

# Periodic formation of epithelial somites from human pluripotent stem cells

Marina Sanaki-Matsumiya <sup>1</sup>✉, Mitsuhiro Matsuda<sup>1</sup>, Nicola Gritti <sup>1</sup>, Fumio Nakaki <sup>1</sup>, James Sharpe<sup>1</sup>, Vikas Trivedi <sup>1,2</sup> & Miki Ebisuya <sup>1</sup>✉

During embryonic development, epithelial cell blocks called somites are periodically formed according to the segmentation clock, becoming the foundation for the segmental pattern of the vertebral column. The process of somitogenesis has recently been recapitulated with murine and human pluripotent stem cells. However, an in vitro model for human somitogenesis coupled with the segmentation clock and epithelialization is still missing. Here, we report the generation of human somitoids, organoids that periodically form pairs of epithelial somite-like structures. Somitoids display clear oscillations of the segmentation clock that coincide with the segmentation of the presomitic mesoderm. The resulting somites show anterior-posterior and apical-basal polarities. Matrigel is essential for epithelialization but dispensable for the differentiation into somite cells. The size of somites is rather constant, irrespective of the initial cell number. The amount of WNT signaling instructs the proportion of mesodermal lineages in somitoids. Somitoids provide a novel platform to study human somitogenesis.

<sup>1</sup>European Molecular Biology Laboratory (EMBL) Barcelona, Barcelona, Spain. <sup>2</sup>EMBL Heidelberg, Developmental Biology Unit, Heidelberg, Germany.  
✉email: [marina.matsumiya@embl.es](mailto:marina.matsumiya@embl.es); [miki.ebisuya@embl.es](mailto:miki.ebisuya@embl.es)

Somites are transient blocks of cells that give rise to a variety of tissues, including the vertebrae, rib cage, skeletal muscle, and part of the skin, in vertebrate embryos<sup>1–3</sup>. Bilateral pairs of somites periodically bud off from the presomitic mesoderm (PSM) along the anterior-posterior axis: as the mesenchymal cells migrate from the posterior PSM region near the tailbud to the anterior PSM region, they undergo a mesenchymal-epithelial transition (MET), acquiring the apical-basal polarity and elongated shapes. The anterior PSM cells eventually form spherical epithelial somites surrounding a core of mesenchymal cells. The somite formation, or somitogenesis, starts around day 20 after fertilization (Carnegie stage 9) in human embryos, and a total of ~40 pairs of somites are formed<sup>4</sup>. The timing of sequential somitogenesis is controlled by the segmentation clock, a molecular oscillator that peaks its activity every 5–6 hours in humans<sup>2,5–9</sup>.

Several aspects of somitogenesis have recently been recapitulated in vitro with pluripotent stem cells. PSM-like flat tissues made from mouse embryonic stem cells (ESCs) display the segmentation clock accompanied by tissue boundary formation<sup>10</sup>. Gastruloids are ESC-derived embryonic organoids that mimic early developmental events, including three-germ layer differentiation and axis patterning<sup>11–13</sup>, and mouse gastruloids embedded in an extracellular matrix (ECM) surrogate, Matrigel, form a string of single somites<sup>14</sup>. Mouse ESC-derived trunk-like structures (TLSs) form both the neural tube and bilateral somites<sup>15</sup>. Regarding human somitogenesis, several groups have induced PSM and somite cell fates from human pluripotent stem cells in 2D cultures and recapitulated the segmentation clock<sup>6–9,16–18</sup>. However, somites or epithelial structures are not formed in those human stem cell-derived models. Although 3D human ‘somitoids’ have recently been reported, the somite-like structures are not sequentially formed<sup>19</sup>. Thus, there is no in vitro model of human somitogenesis so far that recapitulates both the periodic somite formation coupled with the segmentation clock and the maturation into epithelial structures.

In this study, we report human embryonic organoids that periodically form pairs of somite-like structures with the mature epithelial organization along the anterior-posterior axis.

## Results

**Human iPSCs self-organize into somite-like structures.** To make human organoids that form somite-like structures (hereafter referred to as somitoids), we first made aggregates of human induced pluripotent stem cells (iPSCs) in a low attachment U-bottom plate (Fig. 1a). The aggregates were treated for 2 days with a cocktail of signaling molecules and inhibitors that has been used previously to induce the human PSM cell fate in 2D culture conditions<sup>7,8,17</sup>. The cocktail comprises CHIR99021 (a WNT signaling activator through the inhibition of GSK3 $\beta$ ), bFGF, SB431542 (a TGF $\beta$  signaling inhibitor), and DMH1 (a BMP signaling inhibitor), mimicking WNT and FGF activation as well as TGF $\beta$  and BMP inhibition seen in the presumptive PSM region of mouse and chick embryos<sup>17</sup>. After 2-day treatment, the cocktail was gradually diluted by medium changes. The aggregates became oval-shaped around days 3–4 and then further elongated, reminiscent of gastruloids and TLSs<sup>11–15</sup> (Fig. 1b; Supplementary Fig. 1a). Matrigel was added at a 10% concentration on day 4, and the somite-like ball structures first appeared around days 4–5 (Supplementary Fig. 1a). Approximately 10 (pairs of) somites were found per somitoid on day 7 (Fig. 1c–e).

Time-lapse imaging revealed that these somites were sequentially formed (Fig. 1d; Supplementary Movie 1). Either a single somite (12.4%) or a pair of somites (77.6%) was formed at one time, and occasionally (6.2%), the paired somites were completely separated (Fig. 1c; Supplementary Movie 1). Although most

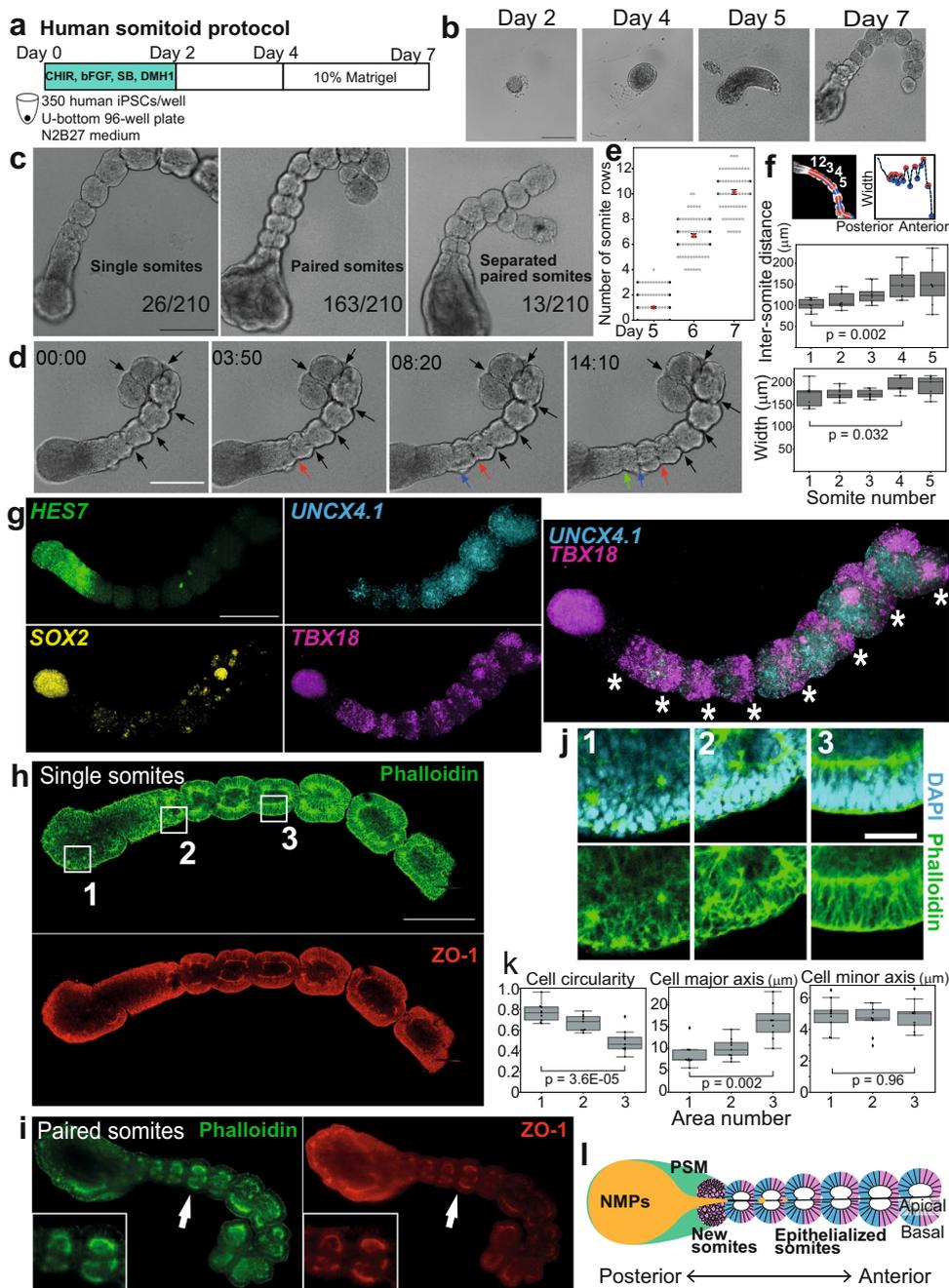
somitoids showed a single anterior-posterior axis, 3.8% of samples showed multiple axes (Supplementary Fig. 1b, c). Overall, the human somitoid protocol was reproducible: 210/210 samples, cultured over different batches, formed at least 3 (pairs of) somites by day 7 (Supplementary Fig. 1b).

