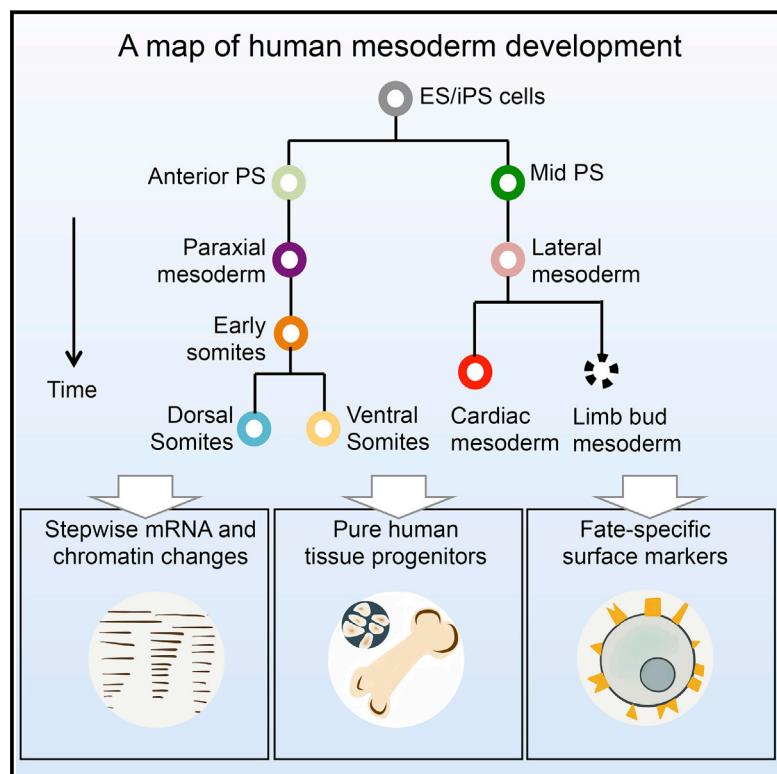


Mapping the Pairwise Choices Leading from Pluripotency to Human Bone, Heart, and Other Mesoderm Cell Types

Graphical Abstract



Authors

Kyle M. Loh, Angela Chen,
Pang Wei Koh, ..., Philip A. Beachy,
Lay Teng Ang, Irving L. Weissman

Correspondence

anglt1@gis.a-star.edu.sg (L.T.A.),
irv@stanford.edu (I.L.W.)

In Brief

The lineage roadmap of human mesoderm development reveals transient developmental processes such as human somite segmentation and enables the generation and isolation of transplantable human bone and heart progenitors.

Highlights

- Stepwise map of competing signals guiding human mesoderm development
- Efficient human mesoderm induction by blocking formation of unwanted fates
- ESC-derived human bone progenitors and heart precursors engraft *in vivo*
- A transient segmentation program in human embryogenesis marked by *HOPX*



Mapping the Pairwise Choices Leading from Pluripotency to Human Bone, Heart, and Other Mesoderm Cell Types

Kyle M. Loh,^{1,7} Angela Chen,^{1,7} Pang Wei Koh,² Tianda Z. Deng,¹ Rahul Sinha,¹ Jonathan M. Tsai,¹ Amira A. Barkal,¹ Kimberle Y. Shen,¹ Rajan Jain,⁵ Rachel M. Morganti,¹ Ng Shyh-Chang,⁶ Nathaniel B. Fernhoff,¹ Benson M. George,¹ Gerlinde Wernig,³ Rachel E.A. Salomon,¹ Zhenghao Chen,¹ Hannes Vogel,³ Jonathan A. Epstein,⁵ Anshul Kundaje,² William S. Talbot,¹ Philip A. Beachy,^{1,4} Lay Teng Ang,^{6,8,*} and Irving L. Weissman^{1,8,*}

¹Department of Developmental Biology, Institute for Stem Cell Biology & Regenerative Medicine, Ludwig Center for Cancer Stem Cell Biology and Medicine

²Departments of Genetics and Computer Science,

³Department of Pathology

⁴Department of Biochemistry, Howard Hughes Medical Institute
Stanford University School of Medicine, CA 94305, USA

⁵Department of Cell and Developmental Biology, Cardiovascular Institute, University of Pennsylvania, Philadelphia, PA 19104, USA

⁶Stem Cell & Regenerative Biology Group, Genome Institute of Singapore, Singapore 138672, Singapore

⁷Co-first author

⁸Co-senior author

*Correspondence: anglt1@gis.a-star.edu.sg (L.T.A.), irv@stanford.edu (I.L.W.)

<http://dx.doi.org/10.1016/j.cell.2016.06.011>

SUMMARY

Stem-cell differentiation to desired lineages requires navigating alternating developmental paths that often lead to unwanted cell types. Hence, comprehensive developmental roadmaps are crucial to channel stem-cell differentiation toward desired fates. To this end, here, we map bifurcating lineage choices leading from pluripotency to 12 human mesodermal lineages, including bone, muscle, and heart. We defined the extrinsic signals controlling each binary lineage decision, enabling us to logically block differentiation toward unwanted fates and rapidly steer pluripotent stem cells toward 80–99% pure human mesodermal lineages at most branchpoints. This strategy enabled the generation of human bone and heart progenitors that could engraft in respective *in vivo* models. Mapping step-wise chromatin and single-cell gene expression changes in mesoderm development uncovered somite segmentation, a previously unobservable human embryonic event transiently marked by *HOPX* expression. Collectively, this roadmap enables navigation of mesodermal development to produce transplantable human tissue progenitors and uncover developmental processes.

INTRODUCTION

Waddington's developmental landscape drawings (Waddington, 1940) depicted how differentiating stem cells negotiate a

cascade of branching lineage choices, avoiding alternate fates at each juncture to decisively commit to a single lineage (Graf and Enver, 2009). To navigate this brachiating landscape and efficiently differentiate stem cells into desired cell-types for regenerative medicine, one must (1) catalog transitional lineage intermediates, (2) map the sequence of pairwise lineage choices through which such intermediates are formed, and (3) discover the positive and negative signals that specify or repress cell fate at each lineage branchpoint. Despite successes in charting lineage intermediates in mammalian tissues, key lineage branchpoints remain controversial and it has been impossible to systematically identify the extracellular signals that control cell fate at each branchpoint.

With the three above goals in mind, here we map the landscape of human mesoderm development in order to coherently guide stem-cell differentiation (Figure 1A). Mesoderm development begins with differentiation of pluripotent epiblast cells into the primitive streak, which then segregates into paraxial and lateral mesoderm, among other lineages (Lawson et al., 1991; Rosenquist, 1970; Tam and Beddington, 1987). Paraxial mesoderm segments into somites, which are fundamental building blocks of trunk tissue (Figure 1A, purple shading) (Pourquié, 2011). Somites are then patterned along the dorsal-ventral axis. The ventral somite (sclerotome) generates the bone and cartilage of the spine and ribs, whereas the dorsal somite (dermomyotome) yields brown fat, skeletal muscle, and dorsal dermis (Christ and Scaal, 2008). Separately, lateral mesoderm (Figure 1B, red shading) gives rise to limb bud mesoderm (Tanaka, 2013) and cardiac mesoderm (Später et al., 2014), the latter of which generates cardiomyocytes and other heart constituents. Various transcription factors (TFs) and signaling molecules regulating mesoderm development in model vertebrates have been identified, broadly outlining the developmental landscape (Kimmelman, 2006; Schier and Talbot, 2005; Tam and Loebel, 2007).

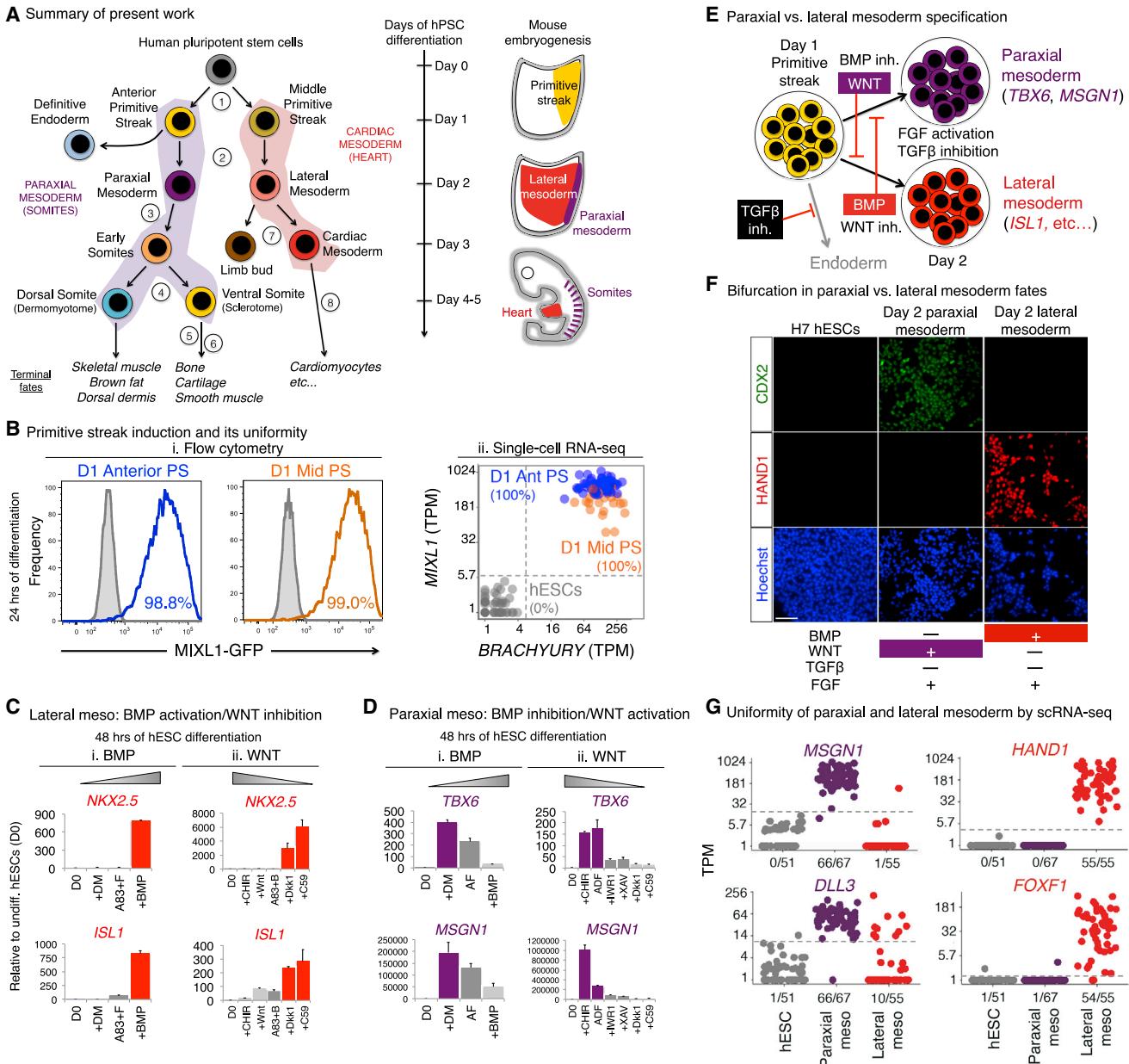


Figure 1. Formation of Human Primitive Streak and Its Bifurcation into Paraxial and Lateral Mesoderm

(A) Each lineage step labeled with a circled number, corresponding to respective sections in the main text and Figure 7A.

(B) FACS (fluorescence-activated cell sorting) of *MIXL1-GFP* hESC (Davis et al., 2008) after 24 hr in anterior or mid PS induction (left). All cells coexpress *BRACHYURY* and *MIXL1* by scRNA-seq. Each dot depicts a single cell (right).

(C) BMP induces, whereas WNT inhibits, lateral mesoderm from the PS on day 1–2. (i) qPCR of day 1 PS treated with BMP4 or a BMP inhibitor (DM3189) for 24 hr (in the context of A8301 + FGF2 [AF]); (2) qPCR of day 1 PS treated with WNT agonists (CHIR99021 or WNT3A) or WNT inhibitors (300 ng/mL Dkk1 or 1 μM C59) for 24 hr (in the context of A8301 + BMP4 [AB]); error bars = SEM for this and all other qPCR experiments.

(D) BMP inhibits, whereas WNT induces, paraxial mesoderm from the PS on day 1–2. (i) qPCR of day 1 PS treated with BMP4 or a BMP inhibitor (DM3189) for 24 hr (in the context of A8301 + FGF2 [AF]); (ii) qPCR of day 1 PS treated with WNT agonist (3 μM CHIR99021) or WNT inhibitors (2 μM IWR1, 1 μM XAV939 or C59, or 300 ng/mL Dkk1) for 24 hr (in the context of A8301 + DM3189 + FGF2 [ADF]).

(E) Lateral versus paraxial mesoderm bifurcation.

(F) CDX2 and HAND1 staining of day 2 H7-derived paraxial or lateral mesoderm populations or undifferentiated hESCs (scale bar, 100 μm), with Hoechst nuclear staining.

(G) scRNA-seq of day 2 lateral mesoderm or *DLL3*⁺ sorted paraxial mesoderm. Each dot depicts a single cell. % of marker-positive cells above the dotted TPM (transcripts per million) threshold.

See also Figures S1 and S2.

Yet gaps in our understanding have been revealed by efforts to differentiate human pluripotent stem cells (hPSCs) to various mesoderm cell types in a dish. Human mesoderm has remained obscure because it first forms in gestational weeks 2–4 (O’Rahilly and Müller, 1987), when it is impermissible to access human embryos. There has been some success in generating human mesoderm derivatives from PSCs, including paraxial (Cheung et al., 2012; Mendjan et al., 2014; Umeda et al., 2012) and heart (Ardehali et al., 2013; Burridge et al., 2014; Chong et al., 2014; Lian et al., 2012; Mendjan et al., 2014) cell types. However, because the sequence of lineage branchpoints and the identity of inductive or repressive signals at every developmental step remain incompletely understood, some mesodermal differentiation protocols take weeks to months and generate heterogeneous mixtures of cell types comprising a subset of the desired lineage along with other contaminating lineages. Prior studies indicated that ACTIVIN/NODAL/TGF β (henceforth referred to as TGF β), BMP, FGF, and WNT broadly induce mesoderm from PSCs (Cheung et al., 2012; Gertow et al., 2013), the importance of dynamic WNT signaling during cardiac induction (Burridge et al., 2014; Lian et al., 2012; Ueno et al., 2007), and that BMP inhibits paraxial mesoderm formation (Cheung et al., 2012; Umeda et al., 2012). Nonetheless, the challenges faced by current differentiation strategies provide an impetus to better understand the complex process of mesoderm development.

Here, we delineate a roadmap for human mesoderm development and define the sequential steps through which pluripotent cells elaborate a diversity of mesodermal progeny. At many developmental steps, we discovered the minimal combinations of signals sufficient to efficiently induce each human mesodermal fate and showed that it was key to define both *inductive* and *repressive* cues at each step (Table S1). It was critical to define how “unwanted” cell fates were specified in order to logically block their formation and steer stem-cell differentiation down a singular developmental path.

This knowledge guided the efficient differentiation of PSCs into a variety of human mesoderm fates within several days, without recourse to gene modification or serum treatment. The authenticity of the induced cells was confirmed by their ability to engraft in vivo and by single-cell RNA-seq to test for lineage identity and homogeneity. Global RNA-seq and ATAC-seq (assay for transposase-accessible chromatin with high throughput sequencing) analyses revealed the stepwise changes in gene expression and sequential opening and closing of chromatin elements at each developmental transition. Collectively, we chart the developmental landscape of human mesoderm formation and uncover the sequential signaling, transcriptional, and chromatin changes at each lineage step. We directly demonstrate the utility of this reference map in guiding stem-cell differentiation, producing transplantable cells for eventual use in regenerative medicine, improving our understanding of human development, and uncovering the putative origins of certain human congenital malformations.

RESULTS

Induction of Anterior and Mid Primitive Streak

Primitive streak (PS) formation from pluripotent cells is the first step in mesoderm development (Figure 1A, step 1). We gener-

ated a >98% pure MIXL1-GFP $^+$ human PS population within 24 hr of PSC differentiation (Figures 1Bi, S1A, and S1B) by activating TGF β , WNT, and FGF and inhibiting PI3K signalling, in the presence or absence of exogenous BMP (Figures S1C–S1E) (Gertow et al., 2013; Loh et al., 2014; Schier and Talbot, 2005; Tam and Loebel, 2007). Attesting to the uniformity of differentiation, single-cell RNA-seq (scRNA-seq) of bulk PS populations revealed that 100% of analyzed cells coexpressed PS TFs *BRACHYURY* and *MIXL1* (Figure 1Bii).

In the vertebrate embryo, different anterior-posterior regions of the PS produce distinct mesoderm derivatives (Lawson et al., 1991; Rosenquist, 1970; Tam and Beddington, 1987). Likewise, hPSC-derived anterior PS induced in the presence of an anteriorizing TGF β signal was competent to form paraxial mesoderm (Figure S1E). By contrast, mid PS induced in the presence of both anteriorizing TGF β and posteriorizing BMP harbored maximal potential to form lateral mesoderm/cardiac progenitors (Figures S1E–S1I). Thus, as in model organisms, human PS is not a single lineage but comprises several subtypes each already partially committed to one type of downstream mesoderm (Figures S1J–SL). Together, these >98% pure human PS populations provided a starting point to understand the subsequent divergence of distinct mesoderm subtypes.

Bifurcation of Paraxial versus Lateral Mesoderm from Primitive Streak by Competing WNT and BMP Signals

In vivo, the PS forms definitive endoderm, paraxial mesoderm, and lateral mesoderm, but how these lineages are segregated is not well understood (Figure 1A, step 2). After PS induction (day 0–1), TGF β specified endoderm (Figures S2A and S2B) (Loh et al., 2014) while TGF β inhibition blocked endoderm formation and instead induced mesoderm (Figure S2B–S2E). Since TGF β inhibition and FGF/ERK activation (Figures S2B and S2F) for 24 hr (day 1–2) created a permissive context for both paraxial and lateral mesoderm formation, we sought how these mutually exclusive mesoderm subtypes become distinguished.

Countervailing BMP and WNT signals respectively induced human lateral versus paraxial mesoderm and each repressed the formation of the mutually exclusive lineage on day 1–2 of PSC differentiation, driving the bifurcation of these two mesoderm subtypes from the PS (summarized in Figure 1E). Exogenous BMP induced lateral mesoderm and repressed paraxial mesoderm (Figures 1Ci, 1Di, S2E, and S2H). By contrast, blocking BMP signaling abrogated lateral mesoderm and instead expanded paraxial mesoderm (Figures 1Ci, 1Di, and S2H). This reveals a key function of BMP in patterning human mesoderm akin to its activity in chick (Tonegawa et al., 1997).

Conversely, WNT promoted human paraxial mesoderm and repressed lateral mesoderm. WNT activation (by GSK3 inhibition) induced paraxial markers while suppressing lateral/cardiac markers (Figures 1Cii and 1Dii). By contrast, WNT inhibition elicited lateral mesoderm while blocking the paraxial fate (Figures 1Cii and 1Dii). Therefore, WNT controls the allocation of human paraxial versus lateral mesoderm, logically linking the requirement for WNT in mouse paraxial mesoderm formation (Aulehla et al., 2008) to the ability of WNT to repress early cardiac mesoderm in *Xenopus* (Schneider and Mercola, 2001).

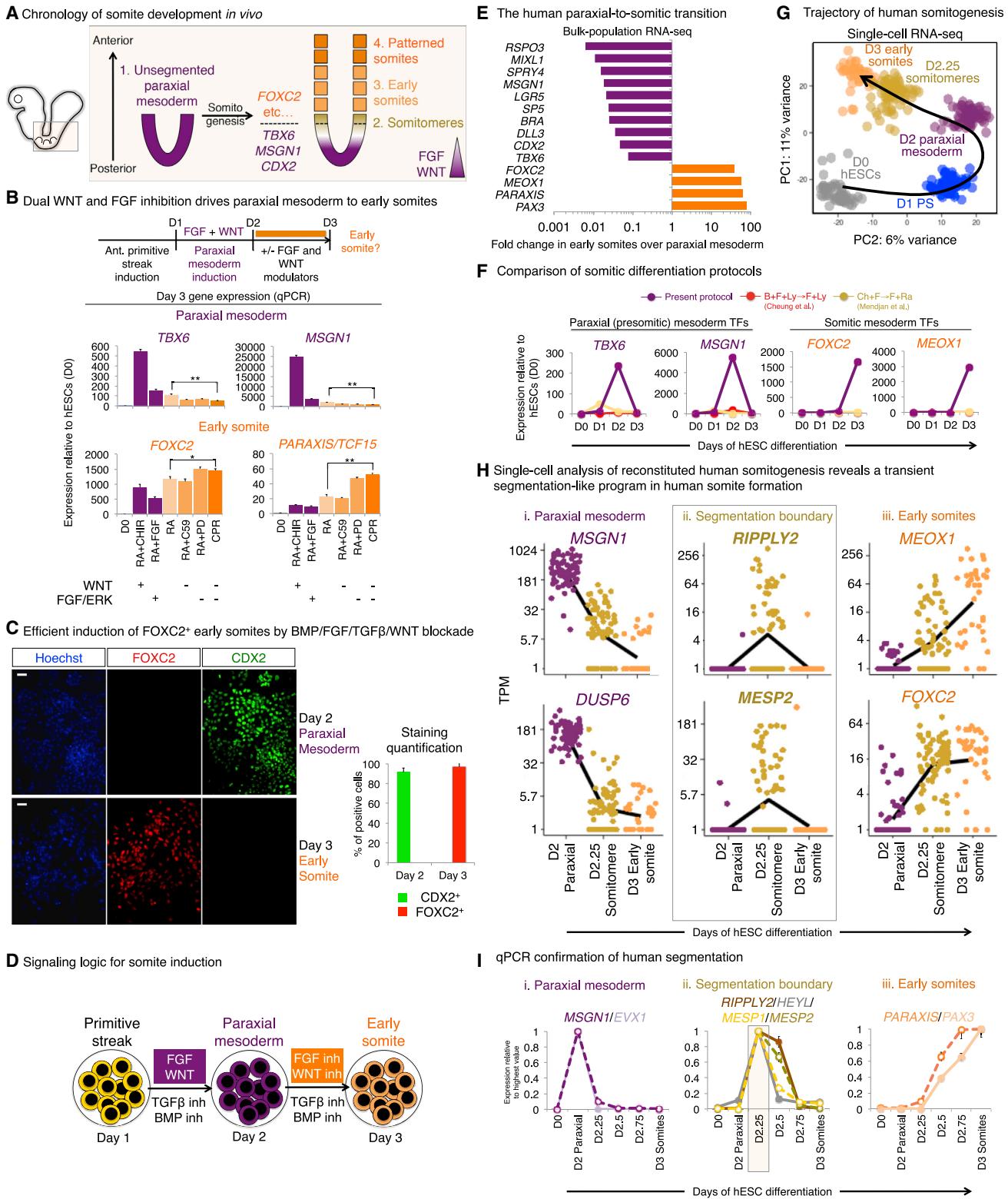


Figure 2. Human Paraxial Mesoderm Differentiation into Early Somites Passes through an Ephemeral Somitomere-Like State

(A) Paraxial mesoderm segmentation into somites *in vivo*.

(B) To reveal how WNT and FGF/ERK control paraxial mesoderm progression to early somites, day 2 H7-derived paraxial mesoderm was treated with RA (2 μ M) for 24 hr, in combination with a WNT agonist (CHIR, 3 μ M), a WNT inhibitor (C59, 1 μ M), FGF2 (20 ng/mL), an ERK inhibitor (PD0325901, 500 nM), or combined

(legend continued on next page)

In summary, on day 1–2 of hPSC differentiation, BMP inhibition and WNT activation induced paraxial mesoderm whereas, conversely, BMP activation and WNT inhibition specified lateral mesoderm from the PS within the permissive context of TGF β inhibition/FGF activation (Figure 1E). These two mutually exclusive signaling conditions produced either CDX2⁺HAND1⁻ paraxial mesoderm or CDX2^{lo/-}HAND1⁺ lateral mesoderm by day 2 of differentiation (Figure 1F and S2I). Tracking the bifurcation of paraxial versus lateral mesoderm fates by scRNA-seq confirmed the mutually exclusive marker expression in the two diverging populations at the level of single cells (Figure 1G). scRNA-seq showed that both human paraxial and lateral mesoderm populations were essentially uniform: 98.5% of paraxial mesoderm cells expressed *DLL3* and *MSGN1* whereas 98.1%–100% of lateral mesoderm cells expressed *HAND1* and *FOXF1* (Figure 1G).

Maturation of Paraxial Mesoderm into Early Somites by Combined BMP, ERK, TGF β , and WNT Inhibition

Having generated $91.2\% \pm 0.1\%$ pure TBX6⁺CDX2⁺ human paraxial mesoderm by day 2 of PSC differentiation (Figure S3A), we next sought to drive these cells into early somite progenitors (Figure 1A, step 3, and Figure 2A). During embryogenesis, the U-shaped sheet of paraxial (presomitic) mesoderm is progressively segmented at its anterior edge to generate spherical early somites (Figure 2a), due to lower anterior levels of FGF and WNT at the “wavefront” (Aulehla et al., 2008; Dubrulle et al., 2001).

Whereas paraxial mesoderm was specified on day 1–2 by FGF and WNT, the inhibition of FGF/ERK and WNT signaling on day 2–3 strongly downregulated paraxial mesoderm genes (e.g., *TBX6*, *MSGN1*) and upregulated early somite markers (e.g., *FOXC2*, *PARAXIS*, *MEOX1*; Figures 2B and S3B). Early somite markers were further upregulated when TGF β and BMP were inhibited (Figures S3C and S3D); therefore, we employed quadruple inhibition of these four pathways to drive near-complete conversion of day 2 CDX2⁺ paraxial mesoderm into $96.8\% \pm 5.7\%$ pure FOXC2⁺ early somite precursors by day 3 (Figures 2C–2E). Together, this identified the minimal signaling cues sufficient to efficiently generate human PS, paraxial mesoderm, and subsequently, early somite progenitors from PSCs within 72 hr of differentiation (Figure 2D) more robustly and rapidly than was previously possible (Cheung et al., 2012; Mendjan et al., 2014) (Figure 2F).

scRNA-seq reaffirmed the homogeneity of these various in-vitro-derived populations at different steps of human somitogenesis. 98.5%–100% of human paraxial mesoderm cells (sorted for *DLL1* expression; Supplemental Experimental Procedures) ex-

pressed paraxial markers *MSGN1* and *DUSP6* at day 2, yet these markers were sharply suppressed within 6 hr of FGF/WNT inhibition (by day 2.25) during differentiation toward somites (Figure 2Hi). Conversely, somite TFs *MEOX1* and *FOXC2* became expressed in the majority (83.3%–88.9%) of human early somite cells by day 3 (Figure 2Hii). Hence, this indicated that human somitogenesis was efficiently reconstituted in culture. We exploited this system to uncover developmental features of this process.

