Dox
Dox-TBX6
Dox-TBX6-dCas9-F1F2


A
2D somite-to-sclerotome differentiation
Harvest RT-qPCR
+/- FOXC1 & FOXC2 dual knockdown
+/- SAG
Day
-1
0
3
6
Seed iPSCs
Early somite
Sclerotome
B
Early somite
TCF15
TBX18
MEOX2
MEOX1
SIX1
C
FOXC1
FOXC2
D
Dermomyotome
PAX3
Sclerotome
TWIST1
PAX9
Hedgehog targets
PTCH1
GLI1
PAX1
SOX9
E
FOXC1 overexpression
TRE3G
FOXC1
mCherry
pEF1α
rtTA
PuroR
cHS4
ITR
T2A
pA
F
2D hTLS differentiation
FOXC1 overexpression
Day
-2
-1
0
2
5
Seed iPSCs
Dissociate and re-seed into 24wp
Harvest for RT-qPCR
G
FOXC1 transgene
0
553
N-TAD
FKH
C-TAD
FOXC1 transgene
TWIST1
MEOX1
PAX9
TWIST2
SOX9
Control
Dox
H
FOXC1 ΔFKH transgene
FOXC1ΔFKH transgene
TWIST1
Gene expression (2^-ΔΔCt)
Dox concentration (μM)
I
3D hTLS differentiation
FOXC1 overexpression
2D culture
3D culture
J
Day 3
Day 4
Day 5
K
FOXC1 overexpression
Day 5
F-actin
[FOXC1]::mCherry
L
PAX3
SOX2
M
PAX8
TWIST1
N
FOXC1 [ΔFKH] overexpression
[FOXC1ΔFKH]::mCherry
FOXC1 [ΔFKH]::mCherry


A
Somite diversification in vivo:
Neural tube
Somite
Noto-chord
HH
EMT
Pre-skin/muscle (dermomyotome)
PAX3+
Pre-cartilage/bone (sclerotome)
FOXC1+ FOXC2+ TWIST1+
PAX1+ PAX9+ SOX9+
B
SHH
IHH
DHH
GLI1
GLI2
GLI3
C
3D hTLS differentiation
+/- SAG (500nM)
Day -2 -1 0 2 5 8
2D culture
3D culture
E
SOX2+ nuclei
SOX2+ nuclei per image
Control
500nM SAG
F
FOXC2+ nuclei
FOXC2+ nuclei per image
G
PAX3+ SOX2- nuclei
PAX3+ SOX2- nuclei per image
D
Day 8
Merge
SOX2 (neural)
FOXC2 (somitic mesoderm)
PAX3 (somite / dorsal neural)
Control
+SAG (500nM)
H
Day 5
[FOXC1]::mCherry
FOXC2 (somite / sclerotome)
SOX9 (sclerotome)
SOX2
SOX9
FOXC2
[FOXC1]:mCherry
I
FOXC2 overexpression
ITR
cHS4
TRE3G
FOXC2
mCherry
T2A
pA
cHS4
pEF1α
rtTA
T2A
PuroR
pA
cHS4
ITR
Day 5
SOX2
PAX8
PAX3
[FOXC2]::mCherry
[FOXC2]::mCherry
PAX8
PAX3


A
TWIST1 overexpression
cHS4
TRE3G
TWIST1
mCherry
cHS4
pEF1α
rtTA
PuroR
cHS4
ITR
T2A
pA
T2A
pA
ITR
B
3D hTLS differentiation
TWIST1 overexpression
Day
-2
-1
0
2
5
2D culture
3D culture
C
[TWIST1]::mCherry
F-actin
FOXC2
SOX2
[TWIST1]::mCherry
FOXC2
F-actin
D
2D somite-to-sclerotome differentiation
Harvest RT-qPCR
+/- TWIST1 & TWIST2 dual knockdown
+/- SAG
Day
-1
0
3
6
Seed iPSCs
Early somite
Sclerotome
E
FOXC1
FOXC2
TWIST1
TWIST2
PAX1
PAX9
SOX9
PTCH1
GL1
Gene expression (2-ΔΔCt)
Control
Dox
SAG
Dox +SAG