The average length (along the anterior-posterior axis) and width of newly formed single somites were 116  $\mu$ m and 157  $\mu$ m, respectively, while those of paired somites, measured as individual somites, were 110  $\mu$ m and 104  $\mu$ m, respectively (Supplementary Fig. 2a). Namely, the somites formed as pairs were smaller than the single somites. Even though precisely measuring the size of somites of human embryos is challenging due to the limited resources, the somites of somitoids were in a comparable size range of in vivo human somites or in vitro mouse somite-like structures<sup>15,20</sup> (Supplementary Fig. 2). The somite width and the somite-to-somite distance showed a gradual increase from the posterior (newly formed) somites to the anterior (older) somites (Fig. 1f), as reported in chicken embryos and mouse gastruloids<sup>14,21</sup>. The older somites also tended to be more spherical (Supplementary Fig. 3; see Fig. 1c, for example), possibly due to dominant tissue surface tension in the absence of constraints from neighboring tissues<sup>15,22,23</sup>.

**Somitoids possess proper polarities.** To molecularly characterize somitoids, we examined lineage marker expressions with qRT-PCR analyses (Supplementary Fig. 4). The expression levels of pluripotency markers, such as *NANOG*, *OCT4/POU5F1*, and *SOX2*, dramatically declined by day 4. The *SOX2* expression recovered after the drop, probably because it is also expressed in neuromesodermal progenitors (NMPs)<sup>24–26</sup>. By contrast, a mesodermal marker *BRACHYURY* (also known as *TBXT*, expressed both in mesoderm and NMPs) showed an expression peak on day 4. PSM markers *TBX6* and *HES7*<sup>27</sup> as well as a segmentation maker *MESP2*<sup>28</sup> also peaked on days 4–5. Somite markers *UNCX4.1* (also known as *UNCX*)<sup>1</sup>, *TCF15*<sup>16</sup>, and *FOXC2*<sup>16</sup>, showed higher expressions after day 5.

To further examine the spatial gene expression patterns, we visualized the marker genes with immunostaining and in situ hybridization chain reaction (HCR)<sup>29</sup>. The co-expression of *BRACHYURY* (a mesodermal marker) and *SOX2* (a neural marker) in somitoids indicates the existence of NMPs, bipotent progenitor cells that give rise to both mesodermal and neural lineages<sup>24–26</sup> (Supplementary Fig. 5). The NMP region at the posterior end of somitoids was surrounded by PSM marker (*TBX6* and *HES7*)-positive cells, and the PSM region extended slightly more anteriorly (Supplementary Fig. 5; Fig. 1g). By contrast, a somite marker *UNCX4.1* was expressed in somite-like structures in the anterior regions of somitoids (Fig. 1g). The boundary between *HES7* and *UNCX4.1* corresponds to the newest somite formed from PSM. *TBX18*, a marker for the rostral halves of somites<sup>1</sup>, showed stripe patterns in somites as reported in the somite-like structures of mouse gastruloids<sup>14</sup>. The alternating patterns of *UNCX4.1* (a marker for the caudal halves of somites) and *TBX18* indicate the rostral-caudal patterning in individual somites<sup>1</sup> (Fig. 1g, right; Supplementary Fig. 6; Supplementary Movie 2), even though the patterns in somitoids are less clear than the in vivo counterparts<sup>30</sup>. These results demonstrate that the NMPs, PSM, and somites are properly located along the posterior-anterior axis of somitoids and that each somite has rostral and caudal compartments.

To examine the maturity of somites, we stained markers for epithelialization<sup>31</sup>. A tight junction marker ZO-1 showed clear localization to the inner surface of somites (Fig. 1h, i), demonstrating the establishment of the apical-basal polarity in somite cells and the formation of the apical lumen in somites. The F-actin staining with Phalloidin also showed a stronger signal on



**Fig. 1** Generation and characterization of human somitoids. **a** Somitoid protocol. 350 human iPSCs were aggregated in a U-bottom plate and treated with CHIR99021, bFGF, SB431542, and DMH1. The cocktail was washed out on day 2, and Matrigel was added on day 4. **b** Time-course images of somitoid development. **c** Classification of morphologies of day 7 somitoids.  $N = 210$  from 12 independent experiments. Only the images with entire somitoid structures were used. **d** Somite formation on day 6. Arrows indicate the somite boundaries. **e** Quantification of the number of somite rows. Mean  $\pm$  SEM.  $N = 284$  (day 5), 113 (day 6), and 77 (day 7). 3–15 independent experiments. **f** The images of day 7 somitoids were segmented, and the width perpendicular to the detected posterior–anterior axis was plotted (top). The somite-to-somite distance and the somite width (middle and bottom). Only somitoids with single somites were measured. Boxplots show median, 75th and 25th percentiles, and max and min.  $N = 9$  (somite number 1–3), 8 (somite number 4), and 7 (somite number 5). P-values are from two-sided paired t-test. **g** HCR images of day 6 somitoids. Asterisks indicate the stripes of *TBX18* expression.  $N = 8$  samples showed similar expression patterns. **h** IHC images of day 6 somitoid that displayed mostly single somites.  $N = 10$  samples showed similar expression patterns. **i** IHC images of day 7 somitoids that displayed typical paired somites.  $N = 5$  samples showed similar expression patterns. Insets: Enlarged images of the regions indicated by the arrows. **j** Enlarged images of the boxed regions in **h**. **k** Cell morphometry of Phalloidin images in **j**. Boxplots show median, 75th and 25th percentiles, and max and min except for outliers.  $N$  of measured cells = 9 (Area 1), 9 (Area 2), and 9 (Area 3). P-values are from two-sided student’s t-test. **l** Schematic diagram of a human somitoid. NMPs: Neuromesodermal progenitors, PSM: Presomitic mesoderm. Scale bars: 300  $\mu\text{m}$  (**b, c, d, g, h, i**) and 50  $\mu\text{m}$  (inset of **j, i**). Microscopes: Opera (**b, c, d**) FV3000 confocal (**h, j**), and Light-sheet (**g, i**). Source data are provided as a Source Data file.

the apical surface of somites (Fig. 1h, i) and demarcated the elongated shapes of epithelial somite cells (Fig. 1j, k, Area 3). The nuclei aligned along the basal surface of mature somites (Fig. 1j, Area 3). By contrast, the F-actin and nuclear stainings showed more random patterns in the PSM/NMP regions, and those immature mesenchymal cells had rounded morphologies (Fig. 1j, k, Area 1). These results demonstrate that a MET happens while PSM cells differentiate into somite cells along the posterior-anterior axis of somitoids and that the formed epithelial somites are mature enough to establish the apical-basal polarity.