Single-Cell RNA-Seq Identifies a Conserved, Segmentation-Like Process in Human Somitogenesis and Infers Transcriptional Cofactor *HOPX* as a Marker

In vertebrate model organisms, somitogenesis entails transient expression of segmentation genes (for 30–120 min in fish and mice, respectively) at the waveform, inducing paraxial mesoderm to segment into somitomeres (prospective somites) and then into somites (Pourquié, 2011). Human somitogenesis begins at week 3 of gestation (O’Rahilly and Müller, 1987), but whether it also entails a transient segmentation molecular program has remained unclear because it is ethically impermissible to access or analyze week 3–5 human embryos.

Indeed, paraxial mesoderm cells in our reconstituted human somitogenesis system passed through an intermediate somitomere stage before differentiating into somites. scRNA-seq of human embryonic stem cells (hESCs) (day 0) differentiating into anterior PS (day 1), paraxial mesoderm (day 2), through an intermediate step (day 2.25), and into early somites (day 3) showed that single cells uniformly moved through a single inferred differentiation trajectory without overtly diverging branch points (Figure 2G), implying the homogeneity of cells at each lineage step. Principal component analysis (PCA) positioned the day 2.25 population between day 2 paraxial mesoderm and day 3 early somites (Figure 2G), suggesting that it is a true intermediate between presomitic/paraxial and somitic states.

Indeed, a subset of day 2.25 cells transiently expressed *MESP2*, *RIPPLY2*, and *HEY1* (Figures 2Hii and S4E), whose homologs are somitomere segmentation markers in model organisms (reviewed by Pourquié, 2011). qPCR confirmed a brief pulse of *MESP2*, *RIPPLY2*, and *HEY1* expression for several hours (~day 2.25–day 2.5) in the interval between paraxial and somitic marker expression (Figure 2I). Hence, there exists a transient somitomere-like transition point in human somitogenesis, arguing that human development entails an evolutionarily conserved segmentation process.

To discover additional somite segmentation markers, single-cell transcriptomes were aligned in “pseudotime” to infer the

WNT/ERK inhibition (C59+PD0325901+RA [CPR]) and qPCR was conducted (${}^{\prime}p < 0.05$, ${}^{**}p < 0.01$), showing WNT/ERK blockade enhances early somite induction (it was later found that exogenous RA was dispensable for early somite formation; Figure S3E).

(C) CDX2 and FOXC2 staining of BJC1-derived paraxial mesoderm (day 2) and early somite (day 3) populations (left) and quantification (right).

(D) FGF and WNT activation, followed by inhibition, induces human paraxial mesoderm and then early somites.

(E) Differentially expressed genes in day 2 paraxial mesoderm versus day 3 early somites (bulk-population RNA-seq).

(F) qPCR time-course comparison of H7 hESCs differentiated into somites using previous protocols (Cheung et al., 2012; Mendjan et al., 2014) or the current method.

(G) PCA of human somitogenesis scRNA-seq. Colors designate cell populations harvested at different time points. Each dot is a single cell.

(H) scRNA-seq of day 2, day 2.25, and day 3 hESC-derived populations. Dots depict single cells. Line indicates mean gene expression in all cells at each time point.

(I) Time-course qPCR of H7-derived cells.

See also Figure S3.

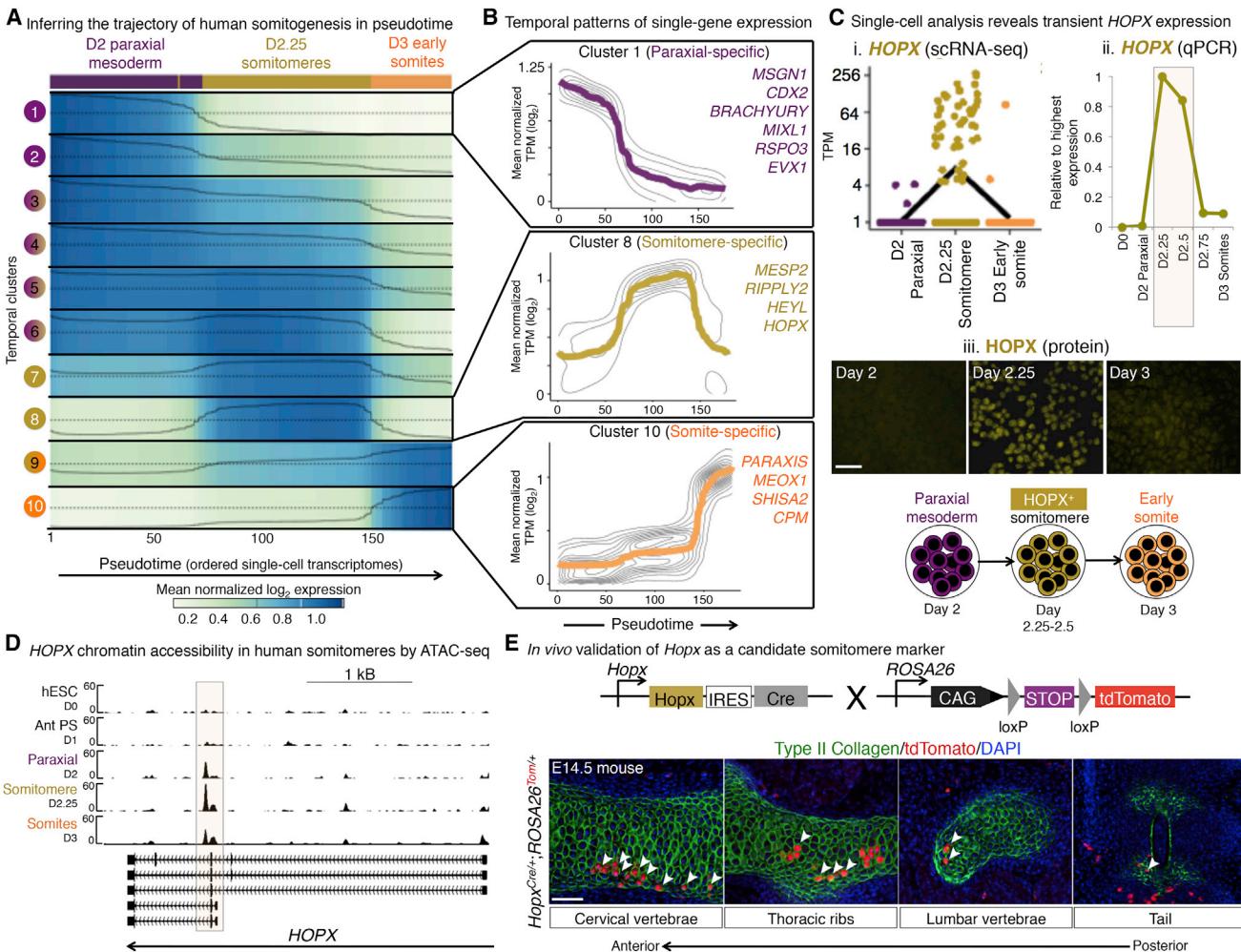


Figure 3. Single-Cell Analysis Captures a Transient *HOPX*⁺ Human Somitomere Progenitor State

(A) Heatmap of normalized scRNA-seq gene expression across the inferred trajectory of human somitogenesis. Each column reflects a single cell, with scRNA-seq paraxial mesoderm, somitomere, and early somite transcriptomes (colored blocks) ordered in pseudotime along the y axis ([Supplemental Experimental Procedures](#)). Genes were clustered into ten clusters (rows) by virtue of their expression kinetics across this pseudotime time course. Line indicates smoothed mean expression of all genes in the cluster across pseudotime.

(B) Mean expression (bold line) of all genes in each temporal cluster across pseudotime (with contours representing density of individual gene expression), with representative genes in each cluster noted.

(C) Transient *HOPX* expression during H7 hESC differentiation toward somites, shown by scRNA-seq (1), qPCR (2), and immunostaining (3); scale bar, 50 μm.

(D) ATAC-seq shows the *HOPX* locus is accessible in day 2.25 hESC-derived somitomeres (signal track, $-\log_{10}$ p values).

(E) Fate mapping progeny of *Hopx*⁺ cells in E14.5 *Hopx*-IRES-Cre;Ai9 embryos reveals contribution to the spine and ribs (labeled by type II collagen); scale bar, 50 μm.

order in which they arose along an inferred developmental path ([Supplemental Experimental Procedures](#); **Figure 3A**). This led to 44 candidate genes with transient somitomere-specific expression (**Figure 3B**), including *HOPX*, a homeodomain-containing transcriptional cofactor. By scRNA-seq, *HOPX* was specifically expressed in a subset of day 2.25 somitomere cells but neither day 2 paraxial mesoderm nor day 3 early somites (**Figure 3Ci**). qPCR (**Figure 3Cii**) and immunostaining (**Figure 3Ciii**) corroborated the transient expression of *HOPX* mRNA and protein during in vitro somitogenesis, paralleling the transitory expression of known somitomere markers (**Figure 2i**). Additionally, open chromatin analysis by ATAC-seq ([Buenrostro et al., 2013](#)) revealed that the *HOPX* locus was accessible in day 2.25 somitomere populations (**Figure 3D**). Thus, *HOPX* is briefly expressed in somitomeres and marks human somite segmentation.

Consistent with the notion that *Hopx* also marks somitomeres in mouse embryos, genetic lineage tracing showed that *Hopx*⁺ cells contributed to the ribcage, cervical, and lumbar vertebrae and tail (**Figure 3E**). Sparse labeling reflects the noted inefficiency of the *Hopx*-IRES-Cre driver allele ([Jain et al., 2015](#)). Collectively, this uncovers a transient molecular signature of somite segmentation conserved between human and other vertebrates and suggests *HOPX* as a possible component of this signature.

Bifurcation of Human Early Somite Progenitors into Sclerotome and Dermomyotome by HH and WNT Signals

After human somites are formed, how do they diverge into distinct derivatives (Figure 1A, step 4)? In vivo, early somites are patterned along their dorsal-ventral axis to generate sclerotome (ventral somite; precursor to bone, cartilage, and smooth muscle) and dermomyotome (dorsal somite; precursor to skeletal muscle, brown fat, and dorsal dermis; Figure 4A).

Starting from day 3 human early somite progenitors, HEDGEHOG (HH) induced sclerotome while WNT specified dermomyotome and each cross-antagonized the effect of the other on days 3–5 of differentiation (Figures 4B and S4A and as summarized in Figure 4H). By activating one signal while inhibiting the other, it was possible to specify one somitic derivative while blocking formation of the alternate fate.

HH activation together with WNT inhibition blocked dermomyotome formation, inducing a fairly uniform SOX9⁺TWIST1⁺ sclerotome population that expressed multiple sclerotome markers (PAX1, PAX9, NKX3.2/BAPX1, FOXC2, SOX9, and TWIST1) (Figures 4B, 4C, S4A, and S4B). Conversely, WNT activation together with HH blockade blocked sclerotome formation and instead exclusively specified dermomyotome (Figures 3B and S4A). If HH and WNT were simultaneously activated, neither sclerotome or dermomyotome was elicited (Figure S4A), indicating the importance of their mutually exclusive activation. Collectively, this shows that HH and WNT pattern human somites, analogous to their role in mice (Fan et al., 1995, 1997; Fan and Tessier-Lavigne, 1994), and demonstrates how their manipulation can separately generate human sclerotome or dermomyotome. Indeed, scRNA-seq showed that hPSC-derived sclerotome and dermomyotome populations formed in these two mutually exclusive signaling conditions constituted largely distinct clusters (Figure 4D) that diverged downstream of a common early somite progenitor state.

Human dermomyotome induction from early somites relied on WNT activation and initial HH inhibition (to block sclerotome induction) in addition to BMP (to induce PAX7; Figure S4C) and later-stage HH activation (to induce EN1; Figures S4D–S4F). In vivo, dermomyotome gives rise to skeletal muscle, and accordingly, human dermomyotome was also capable of generating MYH3⁺ skeletal muscle cells in vitro (Figure S4E).

Human Sclerotome Forms an Ectopic Bone In Vivo

Strikingly, human sclerotome (ventral somite) progenitors could generate an ectopic human bone in vivo (Figure 1a, step 5), reflecting the skeleton-forming fate of embryonic sclerotome (Christ and Scaal, 2008). Upon subcutaneous injection into immunodeficient mice, hESC-derived sclerotome formed ectopic bone-like structures containing both bone and cartilage (Figure 4Ei). Other tissues (i.e., epithelia) were not present (Figure S5A), indicating the developmental lineage restriction of sclerotome. Labeling sclerotome with constitutive GFP (Figure S5B), BCL2-2A-GFP (Figure S5C), or Luciferase (Figure S5D) reporters prior to transplantation confirmed that the subcutaneous GFP⁺ or luciferase⁺ bones (Figure 4E) were not derived from resident mouse cells.

Sclerotome-derived ectopic human bones self-organized themselves even though they were implanted without a patterned

matrix. Specifically, they harbored proliferative chondrocytes that progressed to hypertrophic chondrocytes and finally underwent ossification in a spatially choreographed fashion within column-shaped streams of cells (Figure 4F). They also recruited host blood vessels (Figure S5E), recapitulating a developmental endochondral ossification program. In summary, hESC-derived sclerotome harbored bone/cartilage progenitor activity in vivo and formed bones in a process mimicking natural bone development.

Human sclerotome could also generate cartilage and fibroblasts in vitro (Figure 1A, step 6). Exposure to BMP (Murtaugh et al., 1999) upregulated cartilage structural genes, yielding a fairly uniform COL2A1⁺ population that later secreted glycosaminoglycans (Figures 4Gi and S5E). By contrast, PDGF and TGF β treatment (Cheung et al., 2012) for 3 days yielded a >90% pure SMA α^{hi} fibroblast-like population (Figures 4Gii and S5E). Having defined the signaling logic for PS differentiation into human paraxial mesoderm and somitic derivatives, next we focused on the parallel lineage branch: PS differentiation into lateral mesoderm and cardiac fates (Figure 1A, step 7).

Lateral Mesoderm Patterning into Heart versus Limbs by Competing FGF and WNT Signals

We sought to define how human lateral mesoderm is diversified into multiple derivatives, including NKX2.5⁺ heart-forming anterior-lateral mesoderm and limb-forming PRRX1⁺ posterior-lateral mesoderm (Figure 1A, step 7, and Figure 5A) (Tanaka, 2013).

Bifurcation of day 2 lateral mesoderm into day 3 anterior (cardiac) versus posterior (limb bud) fates was respectively induced by FGF and WNT signals (summarized in Figure 5A). WNT posteriorized lateral mesoderm, inducing limb markers PRRX1 and HOXB5 while suppressing heart field markers NKX2.5 and TBX20 on day 2–3 of hESC differentiation (Figures 5B and S6A). Reciprocally, WNT inhibition suppressed posterior-lateral mesoderm and instead induced cardiac mesoderm (Figures 5B and S6A). Thus, our findings corroborate the clear requirement for WNT blockade for cardiac specification in vivo (Schneider and Mercola, 2001) and in vitro (Burridge et al., 2014; Lian et al., 2012; Mendjan et al., 2014) and further indicate that WNT inhibitors induce heart precursors by restraining limb formation (Figure S6D).

Conversely, FGF anteriorized human lateral mesoderm. Exogenous FGF enhanced NKX2.5-GFP⁺ cardiac mesoderm induction on day 2–3 of hESC differentiation, whereas FGF inhibition abolished cardiac induction (Figure 5C). This demonstrates a conserved role for FGF in cardiac mesoderm specification from humans to zebrafish (Reifers et al., 2000).

Hence, activating pro-cardiac FGF signaling and inhibiting pro-limb WNT signaling (in the permissive context of BMP activation and TGF β inhibition) efficiently directed day 2 lateral mesoderm toward cardiac mesoderm, as tracked using an NKX2.5-GFP knockin hESC line (Elliott et al., 2011). We respectively obtained >80% and >90% NKX2.5-GFP⁺ cardiac mesoderm by days 3 and 4 of hPSC differentiation (Figure 5D), which was more rapid and robust than a WNT-modulator-only cardiac induction protocol (Figure 5E) (Burridge et al., 2014).

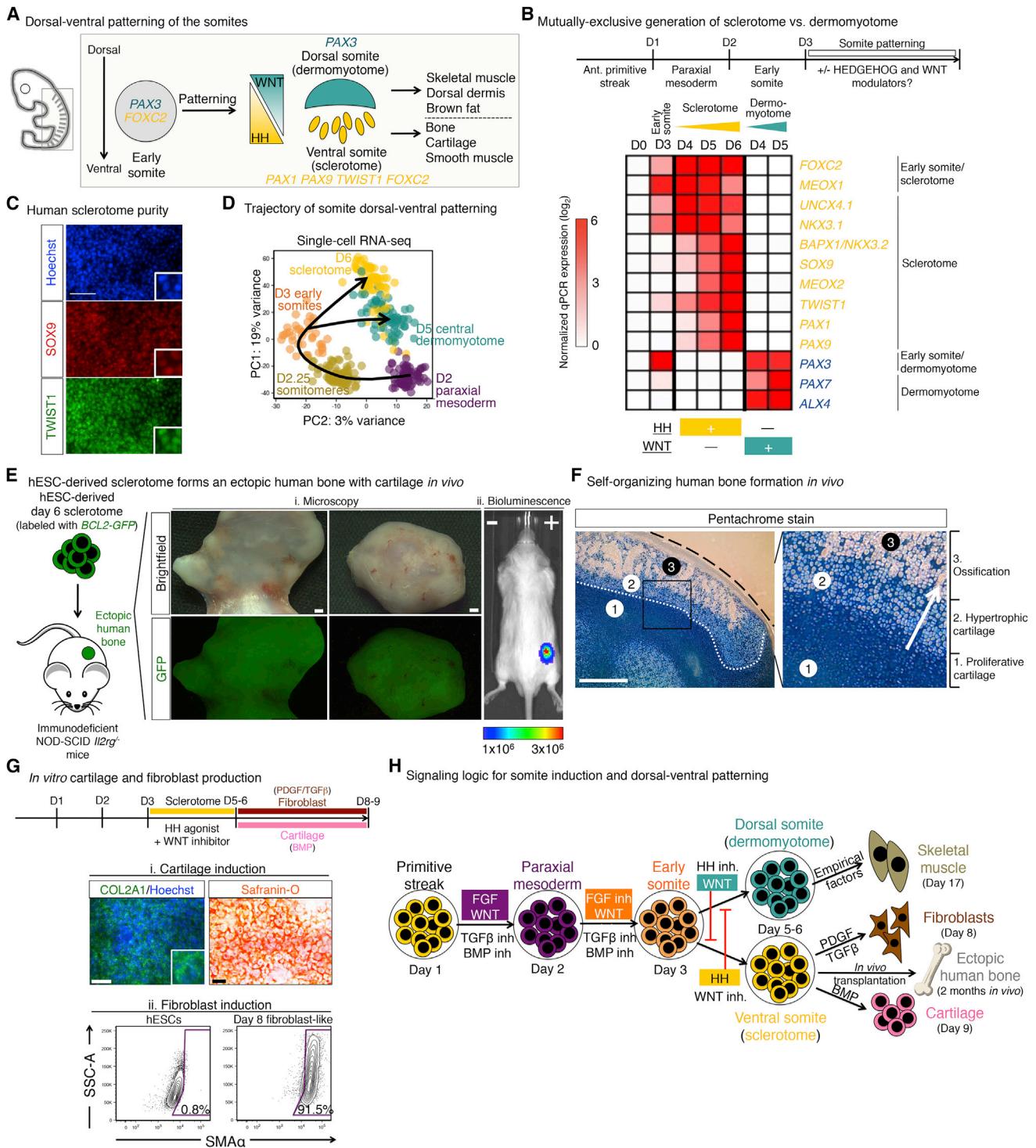


Figure 4. Dorsal-Ventral Patterning of Somite Precursors into Sclerotome and Dermomyotome and Downstream Progeny

(A) Somite patterning *in vivo*.

(B) qPCR heatmap of hESCs (day 0), early somite progenitors (day 3), or those differentiated into sclerotome (day 4, day 5, or day 6, using 21K+C59) or dermomyotome (day 4 or day 5, using BMP4+CHIR+Vismodegib).

(C) SOX9 and TWIST1 staining of day 6 H7-derived sclerotome; scale bar, 100 μm.

(D) PCA of scRNA-seq from indicated populations; each dot depicts a single cell.

(E) EF1A-BCL2-2A-GFP expressing H9-derived sclerotome was subcutaneously injected into NSG mice; 2 months later, ectopic GFP⁺ human bones formed (left); bioluminescent imaging of mice 1 month post-transplantation by UBC-Luciferase-2A-tdTomato H9-derived sclerotome.

(legend continued on next page)

Directing Cardiac Mesoderm into Human Cardiomyocytes that Can Engraft Human Fetal Heart *In Vivo*

Having rapidly generated a >90% pure NKX2.5⁺ cardiac mesoderm population by day 4 of hESC differentiation (Figures 5D and 5E), we next sought to drive these progenitors toward cardiomyocytes (Figure 1A, step 8). Single cardiac mesoderm progenitors can form both myocardium (cardiomyocytes) and endocardium in mice (Devine et al., 2014).

BMP activation together with low FGF levels preferentially induced cardiomyocytes from cardiac mesoderm (Figures S6E and S6F) at the expense of (pro)epicardium or endocardium (Figure S6E). WNT activation seemed to sustain undifferentiated *ISL1*⁺ cardiac progenitors and inhibited maturation into cardiomyocytes (Figure S6E). Therefore, WNT blockade in conjunction with BMP activation enhanced cardiomyocyte differentiation (Figure S6H), which was enhanced by vitamin C (Figure S6I) (Burridge et al., 2014). Treating day 4 cardiac mesoderm with these factors yielded a 72.2% ± 5.6% and a 77.8% ± 1.6% pure TROPONIN⁺ cardiomyocyte population by days 6 and 8 of hPSC differentiation, respectively (Figure 5F), which spontaneously contracted and expressed cardiomyocyte structural genes (Figure S6J).

The authenticity of hESC-derived cardiac lineages was confirmed by their ability to engraft human fetal heart tissue. hESC-derived cardiomyocytes can engraft model organisms (e.g., guinea pigs and non-human primates; Chong et al., 2014) but therapies will require evidence that such cells can engraft *human* heart tissue. To this end, we employed an experimental system whereby ventricular fragments from week 15–17 human fetal heart (Ardehali et al., 2013) were subcutaneously implanted into the mouse ear (Figure 5G). These human heart fragments were revascularized and continued beating for months *in vivo*, as shown by QRS electrocardiogram signals (Figure 5G). Upon transplantation of luciferase⁺/GFP⁺ hESC-derived heart populations, both day 3 cardiac mesoderm and day 8 cardiomyocytes engrafted human ventricular fragments for at least 10 weeks, as indicated by bioluminescence imaging ($n = 10$ successfully engrafted human heart fragments obtained from 2 fetal donors; Figure 5Hi). Within human fetal heart tissue, GFP⁺ hESC-derived cardiomyocytes expressed cytoskeletal protein TROPONIN/TNNT2 and membranous gap junction protein CONNEXIN 43 (Figure 5Hii). Together, these data show that our differentiation strategy can lead to the rapid generation of a ~75% pure cardiomyocyte population capable of engrafting human fetal heart tissue.

Identifying Cell-Surface Markers to Allow Purification of Diverging Human Mesoderm Subtypes

After generating this hierarchy of mesoderm lineages, we also defined lineage-specific cell-surface markers to track different

mesoderm lineages and enable purification of each mesoderm lineage for assessment of biological function and fate and for potential therapeutic purposes in the future. Screening 332 cell-surface markers across hESCs and 7 mesodermal lineages (Figure 6A; Table S2) revealed that certain previously described mesoderm markers were broadly expressed in both paraxial and cardiac mesoderm (Figures S7A and S7B). Therefore, we sought lineage-specific markers.

Surface markers DLL1 and GARP respectively marked paraxial mesoderm and cardiac mesoderm in a mutually exclusive fashion. NOTCH ligand DLL1 was specifically expressed in day 2 paraxial mesoderm, whereas conversely, GARP marked day 3 NKX2.5-GFP⁺ cardiac mesoderm. Neither marker was expressed by undifferentiated hESCs (Figures 6B and S7C), thus tracking a clear bifurcation of PS into paraxial or cardiac mesoderm fates.

DLL1 and GARP were mesodermal markers conserved from human to zebrafish. GARP/LRRC32, a transmembrane protein that tethers TGF β ligands to the cell surface (Tran et al., 2009), was likewise expressed in the heart tube of zebrafish embryos (Figure 6C). Conversely, *deltaC* (a homolog to human DLL1) was likewise expressed in zebrafish paraxial mesoderm (Figure S7D), as reported previously (Smithers et al., 2000).

These surface markers enabled the purification of desired mesoderm lineages from admixed cultures, providing tools to refine stem-cell differentiation. DLL1 was expressed by 91.6% ± 5.4% of cells after 2 days of paraxial mesoderm induction. Paraxial mesoderm-specific TFs (TBX6, MSGN1) were exclusively expressed by the DLL1⁺ fraction (Figure 6D). Sorted DLL1⁺ human paraxial mesoderm cells were essentially uniform as shown by scRNA-seq: 97.0% of cells coexpressed archetypic paraxial markers MSGN1 and DLL3 (Figure 6E). TBX6 mRNA expression in all but a few cells (Figure 6E) may reflect technical dropout in scRNA-seq (Marinov et al., 2014). Hence, sorting for DLL1⁺GARP⁻ cells purifies human paraxial mesoderm attained from either differentiating ESCs or iPSCs (induced pluripotent stem cells; Figure S7E), providing a method to isolate pure human paraxial mesoderm and interrogate its characteristics.

Downstream of paraxial mesoderm during the bifurcation of sclerotome versus dermomyotome fates, surface marker PDGFR α enabled the purification of sclerotome. PDGFR α was expressed by 85.2% ± 8.4% of cells in day 5–6 sclerotome cultures, and only the PDGFR α ⁺ fraction expressed sclerotome markers (FOXC2, PAX1, and PAX9), consistent with *pdgfra* expression in the sclerotome, but not dermomyotome of zebrafish embryos (Figure 6F) (Liu et al., 2002). In vitro, PDGFR α was indeed expressed at higher levels in hPSC-derived sclerotome relative to dermomyotome (Figure S7E), thereby helping to distinguish ventral from dorsal somite fates.

scRNA-seq of PDGFR α ⁺ human sclerotome revealed 86.2% of cells coexpressed chondroprogenitor markers SOX9 and

(F) Russell-Movat's Pentachrome staining of 2-month-old sclerotome grafts revealed zones of chondrogenesis and ossification, with cartilage stained blue; black line denotes the edge of the graft; white line denotes boundary of the ossifying region; scale bar, 1 mm (bottom).

(G) COL2A1 (top left) and Safranin-O staining (top right) of day 6+2 or day 6+6 hESC-derived cartilage, respectively; scale bars, 0.1 mm (left) and 1 mm (right); SMA α intracellular FACS of hESCs or day 8 fibroblast-like cells (bottom).

(H) Somite patterning into dermomyotome or sclerotome and downstream differentiation.

See also Figures S4 and S5.

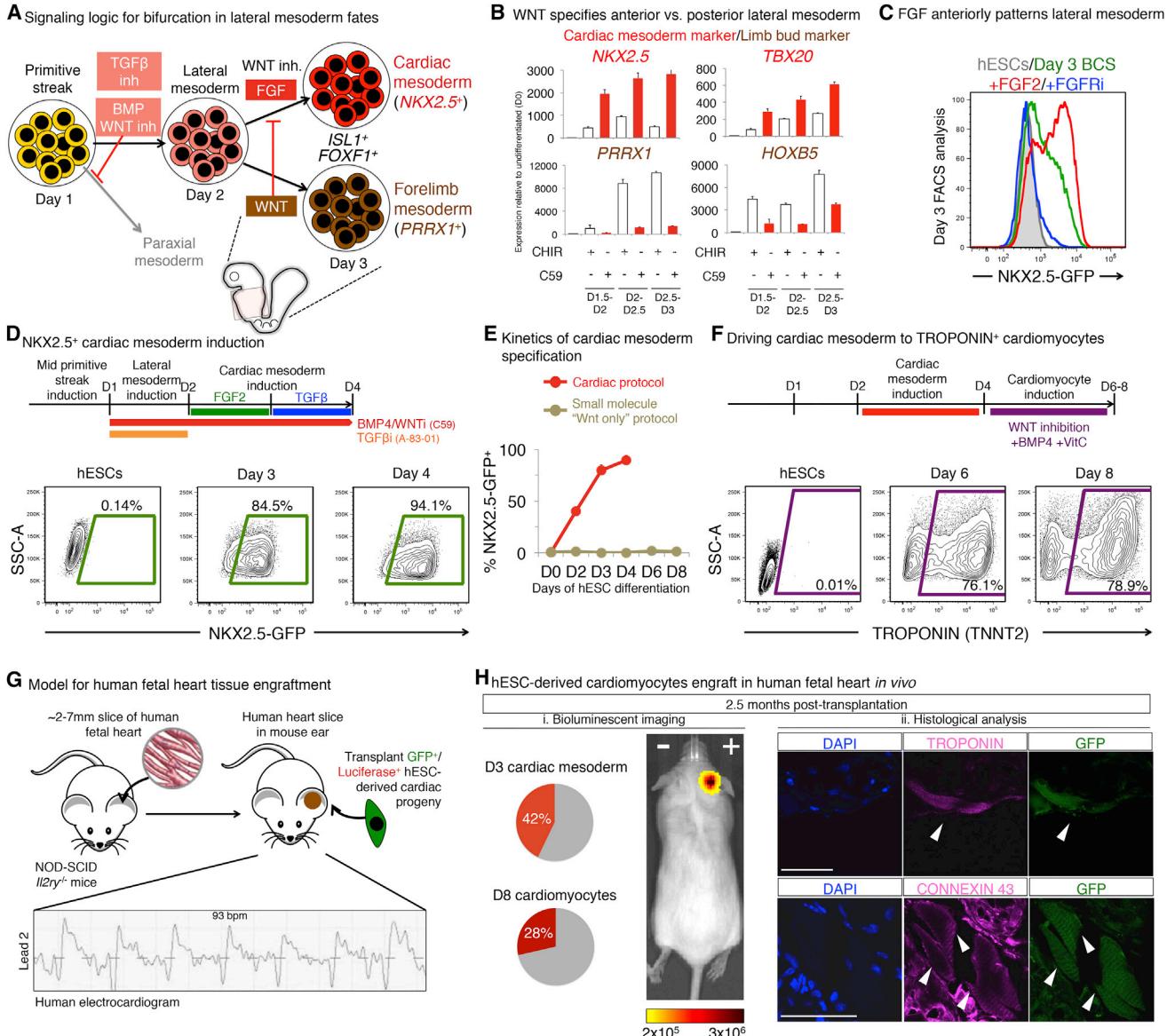


Figure 5. Lateral Mesoderm Patterning into Cardiac versus Limb Mesoderm Fates

- (A) Cardiac versus forelimb bifurcation.
- (B) To assess the role of WNT in lateral mesoderm patterning, day 1 PS was differentiated to lateral mesoderm (30 ng/mL BMP4 + 1 μ M C59 + 2 μ M SB505124 [BS]) for varying lengths of time (until day 2, day 2.5, or day 3) and, for the last 12 hr, was treated with C59 or 3 μ M CHIR (in addition to BS), and qPCR was conducted.
- (C) To assess the role of FGF in lateral mesoderm patterning, day 2 *NKX2.5-GFP* lateral mesoderm was treated with BMP4 + C59 + SB505124 (BCS) with or without FGF2 (20 ng/mL) or FGFR inhibitor PD173074 (100 nM) for 24 hr, and FACS was conducted on day 3.
- (D) Time-point FACS of *NKX2.5-GFP* hESC (Elliott et al., 2011) differentiation using cardiac mesoderm protocol.
- (E) Comparison of *NKX2.5-GFP*⁺ cell percentages (determined by FACS) on days of differentiation, using the current protocol or a previous method (Burridge et al., 2014).
- (F) Intracellular TNNT2 FACS of H7-derived cardiomyocytes (bottom).
- (G) Electrocardiogram of human fetal heart implanted in the mouse ear, >1 month post-implantation.
- (H) 2.5 months post-transplant of *EF1A-BCL2-2A-GFP;UBC-tdTomato-Luciferase* H9 hESC-derived cardiac lineages into human fetal heart grafts, luciferase⁺ donor cells were detected (i); engrafted hESC-derived cardiomyocytes were TROPONIN⁺ and CONNEXIN 43⁺; scale bar, 40 μ m (ii).
- See also Figure S6.

COL2A1 (Figure 6G), reaffirming the skeletal stem-cell-like nature of these cells and consistent with their phenotypic ability to form ectopic bone grafts (see above). However, *PAX9* and

FOXC2 expression was more heterogeneous (Figure 6G), which may reflect distinct anterior-posterior and medial-lateral sclerotome subdomains (Christ and Scaal, 2008).

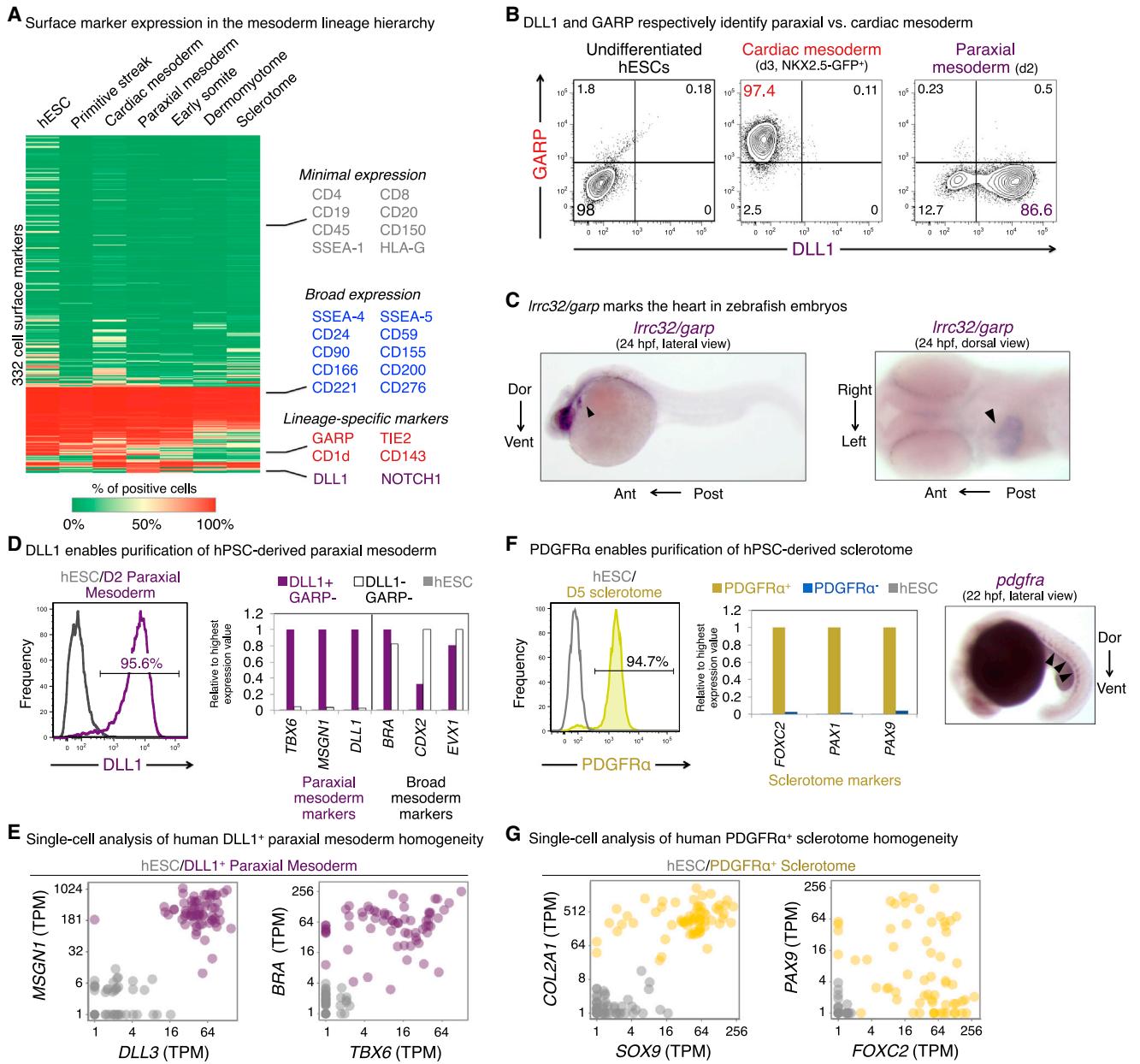


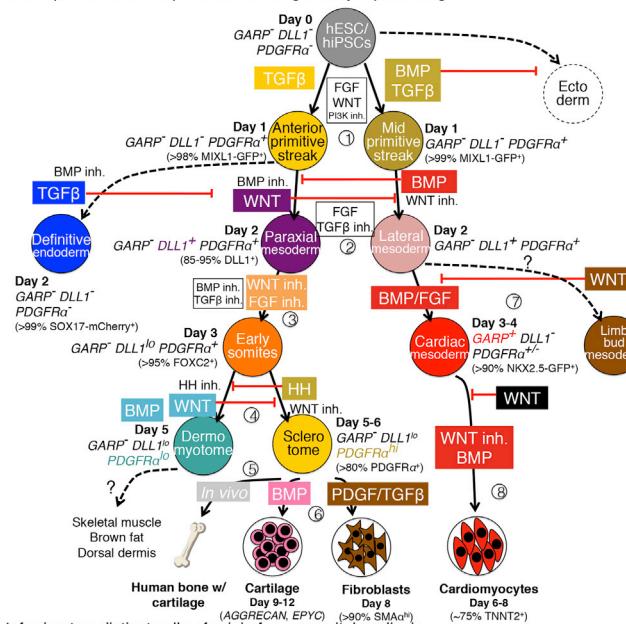
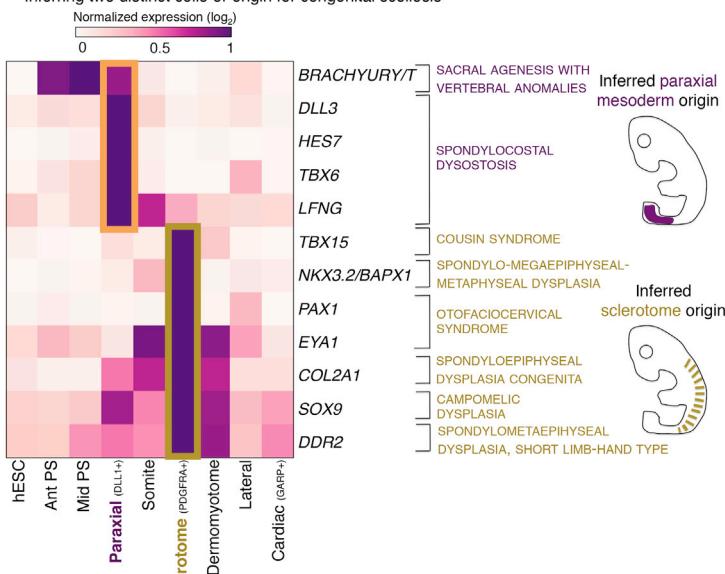
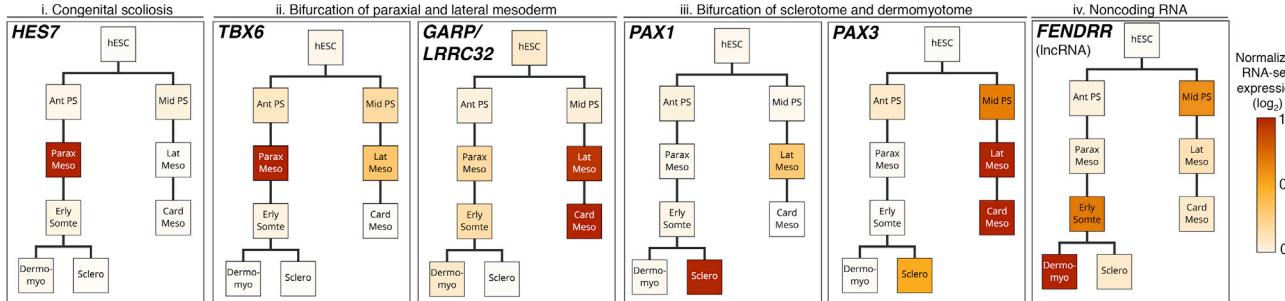
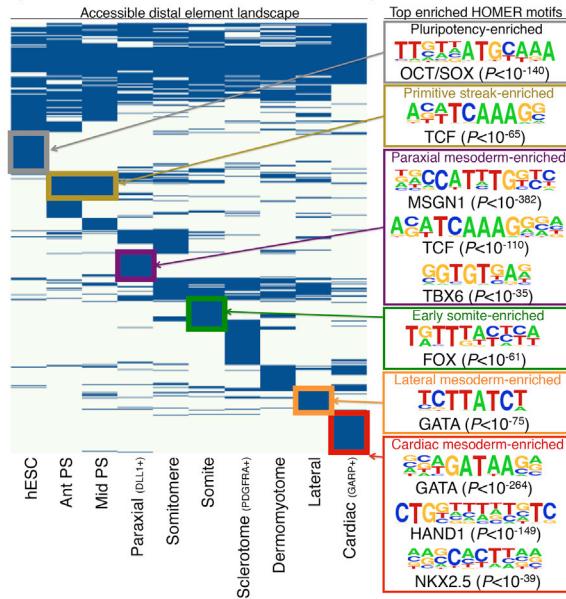
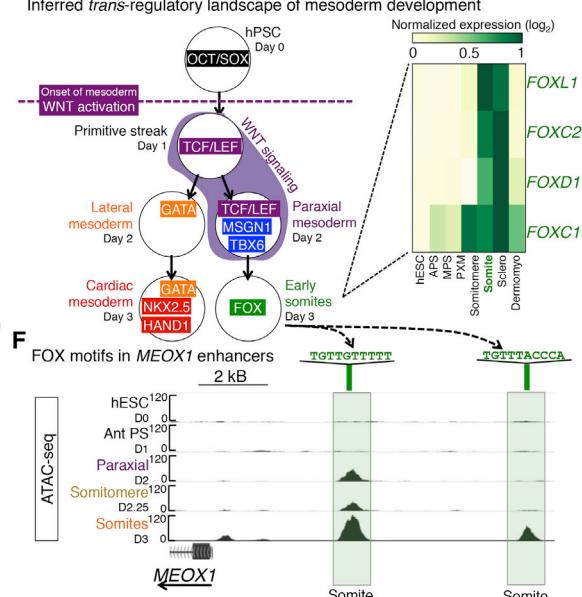
Figure 6. High-Throughput Screen for Lineage-Specific Mesoderm Surface Markers

- (A) Clustered heatmap of surface marker expression in hESCs and 6 mesoderm derivatives. Each row represents an individual surface marker and color denotes the percentage of cells positive for a given marker. For PS and cardiac mesoderm, marker expression was analyzed after pre-gating on MIXL1-GFP⁺ and NKX2.5-GFP⁺ fractions, respectively.
- (B) GARP and DLL1 FACS in hESCs, day 2 paraxial mesoderm cultures or day 3 NKX2.5-GFP⁺ pre-gated cardiac mesoderm.
- (C) *Irrc32* in situ hybridization in 24 hr post-fertilization zebrafish embryo (arrows denote heart).
- (D) DLL1 FACS of day 2 paraxial mesoderm culture (left); qPCR of sorted populations (right).
- (E) scRNA-seq of sorted DLL1⁺ human paraxial mesoderm; each dot is a single cell.
- (F) PDGFR α FACS of day 5 sclerotome population (left); qPCR of sorted PDGFR α ⁺ and PDGFR α ⁻ populations (center); in situ hybridization for *pdgfra* expression (right) in 22 hpf zebrafish embryo (arrowheads denote ventral staining in sclerotome).
- (G) scRNA-seq of sorted PDGFR α ⁺ human sclerotome; each dot is a single cell.

See also Figure S7 and Table S2.

In summary, these cell-surface markers define a roadmap for mesoderm development (Figure 7A) by identifying mutually exclusive types of mesoderm progenitors, thus enabling one to

track the products of key developmental branchpoints. These markers are likewise expressed by the same cell types in zebrafish embryos (Figures 6 and S7). These markers allowed us to

A Developmental roadmap for mesoderm germ layer patterning**B** Inferring two distinct cells-of-origin for congenital scoliosis**C** RNA-seq analysis of mesoderm lineage bifurcations**D** Open chromatin atlas of human mesoderm development**E** Inferred trans-regulatory landscape of mesoderm development

confirm that hPSC differentiation to various mesoderm lineages was efficient (Figure 6) and enabled further purification of desired lineages for transcriptional and chromatin analysis.

Global Transcriptional Profiling of the Mesoderm Lineage Hierarchy Confirms Lineage Demarcations and Suggests a Cell-of-Origin for Congenital Malformations

To chart a molecular roadmap for mesoderm development, we used bulk-population RNA-seq to capture global transcriptional dynamics during the commitment of ESCs to nine diverging human mesodermal fates spanning multiple developmental stages (Table S3). This mesoderm gene expression atlas provided insight into the potential cell-of-origin of various human congenital malformations.

Congenital scoliosis is a genetically heterogeneous disease that has been mapped to diverse genes in human patients, in which it causes malformations in the spine, scapulae, and/or ribs (reviewed by Giampietro, 2012; Pourquié, 2011). Due to the inaccessibility of early human embryos, it has been difficult to assess when and where scoliosis susceptibility genes are expressed during development to uncover the origins of this disease.

RNA-seq analysis implicated at least two independent cells-of-origin for different subtypes of congenital scoliosis. For spondylocostal dysostosis (mapped to *DLL3*, *HES7*, *TBX6*, and *LFNG*) and sacral agenesis with vertebral anomalies (*BRACHYURY*), their causative genes were largely expressed in PSC-derived human paraxial mesoderm, but not other mesodermal cell types (Figures 7B and 7Ci). This implies a paraxial mesoderm cell-of-origin for these two types of scoliosis. By contrast, for six other types of congenital scoliosis, their causative genes were strongly expressed in human sclerotome, but not paraxial mesoderm (Figure 7B). Hence, congenital scoliosis may have at least two independent cells-of-origin (paraxial mesoderm or sclerotome) depending on the specific genetic lesion.

RNA-seq analyses of cells diverging across multiple lineage branchpoints also provided a clear view of how fates segregate across consecutive developmental steps. After PS formation, there was a clear partitioning of gene expression patterns in paraxial mesoderm (*TBX6*) and lateral/cardiac mesoderm (*GARP/LRRC32*) (Figure 7Cii). Downstream of paraxial mesoderm, *PAX1* and *PAX3* (Fan and Tessier-Lavigne, 1994) were respectively restricted to either sclerotome or dermomyotome (Figure 7Ciii). These lineage-specific expression patterns demarcate differences in developmental fate and lend confidence to our transcriptional dataset.

The human mesoderm gene expression atlas also uncovered lineage-specific long noncoding RNAs (lncRNAs), nominating

them for further study of lncRNA function. By way of example, *Fendrr* is an lncRNA critical for mouse heart field development (Grote et al., 2013), and human *FENDRR* was likewise expressed in hESC-derived cardiac mesoderm (Figure 7Biv).

Mesodermal Distal Regulatory Elements Reflect the Impact of Dynamic Signaling and trans-Regulatory Influences on Chromatin

To track how chromatin is dynamically remodeled during development of hESCs into nine distinct types of mesodermal progeny, we charted open chromatin using ATAC-seq to identify putative regulatory elements. We also inferred active TFs for each mesodermal lineage by cross-referencing TFs that were expressed according to RNA-seq with those whose motifs enriched in lineage-specific open chromatin regions (Supplemental Experimental Procedures).

Though accessible chromatin in pluripotent cells was enriched for OCT/SOX-binding motifs, upon 24 hr of differentiation, the PS chromatin landscape became dominated by motifs of TCF/LEF TFs (the effectors of WNT signaling; Figure 7D), reflecting how WNT drives PS induction (Figure 1B). As differentiation progressed, the *trans*-regulatory landscape of day 2 paraxial mesoderm seemed to be built on that of day 1 PS, consistent with how both lineages experience WNT activation (Figure 2): TCF/LEF continued to engage paraxial mesoderm chromatin but was apparently joined by paraxial mesoderm-specific TFs *MSGN1* ($p < 10^{-382}$) and *TBX6* ($p < 10^{-35}$) (Figure 7D). However, within 24 hr of WNT inhibition and the segmentation of paraxial mesoderm into somites, the landscape transitioned from a TCF/LEF-dominated state to one significantly enriched for FOX motifs ($p < 10^{-61}$; Figure 7D). This was evidenced by predicted FOX motifs in two upstream *MEOX1* enhancers that were accessible in somites (Figure 7F). By virtue of RNA-seq expression patterns (Figure 7E), multiple FOX TFs could account for the FOX-driven somite regulatory state, including *FOXC2*, which was indeed expressed at the protein level in somites (Figure 2).

Along the alternate lineage pathway, the 24 hr progression from PS to lateral mesoderm involved a transition from a TCF-driven chromatin landscape to a GATA-dominated one (Figure 7D). This reflects the importance of WNT repression in lateral mesoderm specification (Figure 1E) and expression of multiple GATA TFs in lateral mesoderm (Figure S7K). Upon 24 hr of further differentiation into *GARP⁺* cardiac mesoderm, GATA motifs became accompanied by *HAND1* ($p < 10^{-149}$) and *NKX2.5* ($p < 10^{-39}$) (Figure 7D). These findings provide insight into control of mesoderm development: there is no monolithic “panmesodermal” program, but instead, chromatin is substantially

Figure 7. The Landscape of Mesoderm Development

- (A) Lineage steps with circled numbers correspond to respective sections in the main text and Figure 1A.
 - (B) RNA-seq expression of human congenital scoliosis genes.
 - (C) RNA-seq profiling; color intensity depicts gene expression ($\log_2 \text{TPM}$) normalized to the expression of that gene in all populations profiled, with the highest-expressing lineage assigned the most intense color value.
 - (D) ATAC-seq heatmap. Each horizontal line depicts a single chromatin element (left, non-binariized in Figure S7J), with motifs representative of the top four lineage-enriched motifs shown (right).
 - (E) Inferred *trans*-regulatory lineages programs (left); heatmap of the four FOX TFs most highly expressed in hESC-derived somites (RNA-seq; right).
 - (F) ATAC-seq of the *MEOX1* locus, with FOX motifs centered in two somitic enhancer elements shown.
- See also Figure S7.

remodeled every 24 hr even as closely related mesodermal lineages segue into one another (Figure 7D). Furthermore, the distal element landscape reflects how changes in signaling influences and *trans*-acting regulatory states become physically imprinted on chromatin. Together, this sketches a model for how regulatory states change during mesoderm subtype diversification (Figure 7E).

DISCUSSION

A Roadmap for Human Mesoderm Development

In vitro stem-cell differentiation often yields admixed lineages, possibly due to incomplete suppression of alternate fates or passage through incorrect lineage intermediates. To systematically block production of unwanted lineages in preference to desired fates during stem-cell differentiation, we must map the underlying developmental landscape.

To this end, here we chart a roadmap for human mesoderm development and describe how 12 different human cell types including bone, muscle, and heart emerge from pluripotent cells (Figure 7A). We used scRNA-seq to systematically catalog the diversity of intermediate cell states formed during differentiation, and we tested the minimal combinations of *positive* and *negative* signals that were sufficient for differentiation between each of these intermediate states. Though vertebrate mesoderm development was broadly outlined by pioneering analyses in model organisms that identified certain key genes and signaling pathways (reviewed by Kimelman, 2006; Schier and Talbot, 2005; Tam and Loebel, 2007), it has been difficult to precisely map mesoderm formation due to the large number of mesodermal subtypes and the finely graded, temporally dynamic transitions between them. Throughout consecutive pairwise lineage branches in human mesoderm development, we clearly defined (1) the diverging cell states through scRNA-seq, (2) positive and negative signals inducing each of the mutually exclusive lineages, (3) specific cell-surface markers that identified key mesoderm intermediates, and (4) the chromatin landscapes of the diverging fates (Figure 7A). Besides providing a broad reference map for developmental biology and regenerative medicine, we directly demonstrate the applications of this roadmap to produce engraftable human tissue progenitors and provide insight into developmental signaling dynamics, chromatin remodeling, and congenital disease.

Extrinsic Signals: Logically Blocking Alternative Lineage Formation to Guide Stem-Cell Differentiation

Vertebrate embryology has identified certain signals required for mesoderm formation in model organisms (reviewed by Kimelman, 2006; Schier and Talbot, 2005; Tam and Loebel, 2007), and here, we have tested whether we understand mesoderm development at the level of causation by reconstituting aspects of this process from cultured stem cells. At each lineage transition from human pluripotency to terminally differentiated mesoderm fates, we could identify and test the minimal signaling conditions needed to induce each lineage. The resultant understanding of the underlying signaling logic guided the rapid differentiation of hPSCs into desired mesoderm intermediates (>98% pure MIXL1⁺ primitive streak; >90% pure NKX2.5⁺ car-

diac mesoderm; >90% pure DLL1⁺ paraxial mesoderm; >95% pure FOXC2⁺ early somite progenitors) within several days of differentiation in serum-free, monolayer conditions (without recourse to gene modification). Such efficient, rapid induction relied on the following two principles.