Between paired epithelial somites, we sometimes observed a line of cells (Supplementary Fig. 7a). However, those cells did not display any continuous structure that resembles a neural tube or notochord. Some cells at the center were BRACHYURY and SOX2 double-positive (Supplementary Fig. 7b; Supplementary Movie 3), suggesting that they are NMPs. Thus, we concluded that somitoids display transcriptional and morphological hallmarks of somitogenesis but not those of neural tube or notochord formation and that the formed somites possess organized epithelial structures (Fig. 1l). This conclusion supports the idea that human somitogenesis is largely a tissue-autonomous process that does not require adjacent tissues, including the structured neural tube and notochord, consistent with previous results of mouse explant cultures and gastruloids<sup>14,32</sup>.

**Somitoids contain NMP, PSM, and somite cell populations.** To systematically characterize the cell types consisting of somitoids, we performed single-cell RNA sequencing (scRNA-seq) for day 7 somitoids (Fig. 2; Supplementary Figs. 8–9). Clustering analyses detected populations corresponding to putative NMP, PSM, somite, and late somite cells (Fig. 2a, b; Supplementary Fig. 8d, Sample1). The RNA velocity suggested the differentiation of somite cells from PSM cells (Supplementary Fig. 9a). By contrast, the differentiation of PSM cells from NMPs was not clear (Supplementary Fig. 9a), possibly because most PSM differentiation might happen in the early stages of somitoids in the presence of the cocktail of signaling molecules and inhibitors. Consistent with the lack of morphological features of a neural tube or notochord in somitoids, a clear neural, notochord-, or endoderm-like cell population was not detected (Fig. 2a, c; Supplementary Fig. 9b). In living embryos, the ventral and dorsal portions of somites further become the sclerotome and dermomyotome, respectively<sup>1,3</sup>. A couple of sclerotome markers, including *TWIST1* and *PAX9*, and dermomyotome markers, including *DMRT2* and *PAX3*, were expressed in the late somite cell population of somitoids<sup>15</sup> (Fig. 2c, d; Supplementary Fig. 9e). The sclerotome and dermomyotome markers tended to show mutually exclusive patterns (Fig. 2d). These results suggest that somites in somitoids may have started differentiating into sclerotome and dermomyotome cells, even though somitoids do not show morphological features of sclerotomes and dermomyotomes.

We further characterized the developmental stage and cell proliferation patterns of somitoids by using expression patterns of *HOX* gene clusters and *PCNA*, respectively. The PSM, somite, and late somite cells in somitoids showed the continuous expression of *HOX* genes until *HOX9-10*, suggesting that somitoids mostly recapitulate thoracic and lumbar somitogenesis (Supplementary Fig. 9d). The cell proliferation marker *PCNA* was expressed in most regions of somitoids, but the expression was weaker in late somite cells (Supplementary Fig. 9c). Consistent with this, EdU incorporation into somitoids demonstrated that cells still proliferate in the regions of NMP, PSM, and the first 1–2 somites, while cell proliferation is not active in older somites (Supplementary Fig. 10).

**Somite formation is coupled with the segmentation clock.** Multiple somite-like structures can be induced even in the

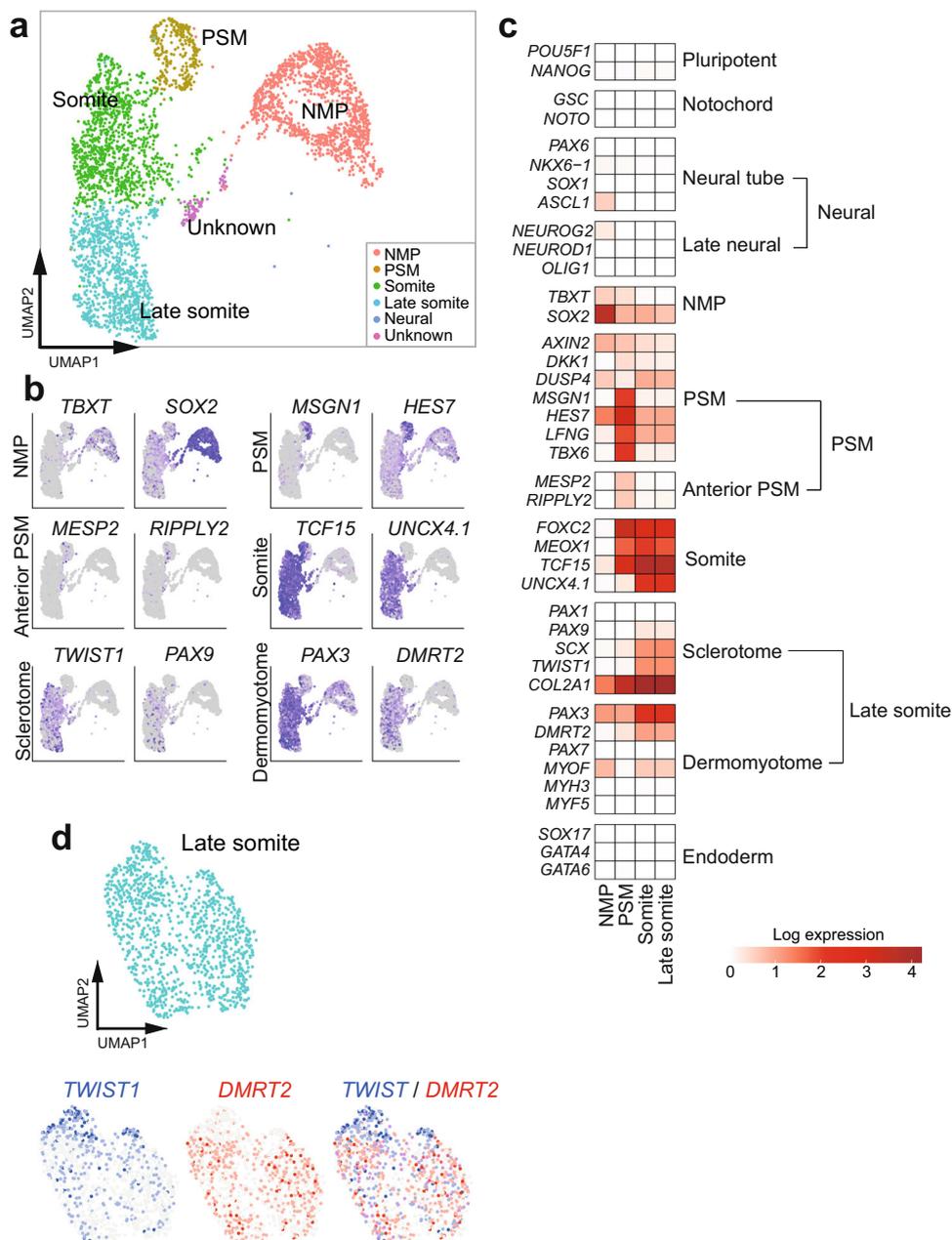
absence of the segmentation clock when signaling pathways are modulated<sup>33,34</sup>. To test whether the somitogenesis in somitoids is regulated by the segmentation clock, we monitored the expression patterns of *HES7*, a core gene of the molecular oscillator<sup>27,35</sup>. The *HES7* promoter-luciferase reporter<sup>7,8</sup> exhibited clear collective oscillations immediately after the medium change on day 2 (Supplementary Fig. 11a). Postponing the medium change postponed the onset of collective oscillations, suggesting that the medium change entrains the oscillation phases of individual cells (Supplementary Fig. 11a). The *HES7* reporter also showed traveling waves from the posterior PSM to the anterior PSM of somitoids (Fig. 3a, b; Supplementary Movie 4). The *HES7* waves were easier to be detected on day 4 when somitogenesis was about to start (Fig. 3a, c): the *HES7* oscillations in the posterior PSM preceded those in the anterior PSM, demonstrating traveling waves (Fig. 3c, d; Supplementary Fig. 11b). The oscillations and waves were disrupted by treatment with a NOTCH signaling inhibitor DAPT (Supplementary Fig. 12), consistent with the *in vivo* segmentation clock<sup>2,3,36</sup>.