First, the principal finding of this work is that, at each lineage bifurcation, stem cells could be exclusively differentiated down a single lineage path by providing the *positive* signal(s) to induce a given fate while repressing *inhibitory* signal(s) that instead induced the alternate fate. Blocking formation of undesired fates was imperative for efficient differentiation. By way of example, efficient differentiation of day 1 PS into day 2 paraxial mesoderm required WNT activation (to specify paraxial mesoderm) together with inhibition of BMP and TGF β pathways (to block lateral mesoderm and endoderm formation, respectively) in order to block differentiation toward unwanted fates and to consolidate differentiation down only a single path. Therefore, hPSC differentiation to a desired lineage cannot solely rely on knowledge of the requisite inductive signal(s), but also on an understanding of signals that induce mutually exclusive fates at each step of the way. This underscores the need for systematic developmental roadmaps.

Second, another highlight was the rapidity with which developmental signals were re-interpreted during hPSC differentiation and the consequent importance of controlling temporal signaling dynamics. In the gastrulating mouse embryo, lineage transitions occur every 12–24 hr, for example: E5.5 epiblast → E6.5 primitive streak → E7–7.5 paraxial mesoderm → E8 early somites → E8.5 sclerotome. In vitro, we found that WNT activation on day 0–1 drove hPSC toward PS; WNT activation on day 1–2 then specified paraxial mesoderm; WNT inhibition on day 2–3 differentiated paraxial mesoderm into early somites; and finally, WNT activation on day 3–4 specified dermomyotome. Thus, over the course of 4 days in vitro, WNT was interpreted four different ways as lineages segued into one another every 24 hr. By contrast, some differentiation methods continuously provide the same signal for days or weeks, potentially explaining why a mélange of lineages is produced. Our system therefore constitutes a venue to understand how extrinsic signals are dynamically interpreted in the context of changing windows of developmental competence.

Single-Cell RNA-Seq: Cataloging Mesodermal Lineages and Transition States in between Them

Mapping a developmental hierarchy hinges on cataloging its constituent progenitor states, which we have done here for the human mesodermal lineage hierarchy. We proposed that scRNA-seq sampling (Table S4) would be a complete method to test the lineage and homogeneity of cells at each developmental step. Indeed, early human primitive streak, lateral mesoderm, and paraxial mesoderm lineages were highly uniform. Starting from human paraxial mesoderm, cells initiated somitogenesis along a single continuous trajectory (Figure 2G), and snapshots of this process uncovered the formation of a transient HOPX⁺ human somitomere intermediate that arose for several hours during differentiation. This argues that human somite development entails passage through an ephemeral segmentation-like state (as described for other vertebrate model

organisms; [Pourquié, 2011](#)), which has been impossible to assess *in vivo* due to its transient nature and the unavailability of early human embryos.

Navigating Mesoderm Development

Identifying lineage-specific cell-surface markers for major mesoderm subtypes (e.g., DLL1 for paraxial mesoderm and GARP for cardiac mesoderm) enables the purification of desired mesoderm subtypes to investigate the biological characteristics of these cells (as embodied by our RNA-seq and ATAC-seq analyses) or for therapeutic transplantation in the future. Moreover, the major surface markers defined here for human mesoderm progenitors were correspondingly expressed in zebrafish embryos, indicating that they are conserved developmental markers.

Collectively, we delineate a clear lineage hierarchy for mesoderm development with prospectively isolatable lineage intermediates at each step, which should be key for understanding human mesoderm development as well as the clinical purification of hPSC-derived tissue stem and progenitor cells for regenerative medicine in the future ([Figure 7A](#)). The ability to produce highly homogeneous populations of human mesodermal progenitors now opens the door to rapid generation and purification of a wealth of different mesodermal cell types from hPSCs—including the engraftable sclerotome and cardiomyocyte populations described here—providing a future foundation for regenerative medicine. Artificially reconstituting aspects of mesoderm development from hPSCs should provide a facile system to study basic developmental processes *in vitro*, including how developmental signals are temporally re-interpreted and combinatorially integrated and how chromatin dynamics are linked to changing windows of developmental competence.

Yet the roadmap remains incomplete. It does not include lineage paths to human axial mesoderm, intermediate mesoderm, or mediolateral derivatives of sclerotome and dermomyotome ([Christ and Scaal, 2008](#)). Finally, though we have identified extracellular signals that specify human mesoderm cell fate *in vitro*, to accompany the jump of complexity from 2D culture dish to 3D embryo we must in turn map the niche cells that produce these signals during embryogenesis and where they are located, thus unraveling differentiation in 3D space.

EXPERIMENTAL PROCEDURES

Mesoderm Differentiation

Monolayer, feeder-free differentiation was conducted in serum-free CDM2 basal medium ([Supplemental Experimental Procedures](#)). hPSCs (mainly H7) were passaged ~1:12–1:20 as fine clumps (using Accutase) onto Geltrex-coated wells and cultured overnight in mTeSR1 + 1 μM thiazovivin. The next morning, hPSCs were differentiated toward anterior PS (30 ng/mL Activin + 4 μM CHIR + 20 ng/mL FGF2 + 100 nM PIK90; for downstream paraxial differentiation) or mid PS (30 ng/mL Activin + 40 ng/mL BMP4 + 6 μM CHIR + 20 ng/mL FGF2 + 100 nM; for downstream lateral differentiation) for 24 hr. Day 1 anterior PS were differentiated toward day 2 paraxial mesoderm (1 μM A8301 + 3 μM CHIR + 250 nM LDN193189 [DM3189] + 20 ng/mL FGF2) for 24 hr. Day 1 mid PS were differentiated toward day 2 lateral mesoderm (1 μM A8301 + 30 ng/mL BMP4 + 1 μM C59) for 24 hr.

Day 2 paraxial mesoderm were differentiated toward day 3 early somites (1 μM A8301 + 1 μM C59 [or alternately, 1 μM XAV939] + 250 nM

LDN193189 + 500 nM PD0325901; 24 hr). Day 3 early somites were differentiated toward either day 5–6 sclerotome (5 nM 21K + 1 μM C59) or day 5 dermomyotome (3 μM CHIR + 150 nM Vismodegib, optionally with 50 ng/mL BMP4) for 48–72 hr. Day 5 sclerotome were differentiated toward day 8 fibroblast-like cells (10 ng/mL TGFβ1 + 2 ng/mL PDGF-BB; [Cheung et al., 2012](#)) for 72 hr. Day 6 sclerotome were differentiated toward day 9–day 12 cartilage (20 ng/mL BMP4) for 3–6 days.

Ectopic Human Bone Formation

$\sim 1.5 \times 10^7$ day 6 sclerotome cells were subcutaneously transplanted in 1:1 CDM2/Matrigel mixture into NOD-SCID *Il2rg*^{-/-} mice, which were sacrificed in ~2–3 months.

Human Fetal Heart Graft Construction

2 × 7 mm strips of week 15–17 human fetal heart were subcutaneously implanted into the ear of NOD-SCID *Il2rg*^{-/-} mice using a trocar. 1 month later, $1.5\text{--}2 \times 10^6$ day 3 cardiac mesoderm or day 8 cardiomyocytes were directly injected into the fetal heart tissue (in 1:1 CDM2/Matrigel mixture) and analyzed ~2.5 months later.

RNA-Seq

RNA was purified from H7-derived mesoderm lineages, either from bulk populations or from 651 single cells spanning ten lineages (Fluidigm C1 system). RNA-seq libraries were prepared (bulk; Ovation RNA-seq System V2 and NEBNext Ultra DNA Library Prep Kit and single cell; SMARTer Ultra Low RNA Kit) and sequenced (Next-Seq 500) to obtain 150 bp paired-end reads, which were processed using the ENCODE long RNA analysis pipeline ([Supplemental Experimental Procedures](#)). Collated data is viewable at http://cs.stanford.edu/~zhenghao/mesoderm_gene_atlas.

ACCESSION NUMBER

The accession number for the RNA-seq and ATAC-seq data reported in this paper is SRA: SRP073808.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, and seven tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2016.06.011>.

A video abstract is available at <http://dx.doi.org/10.1016/j.cell.2016.06.011#mmc6>.

AUTHOR CONTRIBUTIONS

K.M.L., A.C., T.Z.D., J.M.T., N.S.-C., N.B.F., B.M.G., R.E.A.S., and L.T.A. differentiated and transplanted hPSC; P.W.K., R.S., A.A.B., R.M.M., Z.C., and A.K. executed RNA-seq and ATAC-seq; K.Y.S. and W.S.T. stained zebrafish; R.J. and J.A.E. conducted Hopx staining; G.W. and H.V. analyzed heart grafts; and L.T.A., K.M.L., A.C., P.A.B., and I.L.W. oversaw the project.

ACKNOWLEDGMENTS

We thank O. Curreri, A. Dainis, B. Brady, L. Stanton, G. Gulati, W. Wenderski, M. Nichane, K.L. Lee, E. Lujan, C. van Neste, N. Neff, and K. Roberts for input. M. Inlay, N. Sever, E.S. Ng, A. Elefantiy, and E. Stanley graciously shared reagents. T. Storm, A. McCarty, P. Lovelace, J. Coller, T. Doyle, and the Stanford Stem Cell Institute FACS, Genomics, Functional Genomics, and In Vivo Imaging Cores provided support. This work was supported by CIRM (RT2-02060, RT3-07683, TB1-01195), NIH (HL125040, GM007365, HL119553, HL071546, HL100405, NS069375, RR029338, OD018220), HHMI, the Siebel Stem Cell Institute, anonymous donors, and A*STAR. K.M.L. was supported by the Hertz Foundation, U.S. National Science Foundation, and Davidson Institute for Talent Development. J.M.T. by the Paul and Daisy Soros Fellowship, R.J. by the Burroughs Wellcome Fund Career Award for Medical Scientists, and A.K. by the Sloan Foundation Research Fellowship.

Received: February 20, 2015

Revised: April 25, 2016

Accepted: June 1, 2016

Published: July 14, 2016

REFERENCES

- Ardehali, R., Ali, S.R., Inlay, M.A., Abilez, O.J., Chen, M.Q., Blaukamp, T.A., Yazawa, M., Gong, Y., Nusse, R., Drukker, M., and Weissman, I.L. (2013). Prospective isolation of human embryonic stem cell-derived cardiovascular progenitors that integrate into human fetal heart tissue. *Proc. Natl. Acad. Sci. USA* **110**, 3405–3410.
- Aulehla, A., Wiegaebe, W., Baubet, V., Wahl, M.B., Deng, C., Taketo, M., Lewandoski, M., and Pourquié, O. (2008). A β -catenin gradient links the clock and wavefront systems in mouse embryo segmentation. *Nat. Cell Biol.* **10**, 186–193.
- Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218.
- Burridge, P.W., Matsa, E., Shukla, P., Lin, Z.C., Churko, J.M., Ebert, A.D., Lan, F., Diecke, S., Huber, B., Mordwinkin, N.M., et al. (2014). Chemically defined generation of human cardiomyocytes. *Nat. Methods* **11**, 855–860.
- Cheung, C., Bernardo, A.S., Trotter, M.W.B., Pedersen, R.A., and Sinha, S. (2012). Generation of human vascular smooth muscle subtypes provides insight into embryological origin-dependent disease susceptibility. *Nat. Biotechnol.* **30**, 165–173.
- Chong, J.J.H., Yang, X., Don, C.W., Minami, E., Liu, Y.-W., Weyers, J.J., Mahoney, W.M., Van Biber, B., Cook, S.M., Palpant, N.J., et al. (2014). Human embryonic-stem-cell-derived cardiomyocytes regenerate non-human primate hearts. *Nature* **510**, 273–277.
- Christ, B., and Scaal, M. (2008). Formation and differentiation of avian somite derivatives. In *Somitogenesis*, M. Maroto and N. Whittock, eds. (Austin, Texas: Landes Bioscience), pp. 1–41.
- Davis, R.P., Ng, E.S., Costa, M., Mossman, A.K., Sourris, K., Elefanty, A.G., and Stanley, E.G. (2008). Targeting a GFP reporter gene to the MIXL1 locus of human embryonic stem cells identifies human primitive streak-like cells and enables isolation of primitive hematopoietic precursors. *Blood* **111**, 1876–1884.
- Devine, W.P., Wythe, J.D., George, M., Koshiba-Takeuchi, K., and Bruneau, B.G. (2014). Early patterning and specification of cardiac progenitors in gastrulating mesoderm. *eLife* **3**, e03848.
- Dubrulle, J., McGrew, M.J., and Pourquié, O. (2001). FGF signaling controls somite boundary position and regulates segmentation clock control of spatio-temporal Hox gene activation. *Cell* **106**, 219–232.
- Elliott, D.A., Braam, S.R., Koutsis, K., Ng, E.S., Jenny, R., Lagerqvist, E.L., Biben, C., Hatzistavrou, T., Hirst, C.E., Yu, Q.C., et al. (2011). NKK2-5(eGFP/w) hESCs for isolation of human cardiac progenitors and cardiomyocytes. *Nat. Methods* **8**, 1037–1040.
- Fan, C.M., and Tessier-Lavigne, M. (1994). Patterning of mammalian somites by surface ectoderm and notochord: evidence for sclerotome induction by a hedgehog homolog. *Cell* **79**, 1175–1186.
- 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.
- Fan, C.M., Lee, C.S., and Tessier-Lavigne, M. (1997). A role for WNT proteins in induction of dermomyotome. *Dev. Biol.* **191**, 160–165.
- Gertow, K., Hirst, C.E., Yu, Q.C., Ng, E.S., Pereira, L.A., Davis, R.P., Stanley, E.G., and Elefanty, A.G. (2013). WNT3A promotes hematopoietic or mesenchymal differentiation from hESCs depending on the time of exposure. *Stem Cell Reports* **1**, 53–65.
- Giampietro, P.F. (2012). Genetic aspects of congenital and idiopathic scoliosis. *Scientifica (Cairo)* **2012**, 152365.
- Graf, T., and Enver, T. (2009). Forcing cells to change lineages. *Nature* **462**, 587–594.
- Grote, P., Wittler, L., Hendrix, D., Koch, F., Währisch, S., Beisaw, A., Macura, K., Bläss, G., Kellis, M., Werber, M., and Herrmann, B.G. (2013). The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse. *Dev. Cell* **24**, 206–214.
- Jain, R., Li, D., Gupta, M., Manderfield, L.J., Ifkovits, J.L., Wang, Q., Liu, F., Liu, Y., Poleshko, A., Padmanabhan, A., et al. (2015). HEART DEVELOPMENT. Integration of Bmp and Wnt signaling by Hopx specifies commitment of cardiomyoblasts. *Science* **348**, aaa6071.
- Kimelman, D. (2006). Mesoderm induction: from caps to chips. *Nat. Rev. Genet.* **7**, 360–372.
- Lawson, K.A., Meneses, J.J., and Pedersen, R.A. (1991). Clonal analysis of epiblast fate during germ layer formation in the mouse embryo. *Development* **113**, 891–911.
- Lian, X., Hsiao, C., Wilson, G., Zhu, K., Hazeltine, L.B., Azarin, S.M., Raval, K.K., Zhang, J., Kamp, T.J., and Palecek, S.P. (2012). Robust cardiomyocyte differentiation from human pluripotent stem cells via temporal modulation of canonical Wnt signaling. *Proc. Natl. Acad. Sci. USA* **109**, E1848–E1857.
- Liu, L., Chong, S.-W., Balasubramanyan, N.V., Korzh, V., and Ge, R. (2002). Platelet-derived growth factor receptor alpha (pdgfr-alpha) gene in zebrafish embryonic development. *Mech. Dev.* **116**, 227–230.
- Loh, K.M., Ang, L.T., Zhang, J., Kumar, V., Ang, J., Auyeong, J.Q., Lee, K.L., Choo, S.H., Lim, C.Y.Y., Nichane, M., et al. (2014). Efficient endoderm induction from human pluripotent stem cells by logically directing signals controlling lineage bifurcations. *Cell Stem Cell* **14**, 237–252.
- Marinov, G.K., Williams, B.A., McCue, K., Schroth, G.P., Gertz, J., Myers, R.M., and Wold, B.J. (2014). From single-cell to cell-pool transcriptomes: stochasticity in gene expression and RNA splicing. *Genome Res.* **24**, 496–510.
- Mendjan, S., Mascetti, V.L., Ortmann, D., Ortiz, M., Karjosukarso, D.W., Ng, Y., Moreau, T., and Pedersen, R.A. (2014). NANOG and CDX2 Pattern Distinct Subtypes of Human Mesoderm during Exit from Pluripotency. *Cell Stem Cell* **15**, 310–325.
- Murtaugh, L.C., Chyung, J.H., and Lassar, A.B. (1999). Sonic hedgehog promotes somitic chondrogenesis by altering the cellular response to BMP signaling. *Genes Dev.* **13**, 225–237.
- O'Rahilly, R., and Müller, F. (1987). Developmental stages in human embryos (Baltimore, Maryland: Carnegie Institute of Washington).
- Pourquié, O. (2011). Vertebrate segmentation: from cyclic gene networks to scoliosis. *Cell* **145**, 650–663.
- Reifers, F., Walsh, E.C., Léger, S., Stainier, D.Y., and Brand, M. (2000). Induction and differentiation of the zebrafish heart requires fibroblast growth factor 8 (fgf8/acerebellar). *Development* **127**, 225–235.
- Rosenquist, G.C. (1970). Location and movements of cardiogenic cells in the chick embryo: the heart-forming portion of the primitive streak. *Dev. Biol.* **22**, 461–475.
- Schier, A.F., and Talbot, W.S. (2005). Molecular genetics of axis formation in zebrafish. *Annu. Rev. Genet.* **39**, 561–613.
- Schneider, V.A., and Mercola, M. (2001). Wnt antagonism initiates cardiogenesis in *Xenopus laevis*. *Genes Dev.* **15**, 304–315.
- Smithers, L., Haddon, C., Jiang, Y.J., and Lewis, J. (2000). Sequence and embryonic expression of deltaC in the zebrafish. *Mech. Dev.* **90**, 119–123.
- Später, D., Hansson, E.M., Zangi, L., and Chien, K.R. (2014). How to make a cardiomyocyte. *Development* **141**, 4418–4431.
- 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.
- Tam, P.P.L., and Loebel, D.A.F. (2007). Gene function in mouse embryogenesis: get set for gastrulation. *Nat. Rev. Genet.* **8**, 368–381.
- Tanaka, M. (2013). Molecular and evolutionary basis of limb field specification and limb initiation. *Dev. Growth Differ.* **55**, 149–163.

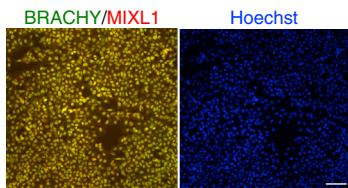
- Tonegawa, A., Funayama, N., Ueno, N., and Takahashi, Y. (1997). Mesodermal subdivision along the mediolateral axis in chicken controlled by different concentrations of BMP-4. *Development* 124, 1975–1984.
- Tran, D.Q., Andersson, J., Wang, R., Ramsey, H., Unutmaz, D., and Shevach, E.M. (2009). GARP (LRRC32) is essential for the surface expression of latent TGF- β on platelets and activated FOXP3+ regulatory T cells. *Proc. Natl. Acad. Sci. USA* 106, 13445–13450.
- Ueno, S., Weidinger, G., Osugi, T., Kohn, A.D., Golob, J.L., Pabon, L., Reincke, H., Moon, R.T., and Murry, C.E. (2007). Biphasic role for Wnt/ β -catenin signaling in cardiac specification in zebrafish and embryonic stem cells. *Proc. Natl. Acad. Sci. USA* 104, 9685–9690.
- Umeda, K., Zhao, J., Simmons, P., Stanley, E., Elefanty, A., and Nakayama, N. (2012). Human chondrogenic paraxial mesoderm, directed specification and prospective isolation from pluripotent stem cells. *Sci. Rep.* 2, 455.
- Waddington, C.H. (1940). *Organisers and Genes* (Cambridge, UK: Cambridge University Press).

Supplemental Figures

Cell

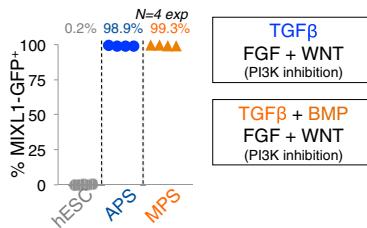
A Highly-efficient primitive streak induction

24 hours of hESC differentiation (immunostaining)



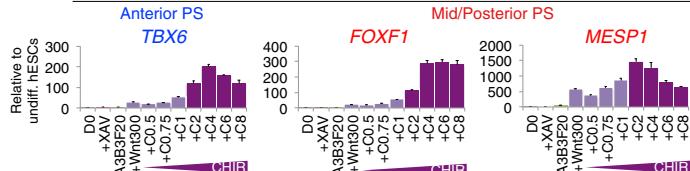
B Consistent primitive streak induction

24 hours of hESC differentiation (FACS analysis)



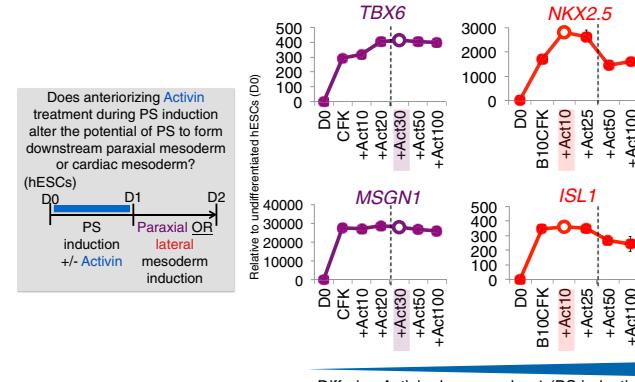
C WNT broadly induces different types of primitive streak

Day 1 gene expression (qPCR)



F High TGFβ instills paraxial, whereas low TGFβ endows cardiac mesoderm potential

i. Paraxial mesoderm (d2) ii. Cardiac mesoderm (d2)

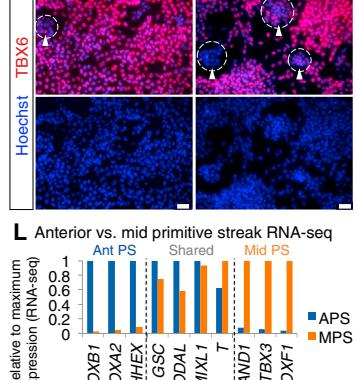


H BMP reduces paraxial mesoderm potential

| Mid TGFβ during PS induction enhances cardiac potential

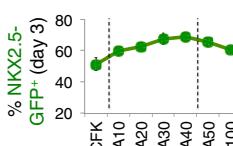
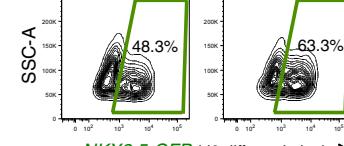


L Anterior vs. mid primitive streak RNA-seq



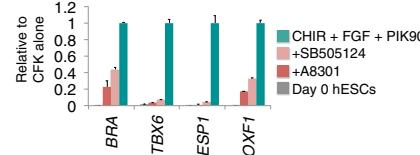
Day 1 PS (no Activin) Day 1 PS+Activin (20 ng/mL)

→ Day 2-3 CarM → Day 2-3 CarM



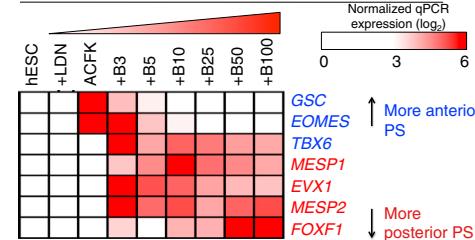
D Endogenous TGFβ critical for primitive streak

Day 1 gene expression (qPCR)



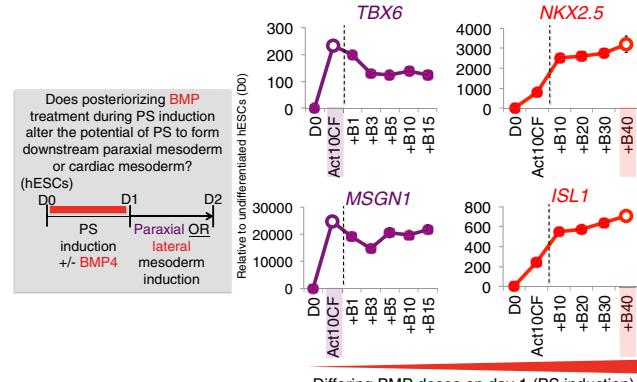
E BMP critical for PS induction and hi BMP posteriorizes the PS

Day 1 gene expression (qPCR)

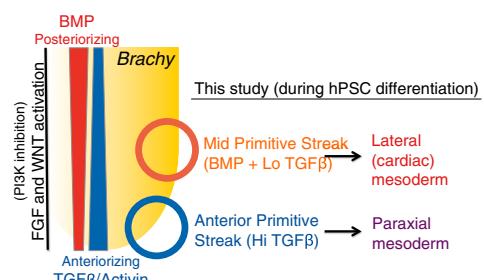


G Exogenous BMP endows cardiac, but inhibits paraxial mesoderm potential

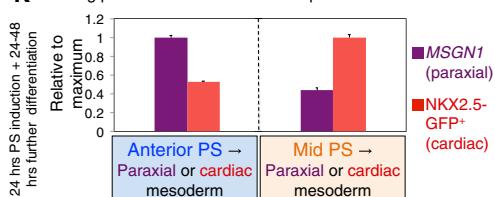
i. Paraxial mesoderm (d2) ii. Cardiac mesoderm (d2)



J Anterior-posterior patterning of primitive streak



K Differing potentials of anterior vs. mid primitive streak



(legend on next page)

Figure S1. Specification of the Anterior or Mid Primitive Streak Is Crucial for Subsequent Generation of Paraxial Mesoderm or Cardiac Mesoderm, Respectively, Related to Figure 1

(A) BRACHYURY and MIXL1 immunostaining of hESCs differentiated for 24 hr in mid primitive streak conditions revealed ubiquitous coexpression of both BRACHYURY and MIXL1, evincing the high efficiency of PS differentiation (and corroborated by FACS of a *MIXL1-GFP* reporter line; see Figure 1B); nuclear counterstaining by Hoechst dye

(B) Efficient *MIXL1⁺* PS specification is highly reproducible in both anterior and mid PS differentiation conditions (“APS” and “MPS,” respectively) across 4 independent differentiation experiments, as assayed by a *MIXL1-GFP* knockin reporter hESC line (Davis et al., 2008). Each circle or triangle represents an individual experiment and the average differentiation efficiency attained across all experiments is indicated.

(C) To test the effects of WNT signaling on PS induction, hESCs were differentiated for 24 hr with 3 ng/mL Activin + 3 ng/mL BMP4 + 20 ng/mL FGF2 (“A3B3F20”) in the presence or absence of a Wnt antagonist (1 μM XAV939) or else Wnt agonists (300 ng/mL WNT3A or 0.5–8 μM CHIR99021) and qPCR was conducted; these results showed that WNT activation broadly induces both anterior (*TBX6*) and mid/posterior (*FOXF1*, *MESP1*) PS markers

(D) To test whether TGFβ signaling was critical for PS induction, hESCs were differentiated toward PS using CHIR99021 + 20 ng/mL FGF2 + 100 nM PIK90, with or without TGFβR inhibitors (2 μM SB-505124 or 1 μM A-83-01); this revealed that TGFβ inhibition strongly impairs PS differentiation (cf. earlier reports that PS formation was still possible even under TGFβ inhibition (Mendjan et al., 2014))

(E) To assess whether inducing the PS in the presence of differing BMP levels would preferentially generate either anterior or posterior PS, hESCs were differentiated toward PS using Activin + CHIR + FGF2 + PIK90 (“ACFK”), in the presence or absence of a BMP antagonist (LDN193189, 250 nM) or increasing doses of BMP4 (3–100 ng/mL); this revealed that whereas BMP inhibition fully abolished PS formation, low (endogenous) levels of BMP supported more anterior PS, whereas higher levels of exogenous BMP induced more posterior PS

(F) To assess whether inducing PS in the presence of TGFβ signaling altered its downstream potential to form different kinds of mesoderm, (i) H7 hESCs were differentiated into PS with CHIR99021 + FGF2 + PIK90 (“CFK”) for 24 hr in the presence or absence of increasing amounts of Activin (10–100 ng/mL), and then were transferred into paraxial mesoderm induction conditions (A-83-01 + CHIR99021 + DM3189 + FGF2) for 24 hr, and qPCR was conducted or (ii) H7 hESCs were differentiated into PS with BMP4 (10 ng/mL) + CHIR99021 + FGF2 + PIK90 (“B10CFK”) for 24 hr in the presence or absence of increasing amounts of Activin (10–100 ng/mL), and then were transferred into lateral/cardiac mesoderm inductions (A-83-01 + BMP4 + C59) for 24 hr and qPCR was conducted. It was found that inducing PS in the presence of higher Activin doses (e.g., 30 ng/mL) instilled paraxial mesoderm potential, whereas PS induced in lower Activin (e.g., 10 ng/mL) had maximal lateral/cardiac mesoderm potential.

(G) To assess whether inducing PS in the presence of BMP signaling altered its downstream potential to form different kinds of mesoderm, H7 hESCs were differentiated into PS with Activin (10 ng/mL) + CHIR99021 + FGF2 + PIK90 (“A10CFK”) for 24 hr in the presence or absence of increasing amounts of BMP4 (1–40 ng/mL), and then were (i) transferred into paraxial mesoderm induction conditions (A-83-01 + CHIR99021 + DM3189 + FGF2) for 24 hr, and qPCR was conducted or were (ii) transferred into lateral/cardiac mesoderm inductions (A-83-01 + BMP4 + C59) for 24 hr and qPCR was conducted. It was found that inducing PS in the presence of BMP4 strongly blocked paraxial mesoderm potential, whereas induction in the presence of high BMP4 (40 ng/mL) maximized lateral/cardiac mesoderm potential.

(H) H7 hESCs were differentiated into anterior primitive streak (30 ng/mL Activin + 4 μM CHIR99021 + 20 ng/mL FGF2 + 100 nM PIK90) in the presence or absence of BMP4 (10 ng/mL) for 24 hr, and then transferred into paraxial mesoderm induction conditions for 24 hr before *TBX6* immunostaining was conducted. Inducing the primitive streak in the presence of BMP4 led to the subsequent generation of more abundant clusters of *TBX6*[−] non-paraxial mesoderm cells.

(I) *NKX2.5-GFP* hESCs were differentiated into primitive streak with BMP4 + CHIR99021 + FGF2 + PIK90 for 24 hr in the presence or absence of increasing doses of Activin (10–100 ng/mL), and then transferred into suboptimal lateral/cardiac mesoderm induction conditions (A-83-01 + BMP4 + C59, without FGF2) for 48 hr and FACS was performed after 72 hr of total differentiation to quantify *NKX2.5-GFP*⁺ cardiac mesoderm, whereas higher Activin doses (50–100 ng/mL) were less effective.

(J) Summary of proposed primitive streak patterning mechanisms, overlaid on a schematized ~E6.5 mouse embryo showing the domain of *Brachyury* expression in the PS region. FGF and WNT activation (together with PI3K inhibition) are permissive for pan-PS specification and though TGFβ and BMP signals are both also required for pan-PS specification (Loh et al., 2014), higher TGFβ levels anteriorize the PS whereas higher BMP levels posteriorize the PS. A combination of both TGFβ and BMP elicits mid PS. Anterior PS (induced in the presence of Activin together with FGF and WNT agonists and a PI3K inhibitor) is preferentially capable of generating paraxial mesoderm, whereas mid PS (induced in the presence of Activin and BMP together with FGF and WNT agonists and a PI3K inhibitor) is preferentially competent to generate lateral/cardiac mesoderm.

(K) Anterior and mid primitive streak populations formed within 24 hr of hESC differentiation have distinct developmental potentials. qPCR of H7 hESC differentiated initially into anterior or mid PS (day 0–1), followed by day 1–2 paraxial mesoderm induction; qPCR reveals that paraxial mesoderm genes are markedly higher in the APS-derived population (left); day 3 FACS of *NKX2.5-GFP* hESC (Elliott et al., 2011) differentiated into anterior or mid PS (day 0–1), followed by day 1–3 lateral/cardiac mesoderm induction; MPS is markedly more competent at generating cardiac progenitors (right)

(L) Bulk population RNA-seq of day 1 anterior and mid PS populations shows that whereas “pan-PS” markers such as *MIXL1* and *T/BRACHYURY* are expressed in both populations, anterior PS markers *FOXA2* and *HHEX* are preferentially expressed in anterior PS whereas mid PS marker *FOXF1* is preferentially expressed in mid PS; RNA-seq TPM values were normalized such that the highest expression value was set to 1.0

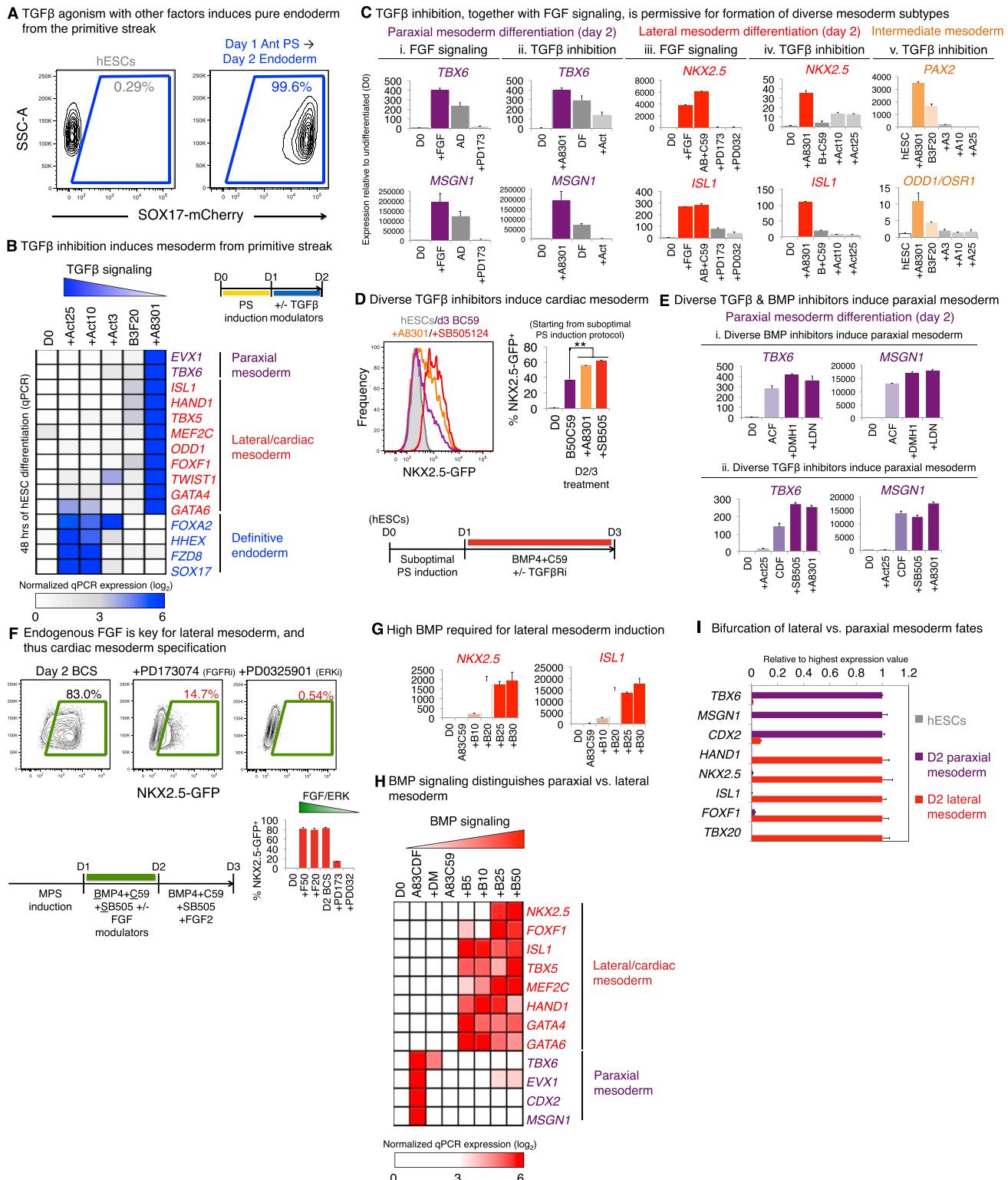


Figure S2. Signaling Logic for Specification of Paraxial Mesoderm or Cardiac Mesoderm from Primitive Streak, Related to Figure 1

(A) To assess whether TGF β activation drives primitive streak toward definitive endoderm, SOX17-mCherry H9 hESCs were differentiated to anterior primitive streak for 24 hr, after which they were treated with a TGF β agonist (Activin) in conjunction with other factors (as described by Loh et al., 2014) for 24 hr; FACS was performed on the day 2 population, which revealed that > 99% of cells were SOX17-mCherry $^{+}$, indicating their endodermal identity

(legend continued on next page)

(B) TGF β inhibition on day 1-2 specifies endoderm from PS. qPCR of day 1 PS treated with 3 ng/mL BMP4 + 20 ng/mL FGF2 ("B3F20") in the presence or absence of TGF β agonists (3-25 ng/mL Activin; "Act") or a TGF β R inhibitor (1 μ M A8301) for 24 hr.

(C) To assess whether FGF/ERK signaling was permissive for both paraxial and lateral/cardiac mesoderm specification, hESCs were differentiated into primitive streak for 24 hr, and then transferred into (i) A-83-01 + DM3189 ("AD"; permissive for paraxial mesoderm) or (ii) A-83-01 + BMP4 + C59 ("AB+C59"; permissive for lateral/cardiac mesoderm) for 24 hr in the presence or absence of a FGF agonist (FGF2, 10-20 ng/mL), an FGFR inhibitor (PD173074, 100 nM) or a MAPK/ERK inhibitor (PD0325901, 500 nM) for 24 hr and qPCR for all experiments was conducted on day 2 (that is, after 48 total hours of hESC differentiation). To assess whether TGF β signaling altered paraxial, intermediate, and lateral/cardiac mesoderm specification, hESCs were differentiated into primitive streak for 24 hr, and then transferred into (iii) DM3189 + FGF2 ("DF"; permissive for paraxial mesoderm) or (iv) BMP4 + C59 ("B+C59"; permissive for lateral/cardiac mesoderm) for 24 hr in the presence of a TGF β agonist (Activin, 10-25 ng/mL) or a TGF β R inhibitor (A-83-01, 1 μ M) for 24 hr or (v) BMP4 (3 ng/mL) + FGF2 (20 ng/mL) ("B3F20") for 24 hr in the presence of a TGF β agonist (Activin, 3-25 ng/mL) or a TGF β R inhibitor (A-83-01, 1 μ M) and qPCR for all experiments was conducted on day 2 (that is, after 48 total hours of hESC differentiation).

(D) To test whether distinct TGF β inhibitors truly induced cardiac mesoderm, *NKX2.5-GFP* hESCs were differentiated into primitive streak for 24 hr, and then treated with BMP4 (50 ng/mL) + C59 ("B50C59") for 48 hr in the presence or absence of one of two TGF β R inhibitors (SB-505124, 2 μ M or A-83-01, 1 μ M) for 48 hr; FACS was conducted on day 3, which confirmed that both TGF β R inhibitors increased *NKX2.5-GFP*⁺ cardiac mesoderm induction (* $p < 0.01$).

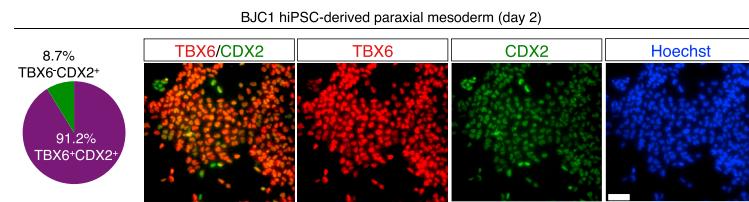
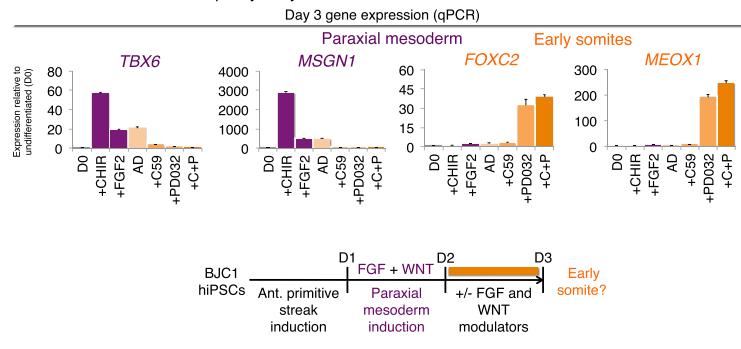
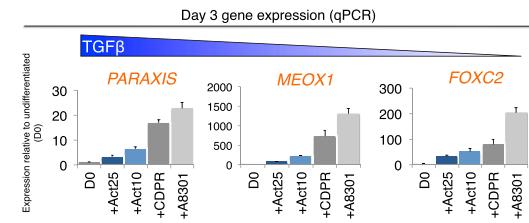
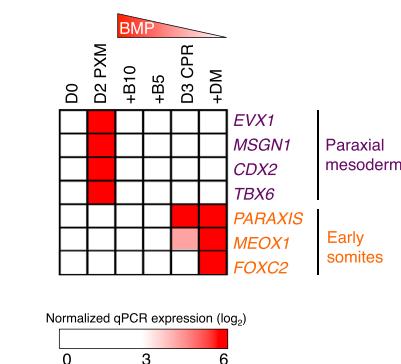
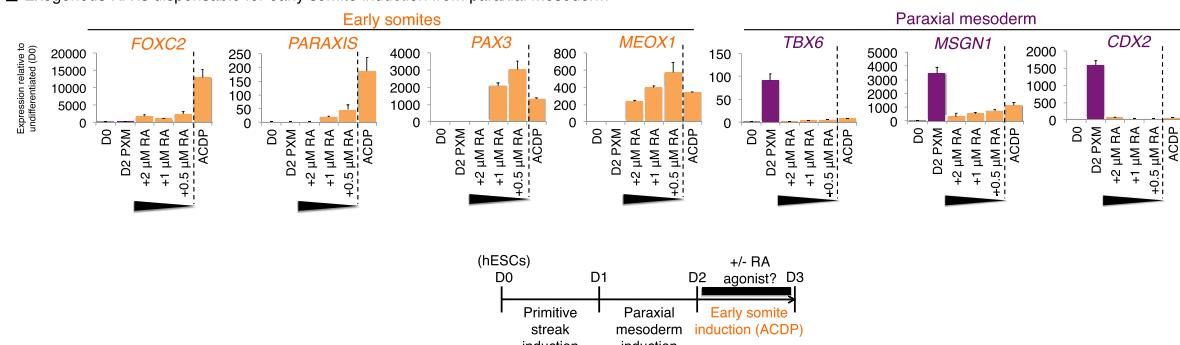
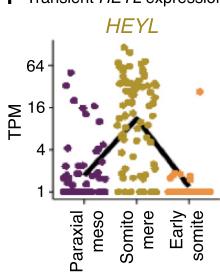
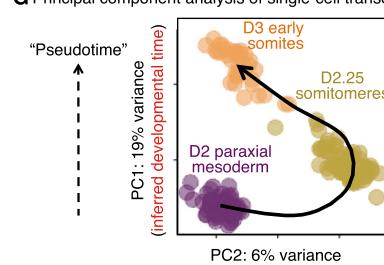
(E) To ensure that blockade of BMP and TGF β signaling truly induced paraxial mesoderm, multiple small-molecule inhibitors to these signaling pathways were used to ensure that they phenocopied one another. To this end, hESCs were differentiated toward primitive streak for 24 hr, and then subsequently transferred into (i) A-83-01 + CHIR99021 + FGF2 ("ACF") in the presence or absence of one of two BMPR inhibitors (DMH1, 1 μ M or LDN193189, 250 nM) for 24 hr and qPCR was conducted on day 2. Alternately, hESC-derived PS was transferred into (ii) CHIR99021 + DM3189 + FGF2 ("CDF") for 24 hr in the presence or absence of a TGF β agonist (25 ng/mL Activin) or one of two TGF β R inhibitors (SB-505124, 2 μ M or A-83-01, 1 μ M) for 24 hr and qPCR was conducted on day 2. These results affirmed that structurally distinct inhibitors against the BMP and TGF β pathways all induced paraxial mesoderm.

(F) To test the role of FGF in lateral mesoderm induction (on day 1-2 of hESC differentiation), *NKX2.5-GFP* hESCs were differentiated into mid primitive streak for 24 hr, and then were treated with BMP4 + C59 + SB505124 in the presence of an FGF agonist (FGF2, 20-50 ng/mL) or a FGFR inhibitor (PD173074, 100 nM) or a MAPK/ERK inhibitor (PD0325901, 500 nM) for 24 hr, and then all cultures were treated with BMP4 + C59 + SB505124 + FGF2 for 24 hr more before being harvested on day 3 for FACS. This showed that FGF/ERK signaling during lateral mesoderm formation (day 2 of hESC differentiation) is critical for the subsequent generation of cardiac mesoderm; FGF/ERK blockade on day 2 abolishes lateral mesoderm, and thus, cardiac mesoderm.

(G) To test the requirement for BMP signaling in lateral mesoderm induction (on day 2 of hESC differentiation), H7 hESCs were differentiated into primitive streak for 24 hr, and then were treated with A8301 + C59 for 24 hr in the presence or absence of BMP4 (10-30 ng/mL) and qPCR was conducted; this showed that high amounts of exogenous BMP were required to drive lateral mesoderm specification.

(H) BMP on day 1-2 specifies lateral mesoderm, and inhibits paraxial mesoderm from PS. qPCR of day 1 PS treated with 1 μ M A8301 + 1 μ M C59 ("A83C59") in the presence or absence of BMP4 (5-50 ng/mL) or BMP inhibitor (250 nM DM3189) for 24 hr; also included is another BMP inhibitor-containing condition (A83CDF containing DM3189).

(I) qPCR of day 2 H7 hESC-derived paraxial mesoderm or lateral mesoderm populations (as well as undifferentiated hESCs); gene expression normalized to the highest value = 1.0

A Efficient specification of TBX6⁺CDX2⁺ paraxial mesoderm**B WNT and ERK inhibition specify early somites from hiPSCs****C TGF β blockade enhances early somite formation****D BMP blockade enhances early somite formation****E Exogenous RA is dispensable for early somite induction from paraxial mesoderm****F Transient HEYL expression in somitomeres****G Principal component analysis of single-cell transcriptomes****Figure S3. Signaling Logic for Specification of Early Somite Progenitors from Paraxial Mesoderm, Related to Figure 2**

(A) TBX6 and CDX2 immunostaining of day 2 BJC1-derived paraxial mesoderm populations shows > 90% of cells co-express TBX6 and CDX2 (scale bar, 100 μ m), where *Tbx6* and *Cdx2* are paraxial mesoderm markers in mouse embryos (Beck et al., 1995; Chapman and Papaioannou, 1998)

(B) To determine the effects of FGF and WNT signaling on formation of early somite progenitors, good manufacturing practice (GMP)-compatible BJC1 hiPSCs were differentiated into day 2 paraxial mesoderm, and subsequently were treated with A8301 + DM3189 ("AD") for 24 hr in the presence or absence of a WNT agonist (CHIR99021, 3 μ M), an FGF agonist (FGF2, 20 ng/mL), a WNT inhibitor (C59, 1 μ M), a MAPK/ERK inhibitor (PD0325901, 500 nM) or combined WNT and ERK inhibitors (C59 + PD0325901; "+C+P"), which showed that dual WNT and ERK inhibition most strongly suppresses paraxial mesoderm genes and enhances early somite marker expression

(C) To determine how TGF β signaling influences early somite progenitor production from paraxial mesoderm, hESC-derived day 2 paraxial mesoderm was exposed to C59 + DM3189 + PD0325901 + RA ("CDPR") for 24 hr in the presence or absence of a TGF β inhibitor (A-83-01, 1 μ M) or a TGF β agonist (Activin, 10–25 ng/mL) and qPCR was conducted on day 3; this revealed that TGF β inhibition was permissive for expression of early somite markers

(D) To determine how BMP signaling influences early somite progenitor production from paraxial mesoderm, hESC-derived day 2 paraxial mesoderm was exposed to C59 + PD0325901 + RA ("CPR") for 24 hr in the presence or absence of a BMP inhibitor (DM3189, 250 nM) or a BMP agonist (BMP4, 5–10 ng/mL) and

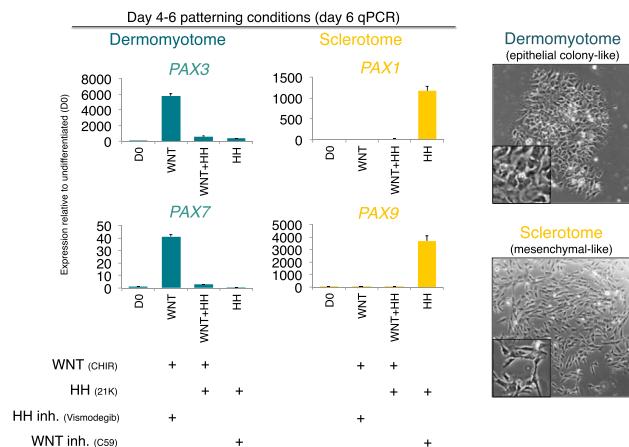
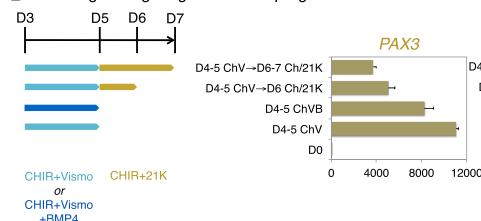
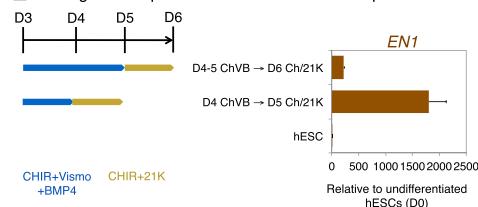
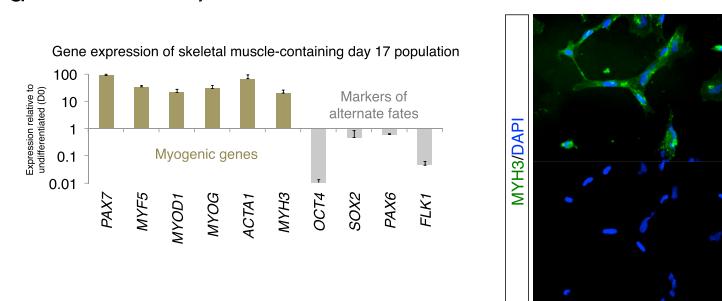
(legend continued on next page)

qPCR was conducted on day 3; for comparison, the gene expression of day 2 paraxial mesoderm was also determined; this revealed that BMP inhibition was permissive for expression of early somite markers

(E) To determine whether exogenous retinoic acid is required for early somite progenitor production from paraxial mesoderm, hESC-derived day 2 paraxial mesoderm was exposed to A8301 + C59 + DM3189 + PD0325901 ("ACDP") for 24 hr in the presence or absence of increasing concentrations of all-*trans* retinoic acid ("RA," 0.5-2 μ M) and qPCR was conducted on day 3 (the gene expression of day 2 paraxial mesoderm was also determined as a control); this revealed that exogenous RA was dispensable for the expression of early somite markers

(F) *HEYL* is transiently expressed during reconstituted human somitogenesis, as shown by single-cell RNA-seq of day 2 paraxial mesoderm, day 2.25 somitomere and day 3 early somite populations; each dot depicts expression in a single cell. Accordingly, *HEYL* is known to be expressed in mouse somitomeres *in vivo* (Leimeister et al., 2000).

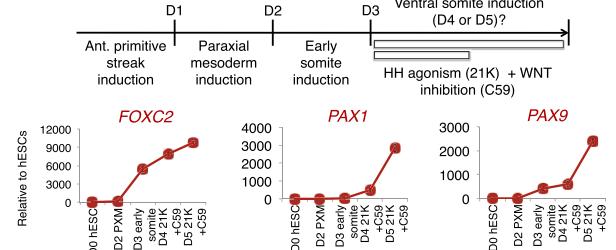
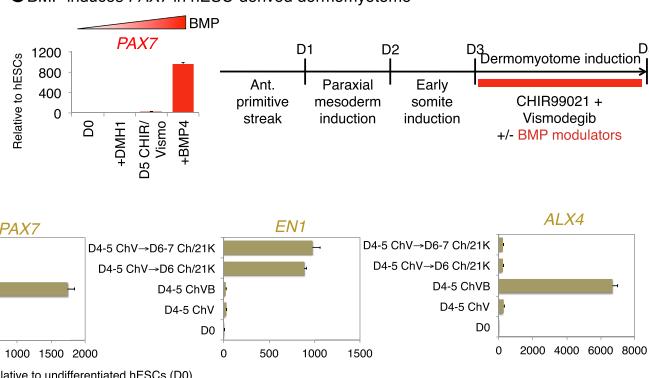
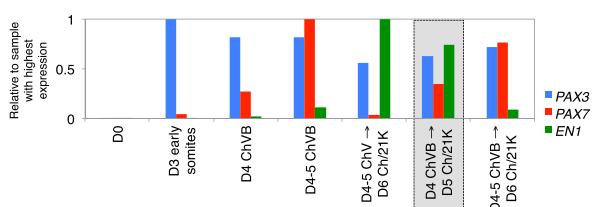
(G) Principal component analysis of single-cell RNA-seq data show that day 2 paraxial mesoderm, day 2.25 somitomere and day 3 early somite populations progress along a single continuous trajectory; each dot depicts a single cell. Moreover these populations are arranged along their "correct" temporal order (from day 2 to day 2.25 to day 3 of *in vitro* differentiation) along the PC1 axis. Hence as described in the Supplemental Experimental Procedures, we used the ordering of cells along PC1 to infer their progression along developmental "pseudotime" for Figures 3A and 3B.