On day 6, when somitogenesis was actively happening, the waves became less visible, but the oscillations were still clear (Fig. 3b, e). Importantly, the timings of *HES7* oscillations coincided with those of somite formation: one somite or one pair of somites was produced during one oscillation cycle (Fig. 3e). The period of *HES7* oscillations was ~5 hours (Fig. 3f), a typical period of the human segmentation clock<sup>2,5–9</sup>. The period calculated from the somite formation timings was also ~5 hours (Fig. 3f). These results indicate that the somite-like structures in somitoids are periodically formed according to the human segmentation clock.

**Matrigel is crucial for somite formation.** We next explored the conditions to generate successful somitoids. Matrigel is a mixture of ECM components, and Matrigel embedding has been demonstrated to be crucial for somite formation in mouse stem cell-derived models<sup>14,15</sup>. The first somites were formed in somitoids ~10 hours after Matrigel addition on day 4 (Supplementary Fig. 13a, b). When Matrigel was added 4 hours later than the usual timing, the first somites were formed ~4 hours later than usual, suggesting that Matrigel addition is the trigger of the first somite formation (Supplementary Fig. 13a, b). In both conditions, the first somites were formed at the ~20% position of somitoids along their midline (Supplementary Fig. 13c, d).

In the absence of Matrigel, our human somitoid protocol (Fig. 1a) never gave rise to a somite-like structure (Supplementary Fig. 14, W/O Matrigel). Instead, hourglass-shaped bodies were formed without Matrigel (Fig. 4a; Supplementary Movie 5, W/O Matrigel), and apical markers ZO-1 and F-actin showed polarized localization in small clusters of cells, especially near the narrow neck of the hourglass (Fig. 4b). However, the cell clusters with apical-basal polarities were small and sporadic, and they did not form sequential epithelial organizations with a consistent apical-basal polarity (Fig. 4b, c).

Interestingly, somitoids without Matrigel still displayed a proper anterior-posterior axis, expressing a somite marker *UNCX4.1* on the opposite side to a PSM marker *HES7* (Fig. 4d). Another somite marker *TBX18* showed a localized expression in the neck region of the hourglass. The expression levels of a MET marker *N-CADHERIN*<sup>37</sup> checked with qRT-PCR analysis were also comparable between the absence and presence of Matrigel (Supplementary Fig. 14b). Our scRNA-seq analyses further confirmed that all the major cell types identified in the control somitoids (10% Matrigel), putative NMP, PSM, somite, and late somite cells, existed in the somitoids without Matrigel (Supplementary Fig. 15). Even some sclerotome and dermomyotome markers were similarly expressed in the absence of Matrigel



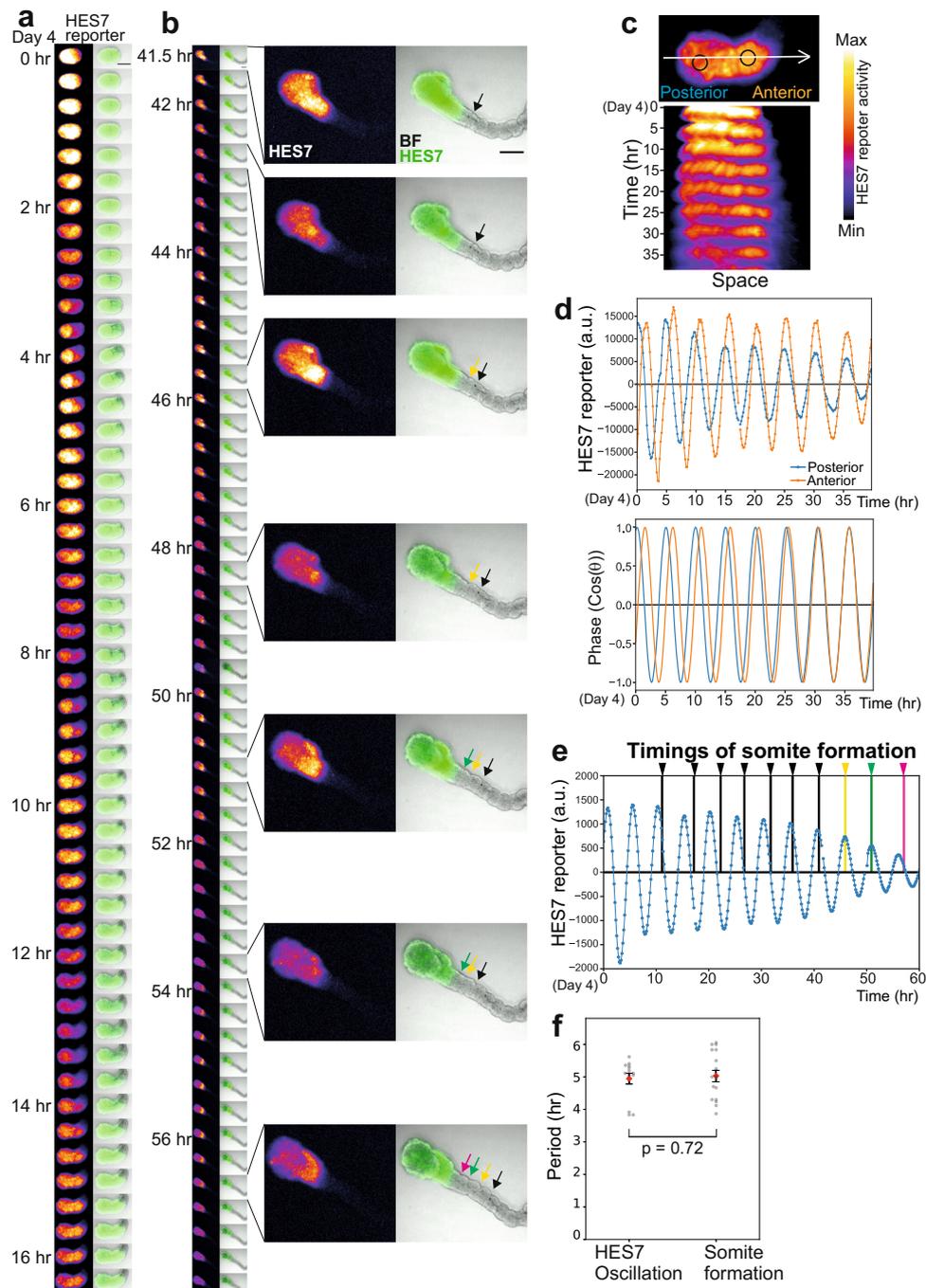
**Fig. 2 scRNA-seq analyses of somitoids.** **a** Uniform manifold approximation and projection (UMAP) plot of day 7 somitoids with 10% Matrigel ( $N = 27$  pooled samples, 3781 cells) colored by the 6 clusters identified in Supplementary Fig. 8a. NMP: Neuromesodermal progenitor, PSM: presomitic mesoderm. The cluster label for Neural is not shown as the cells in this cluster represent less than 0.1% of the total. **b** UMAP plots colored by the expression of selected marker genes. **c** Heatmap of the natural log expression of selected marker genes associated with early embryonic development. **d** Subset analysis of Late somite cells colored by the expression of sclerotome (*TWIST1*) and dermomyotome (*DMRT2*) markers (bottom).