A Dorsal-ventral patterning of early somites by cross-repressive WNT and HH signals**D Late-stage Hedgehog activation upregulates EN1 from dermomyotome****E Prolonged BMP pre-treatment blocks EN1 competence****G hESC-derived dermomyotome can differentiate towards skeletal muscle****Figure S4. Signaling Logic for Differentiation of Early Somite Progenitors into Sclerotome and Dermomyotome, Related to Figure 4**

(A) Day 3 early somite populations were treated with either ventralizing conditions ("HH"; HH agonist/WNT inhibitor), dorsalizing conditions ("WNT"; HH inhibitor/WNT agonist) or a hybrid of these ("WNT+HH"; HH agonist/WNT agonist) for 48 hr, and qPCR was conducted (left); morphology of cell populations after 24 hr of either in vitro sclerotome or dermomyotome differentiation shows a migratory, mesenchymal-like morphology of sclerotome cells and a more epithelial, colony-like morphology of dermomyotome (right), as is known in vivo

(B) To determine the length of time required for sclerotome differentiation, a qPCR time course was performed, encompassing undifferentiated hESCs ("day 0"), day 2 hESC-derived paraxial mesoderm ("day 2 PXM"), day 3 hESC-derived early somite progenitors, and hESC-derived day 3 early somite progenitors that were exposed to sclerotome inductive conditions (5 nM 21K + 1 μ M C59) for 24 or 48 hr

(C) Day 3 hESC-derived early somite progenitors were exposed to dermomyotome-inductive conditions (3 μ M CHIR99021 + 150 nM Vismodegib) for 48 hr in the presence or absence of a BMP agonist (BMP4, 10 ng/mL) or a BMPR inhibitor (DMH1, 1 μ M); this demonstrated that inducing dermomyotome in the presence of BMP greatly enhanced the expression of PAX7

B Minimum of 48 hours is required for sclerotome induction**C BMP induces PAX7 in hESC-derived dermomyotome****F Early transient BMP, followed by late Hh activation, is key for optimal PAX3/PAX7/EN1 expression**

(legend continued on next page)

(D) Day 3 hESC-derived early somite progenitors were exposed to dermomyotome-inductive conditions (either CHIR99021+Vismodegib or CHIR99021+Vismodegib+BMP4) for 24 hr, and then were either harvested for qPCR or further transferred into HEDGEHOG-stimulating conditions (CHIR99021 + 21K, 5 nM) for 24-48 hr; this showed that BMP is important to induce *PAX7* in central dermomyotome and that later stage HEDGEHOG activation is important to induce *EN1*

(E) Day 3 hESC-derived early somite progenitors were exposed to dermomyotome-inductive conditions (CHIR99021+Vismodegib+BMP4) for either 24 or 48 hr before being transferred to CHIR99021+21K in order to induce *EN1*; this showed that only 24 hr of CHIR99021+Vismodegib+BMP4 pre-treatment was optimal for *EN1* induction at later stages

(F) Combining insights from Figures S4D and S4E, it was found that exposure of day 3 hESC-derived early somite progenitors to CHIR99021+Vismodegib+BMP4 from day 3-4 (for 24 hr) followed by exposure to CHIR99021+21K for day 4-5 (for 24 hr) was optimal for the joint upregulation of *PAX3*, *PAX7* and *EN1* to induce central dermomyotome-like cells (dashed box)

(G) Day 5 hESC-derived dermomyotome was exposed to 2% horse serum (Xu et al., 2013) for 12 days to induce myogenic differentiation to some extent (fully described in the Supplemental Experimental Procedures), yielding a heterogeneous day 17 population of cells that contained skeletal muscle cells; qPCR revealed upregulation of myogenic genes whereas markers of alternate fates, including pluripotency (*OCT4*, *SOX2*), neuroectoderm (*SOX2* and *PAX6*) and vasculature (*FLK1*) were minimally expressed (*left*); immunostaining of the day 17 population revealed expression of skeletal muscle marker *MYH3*, with nuclear counterstaining by DAPI (*right*)

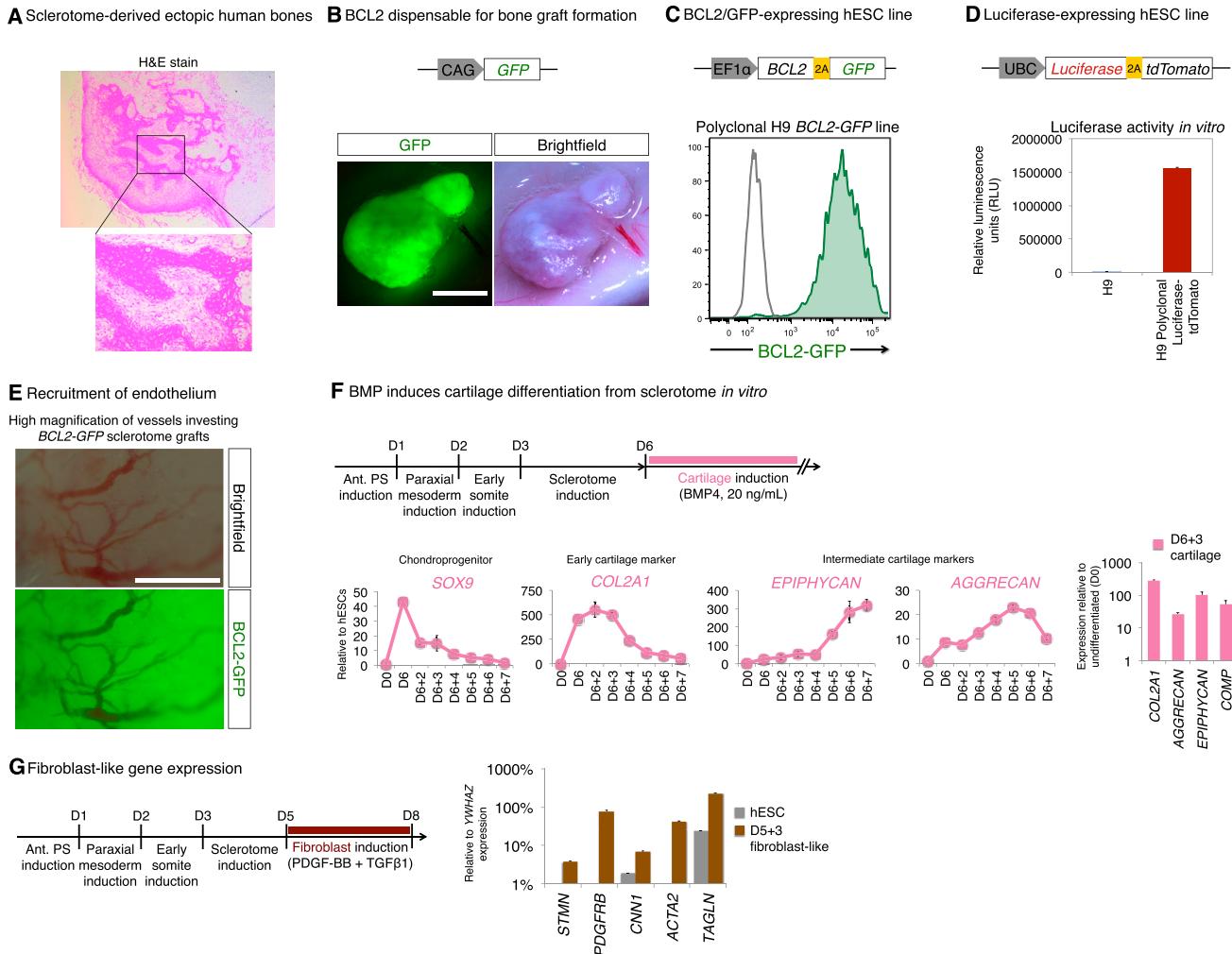


Figure S5. Developmental Potential of Human Sclerotome In Vivo and In Vitro, Related to Figure 4

- (A) H&E staining of ectopic bone graft reveals cartilaginous areas and a zone of a clear ossification; other types of tissue were not patently observed
- (B) Description of pCAG-GFP construct (top); day 6 sclerotome populations derived from pCAG-GFP⁺ H7 hESCs also engrafted when subcutaneously transplanted into NOD-SCID IIrg^{-/-} mice they formed ectopic bone grafts (bottom), therefore demonstrating that exogenous BCL2 expression is dispensable for sclerotome engraftment and bone formation
- (C) Description of EF1 α -BCL2-T2A-GFP lentiviral construct (otherwise known as "C306" (Ardehali et al., 2011); top) and FACS plot of a stably transduced polyclonal H9 hESC line with near-uniform expression of the BCL2-T2A-GFP transgene (green contour) by comparison to uncolored wild-type hESCs (bottom) (see *Supplemental Experimental Procedures* for details of derivation of the H9 BCL2-T2A-GFP hESC line)
- (D) Description of UBC-Luciferase-T2A-tdTomato lentiviral construct (top); bioluminescent assay of cultured H9 UBC-Luciferase-T2A-tdTomato hESCs validates that these cells are luciferase-expressing in vitro (bottom)
- (E) (i) Higher-magnification picture of BCL2-T2A-GFP hESC-derived sclerotome ectopic bone grafts (originally described in Figure 4E), showing that despite the near-ubiquitous expression of GFP throughout the ectopic bone graft (indicating its human provenance), the blood vessels are clearly GFP-negative and therefore presumably derived from the mouse host; scale bar, 1 mm
- (F) Day 5 hESC-derived sclerotome was differentiated toward fibroblast-like cells by exposure to PDGF-BB (10 ng/mL) + TGF β 1 (2 ng/mL) for 72 hr; qPCR of the resultant day 8 fibroblast-like population (brown bars) revealed significant upregulation of fibroblast-affiliated genes by comparison to the expression of these genes in undifferentiated hESCs (gray bars); y axis indicates gene expression relative to the expression of reference gene YWHAZ
- (G) Day 6 hESC-derived sclerotome was exposed to pro-chondrogenic BMP4 (20 ng/mL) for 2 days and qPCR was performed, revealing the upregulation of cartilage markers
- (H) Day 6 hESC-derived sclerotome was exposed to BMP4 (20 ng/mL) for varying lengths of time and time course qPCR was performed, which showed that SOX9 was rapidly downregulated upon BMP treatment, COL2A1 was already expressed in sclerotome to some extent and became downregulated after several days, and finally that EPIPHYCAN and AGGRECAN were only upregulated after prolonged BMP treatment

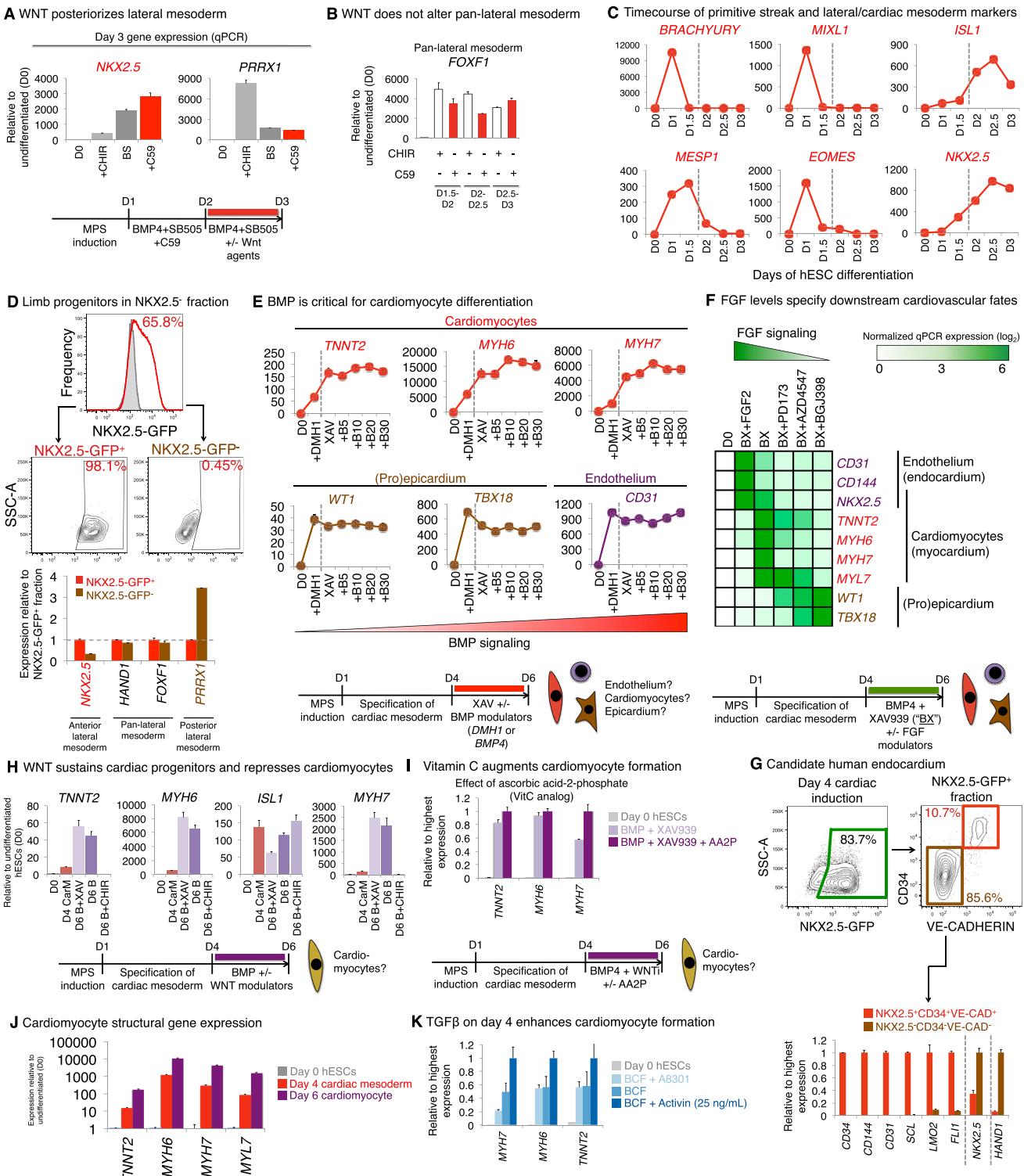


Figure S6. Signaling Logic for Sequential Production of Lateral Mesoderm, Cardiac Mesoderm, and finally, Cardiomyocytes, Related to Figure 5

(A) To assess the effects of WNT signaling on lateral mesoderm patterning on day 3 of hESC differentiation, day 2 hESC-derived lateral mesoderm was treated on day 3 with SB-505124 and BMP4 in the presence or absence of a WNT agonist (CHIR99021) or a WNT inhibitor (C59) for 24 hr and qPCR was performed; this

(legend continued on next page)

revealed that WNT activation specified a posterior lateral mesoderm/forelimb bud fate (*PRRX1*) whereas its inhibition induced an anterior lateral mesoderm/cardiac mesoderm fate

(B) To determine the effects of WNT on lateral mesoderm patterning, day 1 PS was differentiated to lateral mesoderm (30 ng/mL BMP4 + 1 μ M C59 + 2 μ M SB505124) for varying lengths of time (until day 2, day 2.5, or day 3) and for the last 12 hr was treated with C59 or 3 μ M CHIR (in addition to BS) and qPCR was conducted; this revealed that pan-lateral mesoderm marker *FOXF1* was not markedly changed in response to either WNT activation or inhibition (extension of Figure 5B)

(C) Time course of gene expression in hESCs differentiating toward mid primitive streak and lateral mesoderm, where mid PS cells were subsequently differentiated in SB-505124 + BMP4 + C59 for 48 hr. Primitive streak markers were transiently expressed on day 1 of hESC differentiation and sharply downregulated, being succeeded by lateral mesoderm markers that were upregulated by day 2.

(D) To reveal the possible identity of non-cardiac mesoderm lineages that were contaminating cardiac mesoderm-containing cultures, *NKX2.5-GFP* hESCs were differentiated toward mid primitive streak (day 1) and were differentiated toward a heterogeneous population containing cardiac mesoderm (day 3) by treatment with SB-505124 + BMP4 + C59 (without FGF2) on days 2-3 of differentiation. FACS was performed on day 3 to purify either *NKX2.5-GFP⁺* or *NKX2.5-GFP⁻* cells (top subpanels), which were sorted for qPCR (bottom subpanels). Gene expression was normalized to expression of a gene in the *NKX2.5-GFP⁺* fraction (dotted line).

(E) To determine the effects of BMP on downstream differentiation of cardiac mesoderm into various cardiovascular fates, day 4 hESC-derived cardiac mesoderm was treated with XAV939 (1 μ M) for 48 hr in presence or absence of BMP4 (5–30 ng/mL) or the BMP receptor antagonist DMH1 (1 μ M); this showed that BMP treatment preferentially promotes commitment to cardiomyocyte fate (*TNNT2, MYH6, MYH7*) without promoting endothelium/endocardium or (pro)epicardium specification

(F) To determine the effects of FGF on downstream differentiation of cardiac mesoderm into various cardiovascular fates, day 4 hESC-derived cardiac mesoderm was treated with BMP4 (30 ng/mL) + XAV939 (1 μ M) ("BX") for 48 hr in the presence or absence of FGF2 or various FGF receptor antagonists (PD173074, AZD4547 or BGJ398; all at 100 nM); this showed that high FGF signaling promoted endocardium, endogenous FGF levels promoted cardiomyocytes and FGF inhibition promoted (pro)epicardium

(G) In the mouse embryo, endocardium (progenitors to heart-affiliated endothelium) are marked by transient coexpression of pan-heart marker *Nkx2.5* together with endothelial markers (Ferdous et al., 2009; Stanley et al., 2002). Day 4 cardiac mesoderm derived from *NKX2.5-GFP* knockin reporter hESCs was gated on the *NKX2.5-GFP⁺* subset, which was heterogeneous and contained an *NKX2.5⁺CD34⁺VE-CADHERIN⁺* putative endocardium-like population and an *NKX2.5⁺CD34⁻VE-CADHERIN⁻* non-endocardium population (top). qPCR of these two sorted populations revealed that both expressed pan-heart marker *NKX2.5*, but only the *NKX2.5⁺CD34⁺VE-CADHERIN⁺* endocardium-like population expressed high levels of endothelial markers whereas the *NKX2.5⁺CD34⁻VE-CADHERIN⁻* fraction expressed higher levels of *HAND1* (bottom).

(H) To assess whether WNT signaling effects the differentiation of cardiac mesoderm into downstream derivatives, day 4 hESC-derived cardiac mesoderm populations ("day 4 Carm") were exposed to BMP4 (30 ng/mL) in the presence or absence of a WNT agonist (CHIR99021, 3 μ M) or a WNT antagonist (XAV939, 1 μ M) for 48 hr and qPCR was conducted of day 6 populations; this revealed that WNT activation fully blocked cardiac mesoderm differentiation into cardiomyocytes

(I) To assess whether a stabilized vitamin C analog (ascorbic acid-2-phosphate; "AA2P") enhanced cardiac mesoderm differentiation into cardiomyocytes, day 4 hESC-derived cardiac mesoderm was exposed to BMP4 (30 ng/mL) + XAV939 (1 μ M) for 48 hr in the presence or absence of 200 μ g/mL AA2P and qPCR was conducted of day 6 populations; qPCR data were normalized such that expression values in the highest-expressing sample was normalized = 1.0

(J) qPCR of day 4 cardiac mesoderm and day 6 cardiomyocytes reveals high expression of various cardiomyocyte cytoskeletal genes in the latter

(K) Day 3 hESC-derived cardiac mesoderm was treated with BMP4 (30 ng/mL) + C59 (1 μ M) + FGF2 (20 ng/mL) for 24 hr in the presence or absence of Activin (25 ng/mL) or the TGF β inhibitor A-83-01 (1 μ M) and then subsequently differentiated into cardiomyocytes for an additional 48 hr in the presence of BMP4 + XAV939; this showed that TGF β activation at the cardiac mesoderm stage enhanced subsequent cardiomyocyte differentiation, potentially consistent with the requirement for TGF β signaling in zebrafish second heart field development (Zhou et al., 2011)

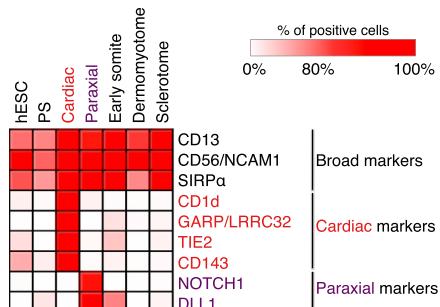
A Specific and non-specific markers of mesodermal subtypes

Figure S7. Stepwise Cell-Surface Marker, Transcriptional, and Chromatin Changes during Human Mesoderm Differentiation, Related to Figure 6 and Figure 7

- (A) Antibody screening by FACS revealed that currently defined mesoderm surface markers including CD56/NCAM1 (Evseenko et al., 2010), SIRP α (Dubois et al., 2011) and CD13 (Ardehali et al., 2013; Drukier et al., 2012) are promiscuously expressed across hESCs, paraxial mesoderm and cardiac mesoderm. Heatmap depicts surface marker expression on indicated lineages (PS and cardiac mesoderm populations were first pre-gated on MIXL1-GFP $^+$ and NKX2.5-GFP $^+$ fractions, as above).
- (B) FACS analysis of CD13, SIRP α or ROR2 expression on undifferentiated NKX2.5-GFP hESCs (gray silhouette) and those that were differentiated in cardiac mesoderm-inducing conditions for 3 days, gating on either NKX2.5-GFP $^+$ cardiac mesoderm (red line) or the NKX2.5-GFP $^-$ non-cardiac fraction (blue line); a fluorescence-minus one (FMO) control was included as a negative control (black line); this revealed that expression levels of CD13, SIRP α or ROR2 in cardiac mesoderm and hESCs were partially overlapping indicating that these markers incompletely discriminate between pluripotent versus cardiac fates.
- (C) Results from high-throughput cell-surface marker screen of NKX2.5-GFP hESC-derived cardiac mesoderm; gating scheme to identify NKX2.5-GFP $^+$ and NKX2.5-GFP $^-$ populations emerging by day 3 of hESC differentiation (left) and expression of selected markers TIE2 and GARP by FACS (right) shows that both markers were uniformly expressed in the NKX2.5-GFP $^+$ cardiac mesoderm lineage by comparison to an isotype control (gray line) and are present at lower levels in NKX2.5-GFP $^-$ mesoderm.
- (D) In situ hybridization for *deltaC* expression in 18 hpf zebrafish embryos (left; hpf: hours post-fertilization) and a cartoon schematic summarizing the *deltaC* expression pattern observed; namely that it was predominately in the U-shaped paraxial mesoderm and to a lesser extent in the posterior halves of the formed somites (right), corroborating earlier studies of *deltaC* expression in zebrafish embryos (Smithers et al., 2000).
- (E) Good Manufacturing Practice (GMP)-compatible BJC1 hiPSCs were differentiated into day 2 paraxial mesoderm populations, which were sorted into either DLL1 $^+$ or DLL1 $^-$ fractions and qPCR was conducted of these fractions as well as undifferentiated hiPSCs; this revealed that paraxial mesoderm transcription factors *TBX6* and *MSGN1* were exclusively expressed in the DLL1 $^+$ fraction, reaffirming the validity of DLL1 as a paraxial mesoderm marker.
- (F) PDGFR α FACS of undifferentiated H7 hESCs (gray) or those that were differentiated into day 5 sclerotome (gold) or day 5 dermomyotome (blue), the latter of which was induced by treating day 3 early somite progenitors with BMP4 + CHIR + Vismodegib ("BCV") for 48 hr; this revealed that expression of PDGFR α was enriched in sclerotome relative to dermomyotome.
- (G) Antibody screening by FACS shows that cell-surface markers present on hESCs tend to persist on their differentiated progeny, thus complicating the use of such markers to confirm or exclude the presence of undifferentiated cells in differentiated cultures. Intensity of color indicates percentage of cells expressing a given surface marker in each distinct population.
- (H) Bulk-population RNA-seq shows that a subset of genes associated with various human congenital heart defects (Fahed et al., 2013) are expressed in human lateral and/or cardiac mesoderm progenitors.
- (I) *PRRX1* and *FOXC2* expression in the human mesoderm hierarchy by RNA-seq; color intensity depicts gene expression (\log_2 TPM) normalized to the expression of that gene in all populations profiled, with the highest-expressing population assigned the most intense color value.
- (J) Non-binarized version of ATAC-seq heatmap shown in Figure 7D; colored boxes denote collections of open chromatin elements that have corresponding transcription factor enrichments indicated in Figure 7D.
- (K) Differentially expressed transcription factors from the *TBX* and *GATA* families across mesodermal lineages as determined by bulk-population RNA-seq; each row represents expression of a single gene and each column reflects expression in a given lineage.

Supplemental Information

Mapping the Pairwise Choices

Leading from Pluripotency to Human

Bone, Heart, and Other Mesoderm Cell Types

Kyle M. Loh, Angela Chen, Pang Wei Koh, Tianda Z. Deng, Rahul Sinha, Jonathan M. Tsai, Amira A. Barkal, Kimberle Y. Shen, Rajan Jain, Rachel M. Morganti, Ng Shyh-Chang, Nathaniel B. Fernhoff, Benson M. George, Gerlinde Wernig, Rachel E.A. Salomon, Zhenghao Chen, Hannes Vogel, Jonathan A. Epstein, Anshul Kundaje, William S. Talbot, Philip A. Beachy, Lay Teng Ang, and Irving L. Weissman

Supplemental Experimental Procedures

Human pluripotent stem cell culture

H7, *MIXL1-GFP* HES3, *NKX2.5-GFP* HES3, *SOX17-mCherry* H9, *pCAG-GFP* H7, *EF1A-BCL2-2A-GFP* H9 and *UBC-Luciferase-2A-tdTomato;EF1A-BCL2-2A-GFP* H9 hESCs and BJC1 hiPSCs were routinely propagated feeder-free in mTeSR1 medium (StemCell Technologies) on cell culture plastics coated with Geltrex basement membrane matrix (Gibco). Undifferentiated hPSCs were maintained at high quality with particular care to avoid any spontaneous differentiation, which would confound downstream differentiation. Unless otherwise indicated, the majority of experiments performed in this study were conducted using H7 hESCs, including all bulk-population RNA-seq, single-cell RNA-seq and ATAC-seq experiments.

Directed differentiation in defined medium

Partially-confluent wells of undifferentiated hPSCs were dissociated into very fine clumps using Accutase (Gibco) and sparsely passaged ~1:12-1:20 onto new Geltrex-coated cell culture plates in mTeSR1 supplemented with 1 µM thiazovivin (Tocris; a ROCK inhibitor to prevent cell death after dissociation) overnight. Seeding hPSCs sparsely prior to differentiation was critical to prevent cellular overgrowth during differentiation, especially during long-duration differentiation. hPSCs were allowed to plate overnight, and the following morning, were briefly washed (in DMEM/F12) before addition of differentiation medium. All differentiation was conducted in serum-free, feeder-free and monolayer conditions in chemically-defined CDM2 basal medium.