(Supplementary Fig. 15a, e). These results suggest that Matrigel is dispensable for the axis elongation and somite cell differentiation but necessary for the epithelialization and the establishment of a consistent apical-basal polarity. In embryos, the PSM and somites are surrounded by the surface ectoderm, and the ECM produced from the surface ectoderm, particularly fibronectin, is essential for somite formation<sup>15,32,38,39</sup>. Matrigel may act as a surrogate for the surface ectoderm that signals the apical-basal polarity to somite-forming cells. Even though fibronectin is not a major component of Matrigel<sup>40</sup>, *FIBRONECTIN* (*FN1*) was expressed in somitoids (Supplementary Fig. 9c), and thus Matrigel may trap FIBRONECTIN secreted by organoids<sup>15</sup>.

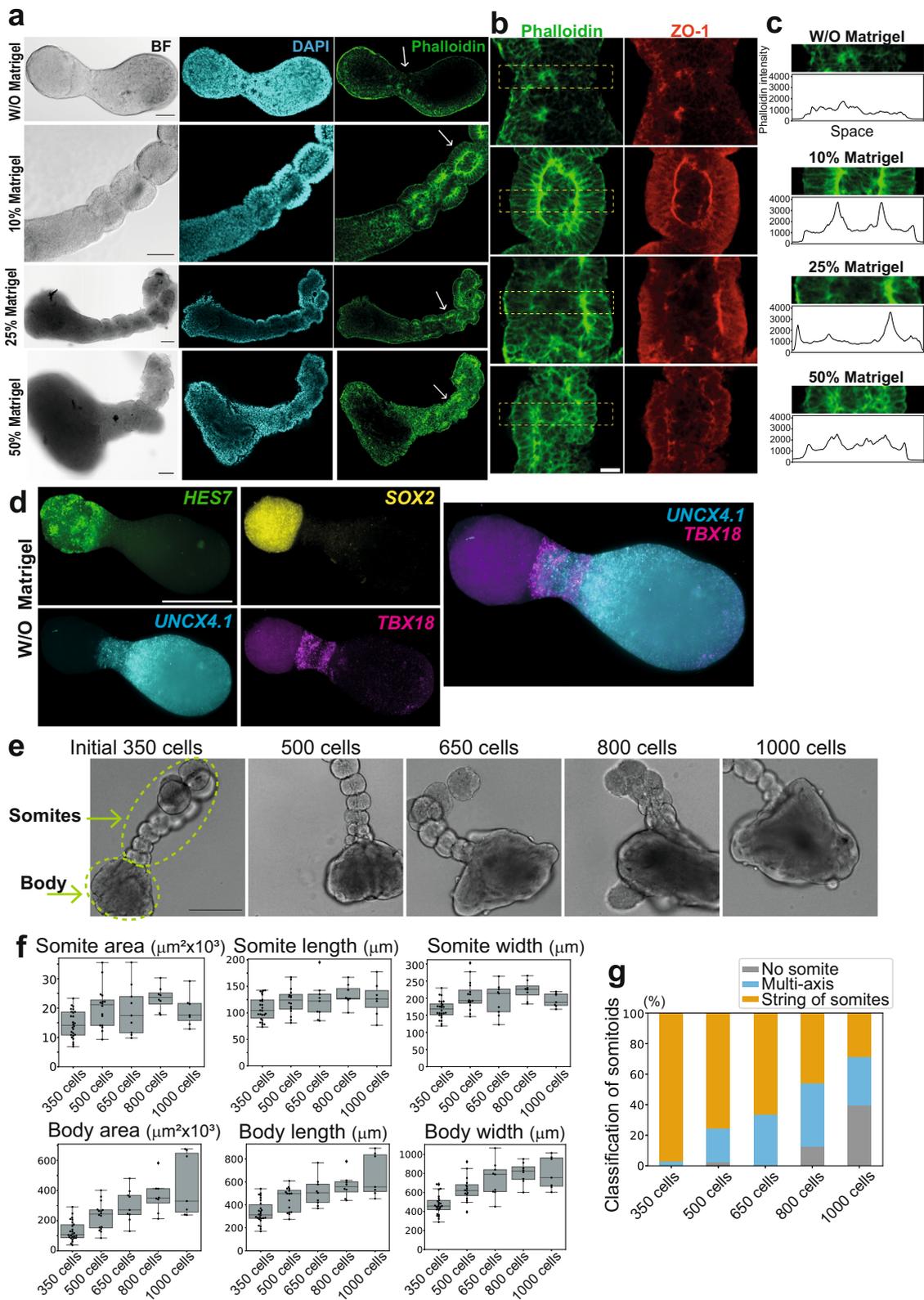
While a low concentration (5%) of Matrigel resulted in similar somitoids as the control 10% Matrigel, high concentrations (25%

and 50%) of Matrigel led to abnormal morphologies (Fig. 4a; Supplementary Fig. 14a; Supplementary Movie 5). Although the cells were still epithelialized with high concentration Matrigel, the somite shapes were not spherical but skewed and convoluted (Fig. 4b, c), and occasionally, multiple small somites were arranged like a ‘bunch of grapes’<sup>15,33</sup> instead of a string of somites (Supplementary Fig. 14a; Supplementary Movie 5, 50% Matrigel). The skewed somites could be due to the stiffness of high concentration Matrigel, and these results highlight the importance of a proper concentration (5–10%) of Matrigel<sup>41</sup>.

**The somites have a preferred size.** As another crucial condition in many types of organoids is the initial cell number<sup>11,42–44</sup>, we



**Fig. 3 Somites are formed according to the segmentation clock. a, b** Time-lapse images of the HES7 promoter-luciferase reporter activity in a representative somitoid. The starting point on day 4 was defined as time 0. See also Supplementary Movie 4. Scale bars: 500  $\mu\text{m}$ . Images were taken by an LV200 luminescent microscope. **b** Enlarged images of several time points are also shown. Arrows indicate the somite boundaries. BF: Bright field. **c** Kymograph of the HES7 reporter activity measured along the posterior-anterior axis (the white arrow) of the same sample shown in **a, b**. **d** Detrended intensity (top) and the oscillation phase (bottom) of the HES7 reporter activity in the same sample shown in **a, b**. Blue and orange lines indicate the signals measured in the posterior and anterior regions of somitoids, respectively, marked in **c**. The gaps in the graph correspond to short halts of imaging to adjust the sample position.  $N = 14$  samples showed similar oscillatory patterns. **e** Relationship between HES7 oscillations and segmentation timings. Detrended HES7 reporter activity of the entire image of the same sample shown in **a** and **b**. Arrows indicate the timings of somite formation, and the colors correspond to the arrows in **b**. **f** Periods calculated from the HES7 oscillations and somite formation. Mean  $\pm$  SEM. Each point in the graph indicates an average period of several oscillation peaks or somite formation timings in one sample from day 5 to day 6.  $N = 14$  (HES7 oscillation) and 17 (Somite formation). 5 independent experiments.  $P$ -value is from two-sided student's  $t$ -test. Source data are provided as a Source Data file.



made somitoids with various numbers of human iPSCs ranging from 200 to 1000 cells. The somitoids created from 200 cells, instead of the control 350 cells, stopped growing early (Supplementary Fig. 16), and thus they were excluded from the following analyses. Above 350 cells, the size of the posterior ‘body’ that includes NMP and PSM regions monotonously increased according to the initial cell number (Fig. 4e, f; see Body

morphometry; Supplementary Fig. 17). Surprisingly, however, the size of somites did not increase accordingly, but it was rather constant (Fig. 4e, f; see Somite morphometry). This suggests that the somites have a preferred size, which might be determined by local cell-cell interactions<sup>33</sup>, the segmentation clock, or other mechanisms. Note that, however, the somitoids with large initial cell numbers tended to form multiple axes and sometimes failed

**Fig. 4 Effects of Matrigel and the initial cell number on somitoid morphologies.** **a** Bright field (BF), DAPI staining, and Phalloidin staining images of the somitoids created without (W/O) or with different concentrations of Matrigel. Matrigel was added on day 4, and the images were taken on day 6. Scale bars: 100  $\mu\text{m}$ . All samples stained ( $N = 6$  (W/O Matrigel), 10 (10% Matrigel), 8 (25% Matrigel), and 5 (50% Matrigel)) showed similar expression patterns. **b** Enlarged images of the regions indicated by the white arrows in **a**. Different Z-planes were used between **a** and **b** to show the images clearly. ZO-1 staining images are also shown. Scale bars: 30  $\mu\text{m}$ . **c** Enlarged images of the boxed regions in **b** and their intensity profiles. **d** HCR images of the day 7 somitoids without Matrigel. Scale bar: 300  $\mu\text{m}$ .  $N = 4$  samples showed similar expression patterns. **e** Representative images of the day 7 somitoids created from 350, 500, 650, 800, and 1000 cells. The light green dashed lines indicate a string of somites and a body that includes the PSM and NMPs. Scale bar: 350  $\mu\text{m}$ . **f** Size measurements of the somite and body parts of the somitoids created with different initial cell numbers. The length and width mean the distances along the anterior-posterior axis and its perpendicular axis, respectively. Only the newest somites were measured. Boxplots show median, 75th and 25th percentiles, and max and min except for outliers.  $N = 26$  (350 cells), 16 (500 cells), 9 (650 cells), 8 (800 cells), and 7 (1000 cells). **g** Classification of the somitoids created with different initial cell numbers.  $N = 150$  (350 cells), 45 (500 cells), 24 (650 cells), 39 (800 cells), and 28 (1000 cells). 5–12 independent experiments. Part of samples of 350 cells is common to Fig. 1c. Microscopes: FV3000 confocal (**a**), Light-sheet (**d**), and Opera (**e**). Source data are provided as a Source Data file.

to form somites (Fig. 4g), indicating an optimal initial number of 350–500 cells.

**WNT signaling instructs lineage specification.** The amount of WNT signaling is known to instruct NMPs to become either the neural tube or PSM/somites<sup>15,25,26,45</sup>. WNT activation promotes mesodermal lineage differentiation at the expense of neural lineages. Thus, we modulated the dosage of the WNT signaling activator CHIR during the initial 2 days of the somitoid protocol (Fig. 5a). High doses of CHIR (8–10  $\mu\text{M}$ ) of the original protocol resulted in usual somitoids with strings of somites (Fig. 5b; Supplementary Fig. 18a). Lowering the dosage decreased the possibility of somite formation (Fig. 5b–d; Supplementary Fig. 18). With 5  $\mu\text{M}$  CHIR, most somite-like structures either disappeared or became sporadic. Instead, an elongated, large epithelial structure was observed (Fig. 5b; Supplementary Fig. 18a). The elongated epithelial structure was SOX2-positive (Fig. 5e, f), suggesting that it may be closer to the neural tube. With intermediate CHIR doses (6–7  $\mu\text{M}$ ), both the elongated epithelial structure and somites were formed (Fig. 5b, e; Supplementary Fig. 18a). The elongated epithelial structure was occasionally flanked by bilateral somites (Fig. 5e, 6–7  $\mu\text{M}$ ), reminiscent of the neural tube and bilateral somites formed in mouse TLSs<sup>15</sup>. The elongated epithelial structure was indeed positive for neural tube markers (SOX2, SOX1, and PAX6)<sup>15</sup>, whereas the somite-like structures were negative (Supplementary Fig. 19). Note that, however, long strings of somites were rarely observed with lower CHIR doses, and sporadic or short rows of somites were often formed (Fig. 5c, d; Supplementary Fig. 18b, c, 5–7  $\mu\text{M}$ ). The WNT sensitivity of somitoids was so high that even a 1  $\mu\text{M}$  increase in the CHIR dosage dramatically changed the outcome (Supplementary Fig. 18a). These results suggest that somite formation in somitoids needs a high amount of WNT signaling to make mesodermal lineages, namely the PSM and somites, at the expense of neural lineages (Fig. 5f).

To confirm the effect of WNT signaling on the cell lineage specification in somitoids, we barcoded the somitoids that were created with 5  $\mu\text{M}$ , 7  $\mu\text{M}$ , and 10  $\mu\text{M}$  CHIR and performed MULTI-seq analyses<sup>46</sup> (Supplementary Figs. 20–22). As expected, the populations of putative NMP, PSM, somite, and late somite cells were identified as the major cell types in the somitoids with the high CHIR dose (Supplementary Fig. 20b, c, 10  $\mu\text{M}$ ). With the low CHIR dose, by contrast, neural tube markers SOX1 and PAX6 were expressed in the somitoids (Supplementary Fig. 21, 5  $\mu\text{M}$ ), and the putative neural cells and NMPs were the major cell types in addition to the minor somite cell population (Supplementary Fig. 20b, c, 5  $\mu\text{M}$ ). RNA velocity demonstrated the differentiation of the neural cells from NMPs (Supplementary Fig. 20e). The neural cell population was not detected in the somitoids with the high CHIR dose (Supplementary Fig. 20b, c, 10  $\mu\text{M}$ ). All

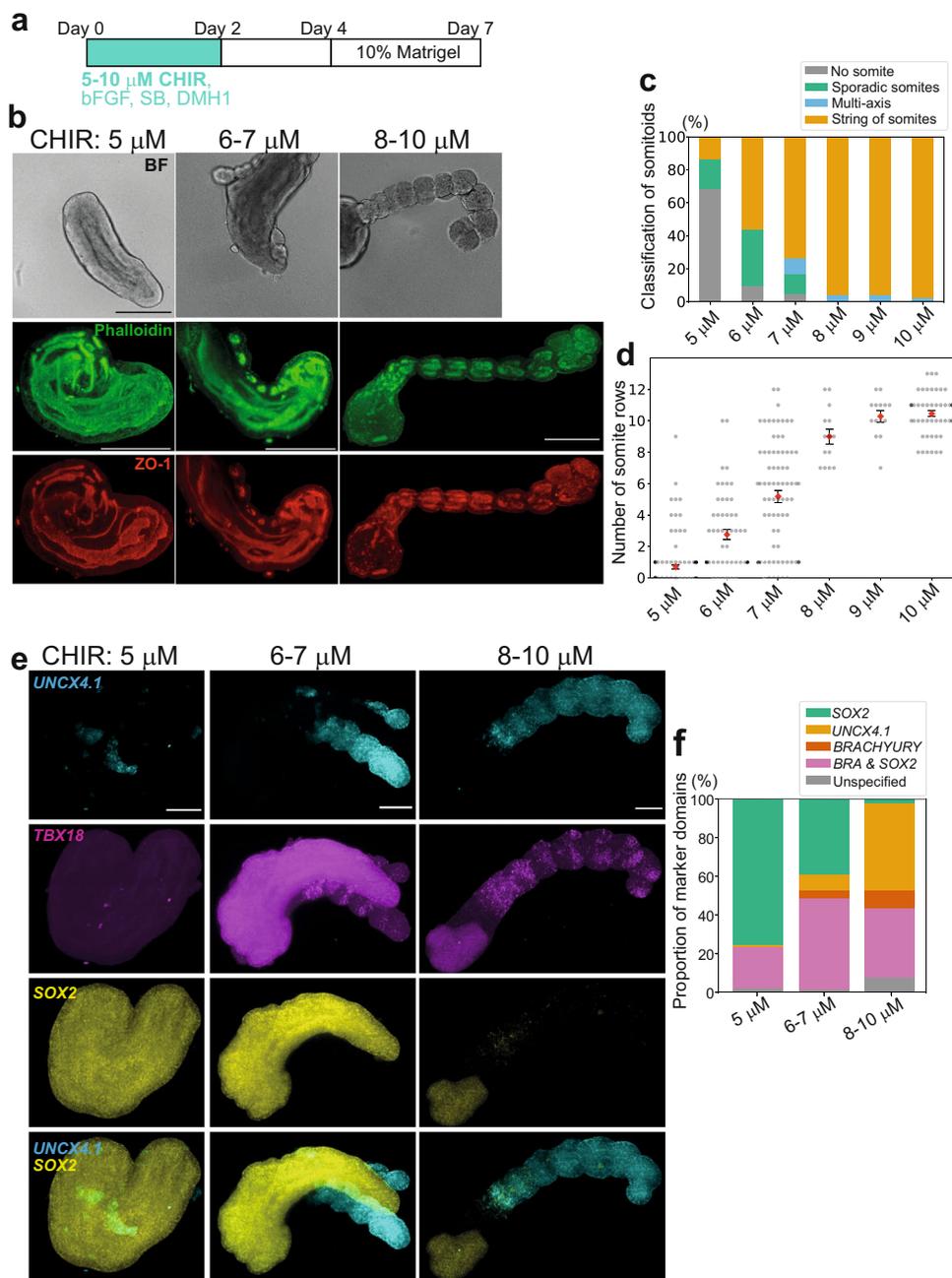
these cell types, including both neural and mesodermal lineages, were detected in somitoids with the intermediate CHIR dose (Supplementary Fig. 20b, c, 7  $\mu\text{M}$ ). Gene ontology (GO) terms related to neural and mesodermal differentiation were enriched in the genes upregulated by 5  $\mu\text{M}$  and 10  $\mu\text{M}$  CHIR, respectively (Supplementary Fig. 20d; Supplementary Data 1). Thus, the MULTI-seq comparison clearly depicted the bias towards neural or mesodermal lineages depending on the low or high amount of WNT signaling, respectively, during the initial 2 days of the somitoid protocol (Supplementary Figs. 21–22). The day 3 somitoids created with 10  $\mu\text{M}$  CHIR already expressed a PSM marker TBX6 while those with 5  $\mu\text{M}$  CHIR did not (Supplementary Fig. 23), suggesting that the bias in the lineage specification in somitoids already starts on day 3 when somites are yet to be formed.