The composition of CDM2 basal medium (Loh et al., 2014) was as follows: 50% IMDM (+GlutaMAX, +HEPES, +Sodium Bicarbonate; Gibco, 31980-097) + 50% F12 (+GlutaMAX; Gibco, 31765-092) + 1 mg/mL polyvinyl alcohol (Sigma, P8136-250G) + 1% v/v concentrated lipids (Gibco, 11905-031) + 450 µM monothioglycerol (Sigma, M6145) + 0.7 µg/mL insulin (Roche, 1376497) + 15 µg/mL transferrin (Roche, 652202) + 1% v/v penicillin/streptomycin (Gibco). Polyvinyl alcohol was brought into solution by gentle warming and magnetic stirring in IMDM/F12 media before addition of additional culture supplements.

Primitive streak induction

As aforementioned, after overnight plating, hPSCs were briefly washed (with DMEM/F12) and then differentiated into either anterior primitive streak (30 ng/mL Activin A + 4 µM CHIR99021 + 20 ng/mL FGF2 + 100 nM PIK90; for subsequent paraxial mesoderm induction) or mid primitive streak (30 mg/mL Activin A + 40 ng/mL BMP4 + 6 µM CHIR99021 + 20 ng/mL FGF2 + 100 nM PIK90; for subsequent cardiac mesoderm induction) for 24 hrs.

Subsequently, day 1 anterior primitive streak was briefly washed (DMEM/F12) and differentiated towards day 2 paraxial mesoderm for 24 hours (1 µM A-83-01 + 3 µM CHIR99021 + 250 nM LDN-193189 [DM3189] + 20 ng/mL FGF2). Separately, day 1 mid primitive streak was differentiated towards day 2 lateral mesoderm for 24 hours (1 µM A-83-01 + 30 ng/mL BMP4 + 1 µM C59; with 2 µM SB-505124 sometimes used instead of A-83-01).

Paraxial mesoderm downstream differentiation

Day 2 paraxial mesoderm was briefly washed (DMEM/F12) and further differentiated into day 3 early somite/somitomere precursors for 24 hours (1 µM A-83-01 + 250 nM LDN-193189 + 1 µM C59 + 500 nM PD0325901). Subsequently, day 3 early somites were dorsoventrally patterned into either ventral somites/sclerotome (5 nM 21K + 1 µM C59) or dorsal somites/dermomyotome (3 µM CHIR99021 + 150 nM Vismodegib). Sclerotome induction was conducted for 48-72 hours (leading to day 5-6 ventral somite progenitors). For dermomyotome induction, sometimes dermomyotome was induced in the presence of 50 ng/mL BMP4 to upregulate *PAX7* after 48 hours of BMP4 + CHIR99021 + Vismodegib differentiation (leading to day 5 dermomyotome progenitors). Media was changed 24 hours for all steps. The small-molecule Hedgehog agonist 21K (Brunton et al., 2009) was commercially synthesized.

Differentiation into downstream somitic fates

- Cartilage: Day 6 sclerotome was briefly washed (DMEM/F12) and treated with BMP4 (20 ng/mL) for 72-144 hours, leading to day 9 or day 12 cartilage-like progeny. Media was changed every 24 hours.
- Fibroblasts: Day 5 sclerotome was briefly washed (DMEM/F12) and treated with TGF β 1 (2 ng/mL) + PDGF-BB (10 ng/mL) (Cheung et al., 2012) for 72 hours, leading to day 8 fibroblast-like progeny. Media was changed every 24 hours.
- Skeletal muscle: For myogenic differentiation, dermomyotome was initially induced via a modified approach. Day 3 early somites were briefly washed (DMEM/F12), treated with BMP4 + CHIR99021 + Vismodegib for 24 hours, washed again (DMEM/F12) and subsequently treated with CHIR99021 (3 μ M) + 21K (5 nM) for 24 hours. This day 5 dermomyotome was further empirically differentiated towards skeletal muscle-like progeny (using medium containing 2% horse serum, as previously described (Xu et al., 2013)) for 12 days, yielding day 17 skeletal muscle-containing populations.

Lateral/cardiac downstream differentiation

Day 2 lateral mesoderm was differentiated into day 4 cardiac mesoderm by treating them with 1 μ M A8301 + 30 ng/mL BMP4 + 1 μ M C59 + 20 ng/mL FGF2 for 48 hrs, or alternatively, with 1 μ M A8301 + 30 ng/mL BMP4 + 20 ng/mL FGF2 for 24 hrs followed by 25 ng/mL Activin + 30 ng/mL BMP4 + 1 μ M C59 for the next 24 hrs (schema in **Fig. 5d**; positive effect of mid-stage Activin shown in **Fig. S6k**). Subsequently, day 4 cardiac mesoderm was briefly washed (DMEM/F12) and treated with 30 ng/mL BMP4 + 1 μ M XAV939 + 200 μ g/mL 2-phospho-ascorbic acid (Sigma) for 48-96 hrs to yield day 6-8 cardiomyocyte-containing populations. Spontaneously contracting cardiomyocyte foci were evident from day 8 onwards. Media was changed 24 hours for all steps.

Immunostaining

Adherent cells were fixed with 4% formaldehyde (Electron Microscopy Services, in PBS) for 15 minutes (4 °C) and washed twice (PBS). They were then permeabilized and blocked in perm/blocking buffer (PBS + 0.1% Triton X100 [Sigma] + 1% donkey serum [Jackson Immunoresearch]) for 1 hour (4 °C) and then stained overnight with primary antibody diluted in perm/blocking buffer. Subsequently, cells were washed twice (perm/blocking buffer) and stained with secondary antibody (diluted in perm/blocking buffer) for 1 hour (4 °C). For nuclear counterstaining, the cells were stained with Hoechst 3342 (1:1000, diluted in perm/blocking buffer) for 5 mins and then washed twice more prior to conducting fluorescent microscopy. A list of primary antibodies used for immunostaining is provided in **Table S5**.

Safranin-O staining for hPSC-derived cartilage

hPSC-derived cartilage progeny were stained with safranin-O to assess their production of acidic proteoglycans, which is a key trademark of cartilage phenotypic function. Safranin-O staining was conducted in accord with standard procedures. Namely, hPSC-derived cartilage progeny at day 12 of differentiation (day 6 sclerotome treated with chondrogenic BMP4 for 6 further days) were fixed (with 4% formaldehyde, for 10 minutes at room temperature), washed twice in PBS, briefly treated with 1M acetic acid for 15 seconds, stained with 1% safranin-O solution for 10 minutes, washed with PBS, and wide-field visualization was performed with a dissecting microscope (Leica M205 FA).

High-throughput cell-surface marker screening

hESCs or their differentiated mesoderm progeny were dissociated (using TrypLE Express) and plated into individual wells of four 96-well plates, each well containing a distinct antibody against a human cell-surface antigen, altogether totaling 332 unique cell-surface markers across multiple 96-well plates (LEGENDScreen PE-Conjugated Human Antibody Plates; Biolegend, 700001). For each LEGENDScreen experiment, approximately 10-70 million cells of each lineage were used. High-throughput cell-surface marker staining was largely done as per the manufacturer's recommendations, and cells were stained with a viability dye (DAPI, 1.1 μ M; Biolegend) prior to analysis on an LSR Fortessa (Stanford Stem Cell Institute FACS Core). Stained cells were not fixed prior to FACS analysis. Sometimes, after lysophilized antibodies were reconstituted in LEGENDScreen plates they were aliquoted into a separate plate to generate replicates of antibody arrays.

The following cell populations were used for LEGENDScreen analyses (**Fig. 6**): H7 hESCs (“undifferentiated hESCs”), H7-derived day 2 paraxial mesoderm (“paraxial mesoderm”), H7-derived day

3 early somite progenitors (“early somite”), H7-derived day 5 dermomyotome (“dermomyotome”), H7-derived day 6 sclerotome (“sclerotome”), *MIXL1-GFP* HES3 hESC-derived day 1 anterior primitive streak (“primitive streak”) and finally, *NKX2.5-GFP* HES3 hESC-derived day 3 cardiac mesoderm (“cardiac mesoderm”).

For the data plotted in **Fig. 6a** and **Fig. 6b**, cell surface marker analysis of day 1 primitive streak was conducted with *MIXL1-GFP* HES3 hESCs and analysis of day 3 cardiac mesoderm was conducted with *NKX2.5-GFP* HES3 hESCs; in each case, pre-gating on the *MIXL1-GFP⁺* and *NKX2.5-GFP⁺* fractions was respectively done before visualizing cell-surface marker expression.

Fluorescence activated cell sorting (FACS) by virtue of cell-surface marker expression

Flow cytometry was conducted largely as previously described (Loh et al., 2014). hESCs or their differentiated derivatives were dissociated using TrypLE Express (Gibco), were washed off the plate with FACS buffer (PBS + 0.1% BSA fraction V [Gibco] + 1 mM EDTA [Gibco] + 1% penicillin/streptomycin [Gibco]) and were pelleted by centrifugation (5 mins, 4 °C). (Sometimes, in order to decrease clumping at downstream steps, dissociated cells were strained through a 70 µm filter prior to pelleting.) Subsequently, cell pellets were directly resuspended in FACS buffer containing pre-diluted primary antibodies (**Table S6**), thoroughly triturated to ensure a single cell suspension, and primary antibody staining was conducted for 30 mins on ice. Afterwards, cells were washed with an excess of FACS buffer and pelleted again, and this was conducted one more time. Finally, washed cell pellets were resuspended in FACS buffer containing 1.1 µM DAPI (Biologend), and were strained through a 30 µm filter. Flow cytometry and sorting was conducted on a BD FACSAria II (Stanford Stem Cell Institute FACS Core).

Intracellular flow cytometry

To quantitatively analyze the expression of SMA α /ACTA2 or TROPONIN/TNNT2 in hPSC-derived fibroblast-like or hPSC-derived cardiomyocyte-like cells, respectively, intracellular flow cytometry was conducted using the Cytofix/Cytoperm kit (BD Biosciences, 554714) largely as per the manufacturer’s instructions. Adherent cell populations were briefly washed with HBSS (lacking Ca²⁺ and Mg²⁺) to remove dead or floating cells, dissociated with TrypLE Express (Gibco), washed off the plate with FACS buffer (composition above), and pelleted. Subsequently, cells were fixed in Cytofix/Cytoperm buffer (20 minutes, on ice), washed twice in 1X Perm/Wash buffer, and directly resuspended in 50–100 µL of 1X Perm/Wash buffer containing staining antibodies (1:10 dilution of α-SMA α PE [R&D Systems] or α-TNNT2 PE [BD Biosciences]; **Table S6**). SMA α staining was conducted for 30 mins on ice. TNNT2 staining was conducted for 45 minutes at room temperature, as previously recommended (Burridge et al., 2014). After antibody staining, cells were washed twice (Perm/Wash buffer), and finally resuspended in FACS buffer before straining through a 70 µm filter and flow cytometry was conducted on a BD FACSAria II (Stanford Stem Cell Institute FACS Core).

In situ hybridization of zebrafish embryos

In situ hybridization was performed generally as previously described (Thisse and Thisse, 2008). Briefly, zebrafish embryos of the indicated stages (18–24 h.p.f.) were fixed in 4% paraformaldehyde, dehydrated in 100% methanol and rehydrated in PBST (PBS + 0.1% Tween 20). They were then treated with proteinase K (diluted 1:2000 from a 10 µg/mL stock) for 1 minute, with care to avoid over-digestion of embryos. They were then re-fixed in 4% paraformaldehyde for 20 mins and pre-hybridized in hybridization buffer (50% formamide + 5x SSC buffer + 0.1% Tween + 9 mM citric acid, pH 6 + 0.5 mg/mL yeast extract) before incubation with digoxigenin-labeled probes *deltaC*, *lrrc32* and *pdgfra* probes overnight at 65 °C. Embryos were washed in SSC buffer before incubation with anti-digoxigenin F_{ab} conjugated to alkaline phosphatase (diluted 1:3000; Roche, 11093274910) overnight in MAB Buffer (0.1 mM Maleic Acid + 150 mM NaCl, pH 7.5) + 0.1% Triton-X100 + 2% blocking reagent (Roche, 11096176001) + 10% serum. Subsequently, embryos were washed six times in MAB Buffer + 0.1% Triton-X100 and developed through alkaline phosphatase treatment for optimized durations of time. Embryos were finally cleared with 0.1M glycine + 0.15% Tween, pH 2.2 and images were captured using the Zeiss AxioCam HRc color camera and processed using AxioVision imaging software.

Primers used for synthesis of *in situ* hybridization probes are as follows: *pdgfra*_forward (GGAGATCTGGTGAATTACCTTC), *pdgfra*_reverse (GAAGCTGGAATCCACAACC), *lrrc32*_forward (AGCTGAGGATTCACCTGTC), *lrrc32*_reverse (TGGAACCTCGTCCTCCATGTC), *deltaC*_forward (GCACGAGCAGTGTGTAAAG), *deltaC*_reverse (TTGCAAGTGCAGTAAAAGG); primers for synthesis of *pdgfra* probes were described previously (Kartopawiro et al., 2014).

Histological analysis of lineage-traced mouse embryos

To genetically trace the progeny of *Hopx*⁺ cells in mouse embryos, heterozygous *Hopx-IRES-Cre* driver mice (i.e., *Hopx*^{IRES-Cre/+}) (Jain et al., 2015) were mated to heterozygous *ROSA26-CAG-LSL-tdTomato* (i.e., *ROSA26*^{lslg/+}) reporter mice (Madisen et al., 2010) to yield E14.5 embryos. Embryos were fixed in 2% paraformaldehyde (4°C overnight), dehydrated through an ethanol series, embedded in paraffin, sectioned and stained with a rabbit anti-RFP antibody (Rockland) to better visualize tdTomato⁺ cells. Sections were imaged (Nikon Eclipse 80i fluorescence microscope) and images were analyzed using Adobe Photoshop, ImageJ and/or FIJI (for sizing, brightness and/or contrast adjustments). Brightness and contrast were adjusted linearly across the entire image for any particular image, including individual color channels for merged images.

Construction of *BCL2-T2A-GFP*-expressing hESC line

H9 hESCs were passaged with accutase, and 1-2 days post-seeding, were infected in mTeSR1 culture media with a *EF1α-BCL2-T2A-GFP* lentivirus (C306, provided by Matt Inlay) in which a constitutive *EF1α* promoter drives the expression of a *BCL2-T2A-GFP* cassette (Ardehali et al., 2011). Over the next 2 weeks, transduced populations were FACS sorted twice to enrich for GFP⁺ cells until a near-homogeneous (>95%) population of GFP⁺ hESCs was generated (Fig. S5c). *BCL2* protein expression was confirmed by intracellular FACS (not shown). During FACS sorting, GFP⁺ hESCs were directly sorted into mTeSR1 in the collection tube and then were subsequently plated in thiazovivin (1 μM) and ciprofloxacin (10 μg/mL) for 2-3 days to enhance recovery at a density of ~1x10⁵ hPSC/well in a 6-well plate.

In vivo transplantation to generate human ectopic bones

hESCs (H9 *EF1A-BCL2-T2A-GFP*-labeled, H9 *EF1A-BCL2-T2A-GFP;UBC-Luciferase-2A-tdTomato*-double labeled or H7 pCAG-GFP) were differentiated towards day 6 sclerotome in 10-cm dishes, and were harvested by briefly washing with PBS followed by dissociation by TrypLE Express. After centrifugation and pelleting, cells were resuspended in a small volume of cold CDM2 basal medium, counted, and diluted with an equal volume of cold Matrigel (BD Biosciences, 354234), yielding a cell suspension in a 1:1 mixture of CDM2/Matrigel that was kept on ice until transplantation.

Immunodeficient NOD-SCID *Il2rγ^{-/-}* (NSG) mice of 2-3 months of age were used for transplantation and were anesthetized by inhaled isoflurane. For subcutaneous transplantation, the skin along the midline was tented using forceps and approximately 100 μL of cell suspension was injected, with care not to puncture the skin at the opposite end of the tent. After the needle was retracted, forceps were used to immediately pinch the site of exit in order to ensure that the cell suspension did not escape. In all, ~1.5-15 million sclerotome cells were transplanted subcutaneously per graft.

In vivo construction of human fetal heart grafts in the mouse ear and subsequent transplantation of hESC-derived cardiac populations

Construction of human fetal heart grafts was done generally as previously described (Ardehali et al., 2013), with modifications as described below. After isolation, de-identified fetal week 15-17 human fetal hearts were transported on ice in UW organ cryopreservation solution and used for experiments within several hours of harvesting. For the purposes of this study, only two human fetal hearts from independent donors were obtained under regulatory approval from Stanford University. Using scissors, the atria were removed, and subsequently, the ventricles were dissected into fragments that were up to approximately 2mm x 7mm in area.

To construct human fetal heart grafts, NSG mice of 2-3 months of age were first anesthetized by inhaled isoflurane. The ear was shaved to the base of the skull and was disinfected. Using fine-point scissors, a small incision was made on the dorsal side of the ear close to the base where the ear meets the skull.

Subsequently a blunt-end forcep was inserted into the incision and then was gently tunneled towards the apex of the ear to create a subcutaneous pocket. Then, a human ventricle fragment (dissected as detailed above) was loaded into a trocar. The trocar was custom-produced by fitting a 16-gauge intravenous plastic catheter with a blunted plunger.

The tip of the trocar was slid into the subcutaneous pocket and then the human ventricle fragment was gently implanted using the plunger. Afterwards, the pocket was closed by gently massaging the ear proximal to the graft with a Q-tip in order to remove residual air and close the tunnel.

~1 month after implanting the human fetal heart graft, hESC-derived cardiac cells were directly injected into the graft. In brief, mice were anesthetized by isoflurane and then $1.5\text{-}2\times10^6$ *EF1A-BCL2-2A-GFP;UBC-tdTomato-2A-Luciferase* H9-derived day 3 cardiac mesoderm or day 8 cardiomyocytes were directly injected into the subcutaneous graft in a 1:1 mixture of CDM2 and Matrigel using a 31-gauge insulin syringe.

Bioluminescence imaging

To non-invasively image luciferase⁺ hESC-derived donor cells after sclerotome or cardiac lineage transplants, bioluminescence imaging was conducted. In brief, mice were injected intraperitoneally with 0.33 mg D-luciferin (in 200 µL volume of PBS) 20 minutes prior to imaging. 5 minutes before imaging, mice were anesthetized by isoflurane and placed in the IVIS Spectrum imaging chamber. A single image was captured using the following settings: autoexposure, small binning and Fstop_2. Data were subsequently analyzed using Living Image software.

Histological analysis of ectopic bone grafts

~2-3 months after subcutaneous transplantation of sclerotome progenitors, recipient mice were humanely sacrificed. Grafts were retrieved, embedded in paraffin, and sectioned. Slides with tissue sections were deparaffinized through sequential washes in histological-grade xylene (Sigma) and ethanol, and were stained with either hematoxylin & eosin (as per standard procedures) or with Russell-Movat's Pentachrome (American MasterTech, KTRMPPT).

Other differentiation protocols

Paraxial mesoderm: FLyB→FLy differentiation was conducted as previously described (Cheung et al., 2012), with 1.5 days of FGF2 (20 ng/mL) + BMP4 (10 ng/mL) + LY294002 treatment followed by 1.5 days of FGF2 (20 ng/mL) + LY294002 treatment (until day 3 of differentiation). CF→FR differentiation was conducted as previously described (Mendjan et al., 2014), with 1.5 days of CHIR99021 (8 µM) + FGF2 (20 ng/mL) treatment followed by 1.5 days of FGF2 (4 ng/mL) + RA (1 µM) (until day 3 of differentiation) (used in **Fig. 2f**).

Cardiac mesoderm: CHIR→C59 differentiation was conducted as previously described (Burridge et al., 2014), with 2 days of CHIR99021 (6 µM) treatment, followed by 2 days of C59 (2 µM) treatment, followed by treatment with basal medium alone. The basal medium used was RPMI1640 + 0.5 mg/mL albumin + 213 µg/mL 2-phospho-L-ascorbic acid (Burridge et al., 2014) (“small molecule Wnt-only protocol” used in **Fig. 5e**).

RNA extraction and reverse transcription

In general, RNA was extracted from undifferentiated or differentiated hPSC populations plated in 12-well format by lysing them with 350 µL RLT Plus Buffer per well. RNA was extracted with the RNeasy Plus Mini Kit (Qiagen) as per the manufacturer’s instructions. Generally 50-200 ng of total RNA was reverse-transcribed with the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems) to generate cDNA libraries for qPCR.

Quantitative PCR (qPCR)

Total cDNA was diluted 1:10-1:30 in H₂O and qPCR was performed with the SensiFAST SYBR Lo-ROX Kit (Bioline) with 10 µL qPCR reactions per well in a 384-well plate: each individual reaction contained 5.0 µL 2x SensiFAST SYBR qPCR Master Mix + 4.6 µL cDNA (totaling ~120 ng of cDNA) + 0.4 µL of

10 μ M primer stock (5 μ M forward + 5 μ M reverse primers). In general, gene-specific primer pairs for qPCR were tested for (1) specificity of amplicon amplification (only one peak on a dissociation curve) and (2) linearity of amplicon amplification (linear detection of gene expression in cDNA samples serially diluted seven times over two orders of magnitude, with 90-110% efficiency of amplification deemed acceptably linear). After qPCR plates were prepared by arraying sample-specific cDNAs and gene-specific primers (**Table S7**), they were sealed and briefly centrifuged (5 mins). 384-well qPCR plates and their adhesive sealing sheets were obtained from Thermo (AB1384 and AB0558, respectively). qPCR plates were run on a 7900HT Fast Real-Time PCR System (Applied Biosystems) with the following cycling parameters: initial dissociation (95 °C, 2 mins) followed by 40 cycles of amplification and SYBR signal detection (95 °C dissociation, 5 seconds; 60 °C annealing, 10 seconds; followed by 72 °C extension, 30 seconds), with a final series of steps to generate a dissociation curve at the end of each qPCR run. During qPCR data analysis, the fluorescence threshold to determine Ct values was set at the linear phase of amplification; other qPCR data analysis methods including qPCR heatmap generation using the GenePattern platform (Broad Institute) have been previously described (Loh et al., 2014).

Bulk-population RNA-seq profiling

For bulk-population RNA-seq, RNA was extracted from undifferentiated H7 hESCs (d0), H7-derived anterior primitive streak populations (day 1), H7-derived mid primitive streak populations (day 1), H7-derived lateral mesoderm (day 2), H7-derived FACS-sorted GARP⁺ cardiac mesoderm (day 3), H7-derived FACS-sorted DLL1⁺ paraxial mesoderm populations (day 2), H7-derived day 3 early somite progenitor populations (day 3), H7-derived dermomyotome populations (day 5, treated with BMP4 + CHIR99021 + Vismodegib on days 4-5), H7-derived FACS-sorted PDGFR α ⁺ sclerotome populations (day 6). Subsequently, the integrity of extracted RNA was assayed by on-chip electrophoresis (Agilent Bioanalyzer) and only samples with a high RNA integrity (RIN) value were used for RNA-seq.

Purified total RNA was reverse-transcribed into cDNA using the Ovation RNA-seq System V2 (NuGEN) and cDNA was sheared using the Covaris S2 system (duty cycle 10%, intensity 5, cycle/burst 100, total time 5 min). Sheared cDNA was cleaned up using Agencourt AMPure XP beads (Beckman Coulter). Subsequently, sheared cDNA was ligated to adapters and sequencing libraries were constructed using the NEBNext Ultra DNA Library Prep Kit (New England Biolabs) using barcoded adapters to enable multiplexing of libraries on the same sequencing lane. For each RNA-seq library, the effectiveness of adapter ligation and effective library concentration was determined by qPCR prior to loading them in multiplexed fashion onto a Next-Seq 500 (Stanford Stem Cell Institute Genomics Core) to obtain 150bp paired-end reads.

We followed the ENCODE long RNA-seq pipeline for expression quantification (<https://www.encodeproject.org/rna-seq/long-rnas/>) (The ENCODE Project Consortium, 2012).

Specifically, reads were aligned to hg38 using STAR (Dobin et al., 2013); gene-level expression was then quantified using RSEM (Li and Dewey, 2011). We used the $\log_2(\text{TPM} + 1)$ values as starting values for the analyses in this paper.

To accurately measure differential expression between cell-types (**Fig. 2e**), we used DESeq2 (Love et al., 2014) to call differentially expressed genes and calculate shrunken fold change estimates. Batch correction for this analysis was also done by DESeq2 via the inclusion of a known batch parameter in the design matrix. Each gene included in **Fig. 2e** was called as differentially expressed at a False Discovery Rate (FDR) of 0.1.

For global comparisons of gene expression levels across cell types (**Fig. 7b,c**) we first filtered out all genes where there was a difference of less than 2 (in \log_2 TPM units, i.e., a 4-fold difference in expression) between the cell types with the highest and lowest expression across all lineages under examination. Next, we used ComBat (Johnson et al., 2007) (as implemented through the *sva* R package (Leek et al., 2015)) to correct for batch effects. This sometimes left small negative values for the expression of some genes, which we set to 0. Finally, for the purposes of visualization, we normalized the expression of all genes such that for each gene, the highest expression was 1.0 (i.e., **Fig. 7c**). Collated bulk-population RNA-seq data are viewable at http://cs.stanford.edu/~zhenghao/mesoderm_gene_atlas.

Single-cell RNA-seq profiling

Cells were briefly washed (DMEM/F12), dissociated (TrypLE Express), strained (100 µm filter), pelleted and resuspended in DMEM/F12 for counting. Before single-cell capture, two quality control steps were implemented. Firstly, cell size was estimated in order to determine whether cells should be loaded onto C1 capture arrays of either 10-17 µm or 17-25 µm size. Arrays were chosen for each lineage by estimating the median cell size of each given population on a flow cytometer on the basis of the FSC-W signal (Tzur et al., 2011) and choosing an array with an appropriate pore size to accommodate such cells. Secondly, to ensure the high viability of *in vitro*-differentiated cells prior to commencing single-cell RNA-seq, for each population a separate aliquot of cells was stained with 1.1 µM DAPI and analyzed by flow cytometry; for all cell populations that were used for single-cell RNA-seq, >98% of cells were viable (i.e., DAPI negative).

For single-cell capture, cells were diluted to a concentration of 1000 cells/µL, diluted 3:2 in C1 Cell Suspension Reagent, and loaded onto a Fluidigm C1 single-cell capture array (a 10-17 µm array was used for hESCs, day 1 anterior PS, day 2 sorted DLL1⁺ paraxial mesoderm, day 2.25 somitomeres, day 3 early somites, day 3 sorted GARP⁺ cardiac mesoderm, day 5 central dermomyotome or day 6 sorted PDGFRA⁺ sclerotome while a 17-25 µm array was used for day 1 mid PS) for automated capture on a Fluidigm C1 Machine (Stanford Stem Cell Institute Genomics Core).