## Discussion

We created human somitoids that recapitulate periodic somite formation coupled with the oscillations of the segmentation clock. The somitoids elongated, and the anterior PSM cells rhythmically budded off to make pairs of somites. The resulting somite tissues were mature enough to display the caudal-rostral compartments and apical-basal polarity.

Although the somite cells in somitoids expressed a couple of sclerotome and dermomyotome markers, we did not notice morphological changes corresponding to the ventral sclerotomes and dorsal dermomyotomes of *in vivo* embryos. We reason that for further somite maturation, inductive signals coming from the neighboring tissues, such as the neural tube, notochord, and lateral plate mesoderm, should be essential<sup>1</sup>. Treating somitoids with additional growth factors and inhibitors may be a way to make more mature sclerotomes and dermomyotomes. As our somitoids displayed the neural tube-like structure flanked by somites with lower doses of the WNT activator, making several neighboring tissues altogether with somites may be an alternative way.

Like many other epithelial organoids, somitoids needed Matrigel to form somites. Matrigel is currently a ‘magic powder’ in the organoid field, and its precise role remains unclear. Since PSM cells differentiated into the somite fate even in the absence of Matrigel, the ECM surrogate may play a role mainly in cell epithelialization and apical-basal polarization. In addition, the fact that epithelialization itself happened with a wide range of Matrigel concentrations (5–50%) suggests that the accurate stiffness of the gel may not be important for epithelialization. Another unexpected finding regarding the somitoid protocol was that the size of somite was relatively constant irrespective of the initial cell number. This was surprising because the somite size has been proposed to scale with the size of the PSM region<sup>47–49</sup>. Further investigation of the effect of cell number on



**Fig. 5 WNT signaling instructs the mesodermal lineage specification.** **a** Somitoid protocol with altered CHIR concentrations during the initial 2 days. The original protocol uses 8–10  $\mu\text{M}$  CHIR. **b** Bright field (BF), Phalloidin staining, and ZO-1 staining images of the day 7 somitoids created with different CHIR concentrations. BF samples are different from Phalloidin- and ZO-1-stained samples. Scale bars: 350  $\mu\text{m}$ . All samples stained ( $N = 3$  (5  $\mu\text{M}$ ), 7 (6–7  $\mu\text{M}$ ), and 5 (8–10  $\mu\text{M}$ )) showed similar expression patterns. **c** Classification of the day 7 somitoids created with different CHIR concentrations. Sporadic somites mean only 1–2 isolated somites, whereas strings of somites mean more than 3 rows of somites.  $N = 113$  (5  $\mu\text{M}$ ), 62 (6  $\mu\text{M}$ ), 126 (7  $\mu\text{M}$ ), 46 (8  $\mu\text{M}$ ), 45 (9  $\mu\text{M}$ ), and 150 (10  $\mu\text{M}$ ). 4–12 independent experiments. **d** Quantification of the number of somite rows in the day 7 somitoids created with different CHIR concentrations. Mean  $\pm$  SEM.  $N = 112$  (5  $\mu\text{M}$ ), 56 (6  $\mu\text{M}$ ), 77 (7  $\mu\text{M}$ ), 14 (8  $\mu\text{M}$ ), 14 (9  $\mu\text{M}$ ), and 49 (10  $\mu\text{M}$ ). 4–12 independent experiments. Only the images with the entire somitoid structures were measured. **c, d** Part of samples of 8–10  $\mu\text{M}$  is common to Fig. 1c, e. **e** HCR images of day 6 somitoids. Scale bars: 150  $\mu\text{m}$ . **f** Volume quantification of lineage marker expression domains using HCR images of day 6 somitoids. SOX2 + & BRACHYURY-: Neural; UNCX4.1 +: Somite; BRACHYURY + & SOX2-: PSM; BRACHYURY + & SOX2 +: NMP. All samples stained ( $N = 3$  (5  $\mu\text{M}$ ), 4 (6–7  $\mu\text{M}$ ), and 6 (8–10  $\mu\text{M}$ )) showed similar patterns as long as somites were present. Microscopes: Opera (BF of panel **b**) and Light-sheet (Phalloidin and ZO-1 of **b, e**). Source data are provided as a Source Data file.

the segmentation clock, cell differentiation, cell movement, and tissue mechanics will be necessary to address the question<sup>23,33,39,50,51</sup>.

Even though the somitoid protocol was reproducible, it was sensitive to experimental materials and cell states. For instance,

we found that the combination of a particular iPSC line, culture medium, and even U-bottom plate was crucial for somitoids. The morphology of iPSC colonies was also important. The incompatible materials and troubleshooting are summarized in a separate protocol<sup>52</sup>.

Human somitoids should provide a novel platform to study congenital abnormalities in vertebral segmentation, including spondylocostal dysostosis (SCD) and congenital scoliosis<sup>5,53</sup>. We previously modeled SCD phenotypes by recapitulating the human segmentation clock from patient-derived iPSCs<sup>7</sup>. Whereas the *in vitro* segmentation clock revealed congenital defects in oscillation patterns, the new somitoid model will enable us to study defects related to somite formation and epithelialization. As somitogenesis is sensitive to developmental environments, such as hypoxia<sup>54</sup>, somitoids may also be useful for high-throughput assessments of environmental factors and teratogens. Another interesting direction is the inter-species comparison of somitogenesis. We have previously compared the human and mouse segmentation clocks, demonstrating the oscillation period difference between the species<sup>8,55</sup>. The new human somitoid model will enable the comparative study on how the tissue dynamics and morphogenesis of somite formation are regulated across species<sup>56</sup>.

## Methods

**Human iPSC cultures.** Human iPSC line (201B7 line, #HPS0063)<sup>57</sup> was provided by the RIKEN BRC through the National BioResource Project of the MEXT/AMED, Japan. The human iPSCs were maintained without feeder cells and cultured on iMatrix-511 silk (Amsbio, 892021)-coated dishes in StemFit medium (Ajinomoto, StemFit Basic04CT). The cells were trypsinized into single cells by TrypLE solution (Gibco, A1285901) and 0.5 mM ethylenediaminetetraacetic acid (EDTA) in phosphate-buffered saline (PBS) (1:1 mixture).  $1.2 \times 10^4$ – $2.4 \times 10^4$  cells were seeded into an iMatrix-coated 3.5 cm dish in the StemFit medium containing 10  $\mu$ M ROCK inhibitor, Y27632 (Sigma, Y0503). The cells were incubated at 37 °C in a humidified atmosphere of 5% CO<sub>2</sub>. The medium was changed the next day into the StemFit medium without Y27632. The cells were passaged every 5–6 days (upon 70–80% confluency). The HES7 promoter-luciferase reporter line was described previously<sup>7</sup>. The fluorescent nuclear reporter (pCAG-mCherry-NLS) was introduced into the cells by using a piggyBac vector<sup>58</sup> and a 4D-Nucleofector (Lonza). After antibiotics selection, a stable clone was isolated. The cells were regularly tested and reported negative for mycoplasma contamination.