After cells were loaded onto C1 arrays, single-cell capture was verified on an automated microscope, followed by cell lysis, reverse transcription and cDNA synthesis on the C1 machine using the SMARTer Ultra Low RNA Kit (Clontech, 634833) generally as per the manufacturers' instructions (Fluidigm, PN 100-7168 version I1). Subsequently, the concentration and integrity of single-cell cDNA libraries was assessed by an electrophoresis-based method to assess DNA concentration and fragment size (Fragment Analyzer, Advanced Analytical). Only single-cell cDNA libraries that (1) were not degraded and (2) originated from wells that were microscopically verified to contain a single cell were carried forward for subsequent library construction.

Using a Mosquito high-throughput robotic pipetter (TTP Labtech), single-cell cDNA libraries from all lineages were approximately diluted to a concentration in the range of 0.1-0.16 ng/µL using C1 Harvest Reagent (Fluidigm) in order to normalize sample concentrations and enhance the consistency of subsequent library construction across all lineages. Then, diluted single-cell cDNA libraries were fragmented and barcoded through the use of the Nextera XT DNA Sample Prep Kit (Illumina, FC-131-1096) and then pooled and cleaned up (Agencourt AMPure XP beads) for deep sequencing, as per the manufacturers' instructions (Fluidigm, PN 100-7168 version I1) such that 384 individual single-cell RNA-seq libraries were sequenced in a single sequencing lane.

After deep sequencing (Next-Seq 500) to obtain 150bp paired-end reads, we quantified single-cell gene expression using the ENCODE long RNA-seq pipeline (described above), with an additional filtering step to ensure that we only kept cells with at least 1 million uniquely mapped reads and with at least 70% of reads uniquely mapping.

Principal component analysis (PCA) for the purpose of visualization (**Fig. 2g**, **Fig. 4d**) was done on the expression values (in log₂ TPM) of the top 500 genes by variance.

To estimate the percentage of cells that were positive for a given differentiation marker by single-cell RNA-seq (as performed for **Fig. 1**), a TPM expression threshold was set such that out of all hESC single-cell transcriptomes analyzed, no more than 1 individual hESC would be regarded positive. Differentiated cells expressing the marker above that threshold were scored as positive and the percentages of positive cells were displayed in **Fig. 1bii**, **1g**.

Temporal analysis of single-cell RNA-seq

To study the kinetics of gene expression in single cells progressing through somitogenesis, we ordered single cells from D2 paraxial mesoderm, D2.25 somitomeres, and D3 early somites along a single inferred temporal trajectory. To do this, we first filtered out genes that did not vary in these cell types by computing

the average \log_2 TPM value of each gene in each of the three cell types and only keeping genes that displayed a difference of at least 2 (in \log_2 TPM units, i.e., a 4-fold difference in expression) between the cell type with the highest and lowest expression. We conducted PCA of the 1,240 retained genes (**Fig. S3g**) and observed that the first principal component (PC1) recovered the correct bulk ordering of D2, D2.25, and D3 cells. Indeed, the three genes with the strongest loadings in PC1 were *MSGN1*, *DKK1*, and *TSPAN7*, of which the first two genes were known to be involved in somitogenesis *in vivo*. Therefore, we simply used the projection of each cell along PC1 to obtain an inferred temporal ordering of the cells along “pseudotime”, as used in **Fig. 3a,b**.

Given the inferred pseudotemporal ordering, we were then interested in clustering the genes by their expression dynamics across pseudotime during reconstituted human somitogenesis. To account for technical dropout and other sources of variation, we first fit a Loess curve to the single-cell expression levels of each gene and used that to impute gene expression values for each cell. Next, we normalized the expression levels for each gene individually by dividing by the maximum expression value for that gene over all cells, so that the resulting values would be reflective of the shape of the trajectory and not its absolute value. Finally, we ran k-medoids clustering with $k=10$ (implemented with the pam method in the R package *cluster* (Maechler et al., 2012)).

ATAC-seq

ATAC-seq was performed generally as previously described (Buenrostro et al., 2013), with minor modifications. In brief, for each replicate, 50,000 cells were lysed in lysis buffer containing 0.01% IGEPAL CA-630 (Sigma, I8896) to obtain nuclei, which were directly used in the Tn5 transposition reaction (reagents from Nextera DNA Sample Preparation Kit; Illumina, FC-121-1030). Immediately following transposition, DNA fragments were purified (MinElute Kit, Qiagen) and PCR amplified for a total of 12-13 cycles using previously-designed primer sequences that contained library barcodes (Buenrostro et al., 2013). Then, libraries were purified (MinElute Kit, Qiagen), pooled and finally concentrations were assessed (Bioanalyzer) prior to next-generation sequencing. The quality of ATAC-seq libraries was confirmed using a MiSeq v3 (Stanford Functional Genomics Facility, 2x75bp reads) before deep sequencing was performed on a NextSeq 500 (2x75bp reads). Two replicates were analyzed per cell-type.

ATAC-seq reads were aligned to hg19 using Bowtie2 (Langmead and Salzberg, 2012). Each replicate was subsampled to a maximum of 35 million uniquely-mapping reads (post-filtering for duplicate, mitochondrial, or low-quality reads) to improve comparability between samples. We then used MACS2 (Zhang et al., 2008) to call peaks for each individual replicate with a relaxed FDR threshold of 0.01. After calling peaks for each individual replicate, we created a unified peak list for each cell-type by selecting only peaks that were reproducible between both replicates. This was done through an irreproducible discovery rate (IDR) analysis (Li et al., 2011), similar to as previously described by the ENCODE Consortium (Gerstein et al., 2012). In brief, the IDR method takes in peak calls from a pair of replicates, filters out all peaks that only appear in one replicate, and then uses a copula mixture model to model the remaining peaks as belonging to either a reproducible “signal” population or an irreproducible “noise” population. We used an IDR threshold of 0.02, i.e., we only retained peaks that were deemed to have come from the “signal” population with a probability of more than 0.98. Finally, we filtered out all peaks that appeared in a curated blacklist of artifactual regions in hg19 (The ENCODE Project Consortium, 2012) (<https://sites.google.com/site/anshulkundaje/projects/blacklists>).

To obtain a universal list of peaks across all cell-types, we used BEDtools (Quinlan and Hall, 2010) to merge the lists of filtered, reproducible peaks for each cell-type. For each cell-type, we then pooled its two biological replicates together and called peaks (MACS2) on the pooled reads. To obtain a single measure of confidence at each peak P in the universal list for each cell-type C, we took the highest $-\log_{10} p$ -value out of all peaks in the pooled replicates for C that intersected with P.

ATAC-seq heatmap and motif enrichment analysis

The pipeline above resulted in a universal list of 87,215 peaks across all cell-types. To account for experimental variation and batch effects between samples, we adopted a simple binarization method in which we only called a peak “active” in a given cell-type if its $-\log_{10} p$ -value was within the top 20% for

that cell-type out of all 87,215 peaks in the universal list. As previously described (Kundaje et al., 2015), we used the *p*-value for the threshold (instead of the fold enrichment or average signal intensity). We then filtered out all peaks that were not “active” in any of the cell-types, resulting in 51,230 peaks. We visualized these 51,230 peaks in the heatmap in **Fig. 7d** after running k-medoids clustering (L1 distance; implemented with the clara method in the R package *cluster* (Maechler et al., 2012)) with *k*=17. The number of clusters was chosen by silhouette analysis (silhouette from the same *cluster* package) from *k*=2 to 20, but the heatmap and subsequent motif enrichment were qualitatively similar for different values of *k*.

To confirm that the binarization was appropriate, we subsequently used a rank-transformation to assign each peak in each cell-type a number between 0 and 1, e.g., for each cell-type, the peak with the highest $-\log_{10}$ p-value was assigned 1.0 and conversely the peak with the lowest $-\log_{10}$ p-value was assigned 0.0. We visualized the ranks of the 51,230 peaks in **Fig. S7j**, using the same clustering obtained on the binarized data. This was qualitatively similar to the binarized plot (**Fig. 7d**); for clarity, we used the binarized plot in subsequent analyses.

Clustering revealed a large cluster of ubiquitously “active” peaks, as well as many smaller clusters of lineage-specific “active” peaks. To find transcription factor motifs enriched in the latter, we ran HOMER (Heinz et al., 2010) using peaks in a given lineage-specific cluster as foreground and all peaks in the remaining 16 clusters as background. All motifs shown in **Fig. 7d** were within the four most statistically-significant motifs identified for each cluster by HOMER. In many cases, the detected motif matched a broad family of transcription factors; where possible, we used bulk RNA-seq expression data and known biological connections to identify a smaller subset of plausible factors (**Fig. 7d,e**). To verify these results, we also ran SeqGL (Setty and Leslie, 2015) on the clusters. In brief, SeqGL is a discriminative motif discovery algorithm that compares peaks to flanking regions and builds a *k*-mer model that can classify peaks vs. flanks; SeqGL relies on HOMER in the backend to generate motifs for visualization and comparison to known motifs. The results obtained with SeqGL (data not shown) matched with what we obtained from direct application of HOMER; thus results from the latter were shown for simplicity.

To identify potential FOX motifs in the two putative *MEOX1* enhancers (**Fig. 7f**), we used PWMscan (<http://ccg.vital-it.ch/pwmtools>) using the motif from the somite-specific cluster that HOMER had associated with the FOX family. PWMscan identified two potential FOX binding sites, one in each enhancer, at an uncorrected *p*-value cutoff of 0.001.

Regulatory and institutional review

All animal experiments, including those involving de-identified human fetal tissue, were conducted pursuant to experimental protocols approved by the Stanford Administrative Panel on Laboratory Animal Care (APLAC). All human pluripotent stem cell experiments were conducted in accord with experimental protocols approved by the Stanford Stem Cell Research Oversight (SCRO) committee.

Supplemental Tables

Table S1: Summary of major developmental signaling pathway modulators used in this study; related to Experimental Procedures

Item name	Description	Company	Catalog no.
2-Phospho-L-ascorbic acid trisodium salt	Stabilized Vitamin C analog	Sigma	49752-10G
21K	Hedgehog agonist	Commercially synthesized	N/A
A-83-01	TGF- β inhibitor	Tocris	2939
Activin A	TGF- β agonist	R&D Systems	338-AC-050
AZD4547	FGFR inhibitor	Cellagen Technology	C2454-5S
BGJ398	FGFR inhibitor	Selleckchem	S2183
BMP4	BMP agonist	R&D Systems	314-BP-050
C59	WNT inhibitor	Cellagen Technology	C7641-2S

CHIR99021	WNT agonist	Tocris	4423
DKK1	WNT inhibitor	R&D Systems	5439-DK-010
DMH1	BMP inhibitor	Tocris	4126
FGF2	FGF agonist	R&D Systems	233-FB-01M
GDC-0449 (Vismodegib)	Hedgehog inhibitor	Cellagen Technology	C4044-5
IWR1	WNT inhibitor	Sigma	I0161-5MG
LDN193189 (DM3189)	BMP inhibitor	Stemgent	04-0074
PD0325901	ERK inhibitor	Tocris	4192
PD173074	FGF inhibitor	Tocris	3044
PDGF-BB	PDGF agonist	R&D Systems	220-BB-010
PIK90	PI3K inhibitor	EMD Millipore	528117-5MG
Retinoic acid	Retinoid agonist	Sigma	R2625-50MG
SB-505124	TGF-β inhibitor	Tocris	3263
TGF-β1	TGF-β agonist	R&D Systems	240-B-010
Thiazovivin	ROCK inhibitor	Tocris	3845
WNT3A	WNT agonist	R&D Systems	5036-WN-010
XAV939	WNT inhibitor	Tocris	3748

Agonists and inhibitors of major developmental signaling pathways employed in this study, including commercial suppliers and catalog numbers where relevant.

Table S5: Antibodies for immunostaining; related to Experimental Procedures

Antibody	Company	Catalog Number	Dilution
BRACHYURY	R&D Systems	AF2085	1:250
CDX2	BioGenex	CDX2-88	1:200
COL2A1	Thermo	MA1-37493	1:100
CONNEXIN43/CX43	Abcam	ab11370	1:100
FOXC2	Abcam	ab5060	1:100
HAND1	R&D Systems	AF3168	10 µg/mL
HOPX	Santa Cruz Biotech	sc-30216	1:500
MIXL1 (clone 6G2)	Courtesy of D. Elliot	N/A	1:50
MYH3	Novocastra	NCL-MHCd	1:20
TBX6	R&D Systems	AF4744	1:40
TROPONIN/TNNT2	Thermo	MS-295-P	1:200

Table S6: Antibodies for flow cytometry; related to Experimental Procedures

Antibody	Company	Catalog Number	Dilution
CD13 APC	eBioscience	8046-0138-025	1:5
DLL1 PE	Biolegend	346404	1:5
GARP APC	Biolegend	352505	1:20
PDGFRα PE	BD Biosciences	556002	1:20
ROR2 APC	R&D Systems	FAB2064A	1:10
SIRPα APC	Biolegend	323810	1:50
SMAα PE	R&D Systems	IC1420P	1:10

Table S7: Primers for quantitative PCR; related to Experimental Procedures

Gene Name	Forward	Reverse
<i>ACTA1</i>	CGACATCAGGAAGGACCTGTAT GCC	GGCCTCGTCGTACTCCTGCTTGG
<i>ACTA2</i>	CTATGAGGGCTATGCCTGCC	GCTCAGCAGTAGTAACGAAGGA
<i>AGGRECAN</i>	CCCCTGCTATTCATCGACCC	GACACACGGCTCCACTTGAT
<i>ALX4</i>	ATGAATGCTGAGACTTGCCTC	GGGAAATGCCCTAAAAGGCG
<i>BAPX1/NKX3.2</i>	GA TTTCAGGCCTGCTGGGA	TTTCGCACCCCTGGTTACA
<i>BRACHYURY</i>	TGCTTCCCTGAGACCCAGTT	GATCACTTCTTCCTTGATCAA G
<i>CD144/VECAD</i>	AACGAGCAGGGCGAGTTCACCT TC	TAGGTGACCAGCTGCTCGTGGAT C
<i>CD31/PECAMI</i>	AACAGTGTGACATGAAGAGCC	TGTAAAACAGCACGTACATCCTT
<i>CD34</i>	TGGCTGTCTGGCATCACTGG	CTGAATGGCGTTCTGGAGGTG G
<i>CDX2</i>	GGGCTCTCTGAGAGGCAGGT	CCTTGCTCTGCGGTTCTG
<i>CNN1</i>	GTCAACCCAAAATTGGCACCA	ACCTTGTTCCTTCGTCTTCG
<i>COL2A1</i>	CCAGATGACCTTCCTACGCC	TTCAGGGCAGTGTACGTGAAC
<i>COMP</i>	GATCACGTTCTGAAAAACACG	GCTCTCCGTCTGGATGCAG
<i>DLL1</i>	ACTCCCGCTTCAGCAACCCAT CAACATAAACGGACTCAATCCC A	TGGGTTTCTGTTGCGAGGTCACTC AGG
<i>EOMES</i>		ACCACCTCTACGAACACATTGT
<i>EPIPHYCAN</i>	AGGAGGGAGGAATCTACTCCA	CAGCGGAGGAATAGCATCAAG
<i>EVX1</i>	AGTGACCAGATGCGTCGTTAC	TGGTTTCCGGCAGGTTTAG
<i>FLK1</i>	TTTTGCCCTTGTCTGTCC	TCATTGTTCCCAGCATTCA
<i>FOXA2</i>	GGGAGCGGTGAAGATGGA	TCATGTTGCTCACGGAGGAGTA
<i>FOXC2</i>	CCTCCTGGTATCTCAACCACA1	GAGGGTCGAGTTCTCAATCCC
<i>FOXF1</i>	AGCAGCCGTATCTGACCAGAA	CTCCTTCGGTCACACATGCTG
<i>FZD8</i>	ATCGGCTACAACACACCTACA	GTACATGCTGCACAGGAAGAA
<i>GATA4</i>	TCCCTCTCCCTCCTCAAAT	TCAGCGTGTAAAGGCATCTG
<i>GATA6</i>	CCCACAAACACAACCTACAGC	GCGAGACTGACGCCTATGTA
<i>GSC</i>	GAGGAGAAAGTGGAGGTCTGG TT	CTCTGATGAGGACCGCTTCTG
<i>HAND1</i>	GTGCGCTCTTAATCCTCTTC	GTGAGAGCAAGCGGAAAAG
<i>HHEX</i>	CACCCGACGCCCTTTACAT	GAAGGCTGGATGGATCGGC
<i>HOPX</i>	GACAAGCACCGGATTCCA	GTCTGTGACGGATCTGCACTC
<i>HOXB5</i>	AACTCCTCTCGGGCGTTAT	CATCCCATTGTAATTGTAGCCGT
<i>ISL1</i>	AGATTATATCAGGTTGTACGGG ATCA	ACACAGCGAACACTCGAT
<i>MEF2C</i>	ATGGATGAACGTAACAGACAG GT	CGGCTCGTTGACTCCGTG
<i>MEOXI</i>	TCTGAGCGCCAGGTCAAAG	CTGAACCTGGAGAGGCTGTGG
<i>MEOX2</i>	GTGGCGGCTACAAGGTATC	CTGGCGCGAACATAAACAA
<i>MESP1</i>	GAAGTGGTCTGGCAGAC	TCCTGCTGCCTCAAAGTGT
<i>MESP2</i>	AGCTTGGGTGCCTCCTTATT	TGCTTCCCTGAAAGACATCA

<i>MIXL1</i>	GGTACCCGACATCCACTTG	TAATCTCCGGCTAGCCAAA
<i>MSGN1</i>	CGGAATTACCTGCCACCTGT	GGTCTGTAGTTCCCCGATG
<i>MYF5</i>	GCCTGAAGAAGGTCAACCAG	CCATCAGAGCAGTGGAGGT
<i>MYH3</i>	CTGGAGGATGAATGCTCAGAGC	CCCAGAGAGTCCTCAGTAAGG
<i>MYH6</i>	GCCCTTGACATTGCACTG	GGTTTCAGCAATGACCTTGCC
<i>MYH7</i>	TCGTGCCTGATGACAAACAGGA GT	ATACTCGGTCTCGGCAGTGACTT T
<i>MYL7</i>	ACATCATCACCCATGGAGACGA GA	GCAACAGAGTTATTGAGGTGCC C
<i>MYOD1</i>	TGCCACAACGGACGACTT	CGGGTCCAGGCTTCGAA
<i>MYOG</i>	AGATGTGTCTGTGGCCTTCC	AGCTGGCTTCCTAGCATCAG
<i>NKX2.5</i>	CAAGTGTGCGTCTGCCTTT	CAGCTCTTCTTTCGGCTCTA
<i>NKX3.1</i>	CCAGCTCAGGTGACAACCAT	CTTGGCCCCTTGTGCTTTTC
<i>OCT4/POU5F1</i>	AGTGAGAGGCAACCTGGAGA	ACACTCGGACCACATCCTTC
<i>ODD1</i>	CAGCTCACCAACTACTCCTTC TTCA	TGCAACCGCCTGAAACCATA
<i>PARAXIS</i>	GAGCTGAGGAGAGTCCCCTG	TGTGCCTCTCTCTAGGTCCA
<i>PAX1</i>	CGCTATGGAGCAGACGTATGGC GA	AATGCGCAAGCGGATGGCGTTG
<i>PAX3</i>	CTCCACGCTCCGGATAGTTC	ATCTTGTGGCGGATGTGGTT
<i>PAX6</i>	GCAGATGCAAAAGTCCAGGTG	CAGGTTGCGAAGAACTCTGTTT
<i>PAX7</i>	TCCAAGATTCTTGCCGCTAC	GGTCACAGTGCCCATCCTTC
<i>PAX9</i>	TGGTTATGTTGCTGGACATGGG TG	GGAAGCCGTGACAGAATGACTAC CT
<i>PDGFRB</i>	GGTGGGCACACTACAATTGC	GGTGGGTAGGCCTCGAAC
<i>PRRX1</i>	TGATGCTTTGTGCGAGAAGA	AGGGAAAGCGTTTTATTGGCT
<i>SOX17</i>	CGCACGGAATTGAACAGTA	GGATCAGGGACCTGTCACAC
<i>SOX2</i>	TGGACAGTTACGCGCACAT	CGAGTAGGACATGCTGTAGGT
<i>SOX9</i>	CGTCAACGGCTCCAGCAAGAAC AA	GCCGCTTCTCGCTCTCGTTAGA AGT
<i>STMN</i>	CGGCCTGCGCGTGTCTAATCC	CTGTGACCTCCAGCAGCTCCGA A
<i>TAGLN</i>	AGTGCAGTCCAAAATCGAGAA G	CTTGCTCAGAATCACGCCAT
<i>TBX20</i>	GGCGACGGAGAACACAATCAA	CTGGGCACAGGACGACTTC
<i>TBX5</i>	TACCACCAACCCATCAAC	ACACCAAGACAGGGACAGAC
<i>TBX6</i>	AAGTACCAACCCCGCATACA	TAGGCTGTCACGGAGATGAA
<i>TNNT2</i>	GGAGGAGTCAAACCAAAGCC	TCAAAGTCCACTCTCTCCATC
<i>TWIST1</i>	CTGCAGCACCGGCACCGTT	CCCAACGGCTGGACGCACAC
<i>UNCX4.1</i>	CTATCCCGACGTGTTCATGC	GAACCTGGGACTCGACCAG

Supplemental References

- Ardehali, R., Ali, S.R., Inlay, M.A., Abilez, O.J., Chen, M.Q., Blaukamp, T.A., Yazawa, M., Gong, Y., Nusse, R., Drukker, M., *et al.* (2013). Prospective isolation of human embryonic stem cell-derived cardiovascular progenitors that integrate into human fetal heart tissue. Proc Natl Acad Sci USA.
- Ardehali, R., Inlay, M.A., Ali, S.R., Tang, C., Drukker, M., and Weissman, I.L. (2011). Overexpression of BCL2 enhances survival of human embryonic stem cells during stress and obviates the requirement for serum factors. Proc Natl Acad Sci USA.

- Brunton, S.A., Stibbard, J.H.A., Rubin, L.L., Guicherit, O.M., Kruse, L.I., Price, S., di Lucrezia, R., Mackinnon, C.H., Avery, A., Park, Y., *et al.* (2009). Potent agonists of the Hedgehog signaling pathway. *Bioorganic & Medicinal Chemistry Letters* *19*, 4308-4311.
- Burridge, P.W., Matsa, E., Shukla, P., Lin, Z.C., Churko, J.M., Ebert, A.D., Lan, F., Diecke, S., Huber, B., Mordwinkin, N.M., *et al.* (2014). Chemically defined generation of human cardiomyocytes. *Nature Methods*.
- Cheung, C., Bernardo, A.S., Trotter, M.W.B., Pedersen, R.A., and Sinha, S. (2012). Generation of human vascular smooth muscle subtypes provides insight into embryological origin-dependent disease susceptibility. *Nat Biotechnol*.
- Consortium, T.E.P. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* *488*, 57-74.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* *29*, 15-21.
- Gerstein, M.B., Kundaje, A.B., Hariharan, M., Landt, S.G., Yan, K.-K., Cheng, C., Mu, X.J., Khurana, E., Rozowsky, J., Alexander, R., *et al.* (2012). Architecture of the human regulatory network derived from ENCODE data. *Nature* *488*, 91-100.
- 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.
- Jain, R., Li, D., Gupta, M., Manderfield, L.J., Ifkovits, J.L., Wang, Q., Liu, F., Liu, Y., Poleshko, A., Padmanabhan, A., *et al.* (2015). Integration of Bmp and Wnt signaling by Hopx specifies commitment of cardiomyoblasts. *Science* *348*, aaa6071.
- Johnson, W.E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* (Oxford, England) *8*, 118-127.
- Kartopawiro, J., Bower, N.I., Karnezis, T., Kazenwadel, J., Betterman, K.L., Lesieur, E., Koltowska, K., Astin, J., Crosier, P., Vermeren, S., *et al.* (2014). Arap3 is dysregulated in a mouse model of hypotrichosis-lymphedema-telangiectasia and regulates lymphatic vascular development. *Hum Mol Genet* *23*, 1286-1297.
- Kundaje, A.B., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J., *et al.* (2015). Integrative analysis of 111 reference human epigenomes. *Nature* *518*, 317-330.
- Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods* *9*, 357-359.
- Leek, J., Johnson, W., Parker, H., Fertig, E., Jaffe, A., and Storey, J. (2015). sva: Surrogate Variable Analysis. R package version 3.18.0. .
- Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* *12*, 323.
- Li, Q., Brown, J.B., Huang, H., and Bickel, P.J. (2011). Measuring reproducibility of high-throughput experiments. *The Annals of Applied Statistics* *5*, 1752-1779.
- Loh, K.M., Ang, L.T., Zhang, J., Kumar, V., Ang, J., Auyeong, J.Q., Lee, K.L., Choo, S.H., Lim, C.Y.Y., Nichane, M., *et al.* (2014). Efficient Endoderm Induction from Human Pluripotent Stem Cells by Logically Directing Signals Controlling Lineage Bifurcations. *Cell Stem Cell* *14*, 237-252.
- Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* *15*, 550.
- Madisen, L., Zwingman, T.A., Sunkin, S.M., Oh, S.W., Zariwala, H.A., Gu, H., Ng, L.L., Palmiter, R.D., Hawrylycz, M.J., Jones, A.R., *et al.* (2010). A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nature Neuroscience* *13*, 133-140.
- Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., and Hornik, K. (2012). Cluster: cluster analysis basics and extensions. . R package version 1.
- Mendjan, S., Mascetti, V.L., Ortmann, D., Ortiz, M., Karjosukarso, D.W., Ng, Y., Moreau, T., and Pedersen, R.A. (2014). NANOG and CDX2 Pattern Distinct Subtypes of Human Mesoderm during Exit from Pluripotency. *Cell Stem Cell*.
- Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* *26*, 841-842.
- Setty, M., and Leslie, C.S. (2015). SeqGL Identifies Context-Dependent Binding Signals in Genome-Wide Regulatory Element Maps. *PLoS Computational Biology* *11*, e1004271.

- Thisse, C., and Thisse, B. (2008). High-resolution *in situ* hybridization to whole-mount zebrafish embryos. *Nat Protoc* 3, 59-69.
- Tzur, A., Moore, J.K., Jorgensen, P., Shapiro, H.M., and Kirschner, M.W. (2011). Optimizing optical flow cytometry for cell volume-based sorting and analysis. *PLoS ONE* 6, e16053.
- Xu, C., Tabebordbar, M., Iovino, S., Ciarlo, C., Liu, J., Castiglioni, A., Price, E., Liu, M., Barton, E.R., Kahn, C.R., *et al.* (2013). A zebrafish embryo culture system defines factors that promote vertebrate myogenesis across species. *Cell* 155, 909-921.
- Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., *et al.* (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9, R137.