**Generation of human somitoids.** Ethical approval for this project was granted by Departament de Salut de la Generalitat de Catalunya (Carlos III Program). Once human iPSCs reached 50–70% confluency, the cells were washed with PBS(-) twice and incubated with 2 ml 0.5 mM EDTA in PBS at 37 °C for 6–7 min. The cells were mechanically dissociated by pipetting and transferred into 8 ml of the N2B27 medium described below. The cell suspension was centrifuged at  $152 \times g$  for 3 min and resuspended into 10 ml of the N2B27 medium. After 1 more wash with the N2B27 medium, the supernatant was completely removed. The cell pellet was resuspended into 150–500  $\mu$ l of the somitoid induction medium described below with 10  $\mu$ M Y27632. The cell concentration was adjusted to 7 cells/ml with an adequate amount of the pre-warmed induction medium with Y27632. Then, 50  $\mu$ l of the cell suspension (=350 cells) was aliquoted into each well of a U-Shaped-Bottom, 96-well-plate (Nunc) Sphera 96U-well plate, Thermo Scientific, 174925) by using a multichannel pipette. The 96-well-plate was centrifuged at  $152 \times g$  for 2 min for the cells to settle down on the bottom of the plate. One day after aggregate formation, 150  $\mu$ l of the fresh somitoid induction medium without Y27632 was added to each well. On days 2 and 3, 150  $\mu$ l of the medium was carefully removed and replaced with 150  $\mu$ l of the fresh N2B27 medium without disturbing the aggregate. On day 4, 150  $\mu$ l of the medium was removed from the well and replaced with 150  $\mu$ l of the fresh N2B27 medium containing 10% Matrigel (growth factor reduced, Corning, 356231). The medium was not changed after the Matrigel addition. The step-by-step instructions and typical problems can be also found in the Protocol Exchange<sup>52</sup>.

**Somitoid induction medium.** The N2B27 medium is a mixture of DMEM/F12 (Gibco, 21331020) with 1x N2 supplement (R&D, AR009) and Neurobasal medium (Gibco, 21103049) with 1x B27 supplement (Gibco, 17504-044) at a 1:1 ratio. The N2B27 medium was also supplemented with 2 mM Glutamax (Gibco, 35050-038), 0.1 mM nonessential amino acids (Gibco, 11140-035), 1 mM sodium pyruvate (Gibco, 11360-039), and penicillin/streptomycin (Gibco, 15140-122).

The Somitoid induction medium is the N2B27 medium containing 10  $\mu$ M SB431542 (Sigma, S4317), 8–10  $\mu$ M CHIR99021 (Sigma, SML1046), 2  $\mu$ M DMH1 (Sigma, D8946), and 20 ng/ml bFGF (PeproTech, AF-100-18B).

**Quantitative RT-PCR.** Total RNA was extracted with RNeasy Micro Kit (Qiagen, 74004). Then cDNA was synthesized from 0.5  $\mu$ g of the total RNA with Quantitect Reverse Transcription kit (Qiagen, 205311). Quantitative PCR was performed with LightCycler 480 SYBR Green I Master (Roche, 4707516001) and the gene-specific

primers<sup>9,16,59</sup> (Supplementary Data 2) with LightCycler 480 II (Roche). The expression levels of the target genes were normalized by *GAPDH* expression.

**In situ hybridization chain reaction (HCR).** HCR was performed as described<sup>29</sup> using probes against the following list of target genes: *HES7* (Accession NM\_001165967, hairpin B4), *BRACHYURY* (Accession NM\_003181, hairpin B5), *UNCX4.1* (Accession NM\_001080461, hairpin B1), *SOX2* (Accession NM\_003106, hairpin B3), and *TBX18* (Accession NM\_001080508, hairpin B2). Hairpin B1 was labeled with Alexa 546, B2 was labeled with Alexa 488 or Alexa 647, B3 was labeled with Alexa 700, B4 was labeled with Alexa 488, and B5 was labeled with Alexa 647. Images were taken with a MuVi-SPIM Light-Sheet Microscope (Luxendo/Bruker, Luxendo processor software v3.0) The images from the opposing camera views were fused with the Luxendo image processor pipeline using content-based registration.

Analysis of HCR somitoid images was performed using a custom-written napani<sup>60</sup> plugin which is available via GitHub (Fig. 5). Images were loaded in the plugin, and the binarized image for each available channel was computed using the Otsu thresholding algorithm. Next, the masks were visually inspected to confirm the accurate segmentation and occasionally correct segmentation errors that were caused, for instance, by the low fluorescence signals in some of the images. To compute the overall volume of the somitoid, we made use of the fact that the genes chosen displayed patterns of expression that jointly covered the whole somitoid structure. An artificial channel which was the sum of all the individual fluorescence channels for the same somitoid was created, and it was binarized using the Otsu algorithm. Given the anisotropic resolution of the 3D stack ( $Z = 2 \mu$ m,  $XY = 0.391 \mu$ m), the positive pixel values within each mask were counted, and the pixel counts were converted into volumes.

**Immunohistochemistry (IHC).** Somitoids were collected with wide-bore tips into 2 ml Eppendorf tubes. After 2 washes with PBS(-), 300  $\mu$ l of 4% paraformaldehyde (PFA) was applied, and the samples were incubated at 4 °C overnight. After 2 washes with PBS(-), they were further washed twice with the blocking buffer (PBS containing 10% fetal bovine serum (FBS) and 0.2% Triton X-100) for 5 min each. Then, the samples were preincubated in the fresh blocking buffer at room temperature for a few hours before being incubated with an anti-ZO-1 antibody (1/300 dilution, Invitrogen, 61-7300), anti-TBX6 antibody (1/300 dilution, Abcam, ab38883), anti-SOX2 antibody (1/300 dilution, R&D, MAB2018), anti-BRACHYURY antibody (1/300 dilution, R&D, AF2085), anti-SOX1 antibody (1/300 dilution, R&D, AF3369), or anti-PAX6 antibody (1/300 dilution, Abcam, ab195045) in the blocking buffer at 4 °C overnight. The next day, the samples were washed 3 times with the blocking buffer for 15 min each. Then they were incubated with Alexa Fluor 488 Phalloidin (1/300 dilution, Invitrogen, A12379), Alexa Fluor 594 Goat anti-Rabbit IgG (H + L) (1/500 dilution, Invitrogen, A-11037), Alexa Fluor 647 Donkey anti-Goat IgG (H + L) (1/500 dilution, Invitrogen, A32849), Alexa Fluor 647 Goat anti-Mouse IgG (H + L) (1/500 dilution, Invitrogen, A-21236), or Alexa Fluor 488 Goat anti-Mouse IgG (H + L) (1/500 dilution, Invitrogen, A-11029) together with DAPI (1/1000 dilution, Invitrogen, 62247) in the blocking buffer at room temperature for a few hours. The samples were washed 3 times with PBST (0.2% Triton X-100 in PBS) for 15 min each, and images were taken with an FV3000 confocal microscope (Olympus, FV3000 Fluoview RS software), a MuVi-SPIM Light-Sheet Microscope, an LSM 980 Confocal Laser Scanning Microscope (Zeiss, ZEN software), or an Opera Phenix HSC system (PerkinElmer, Harmony software v4.9). When a somitoid spanned multiple images, the tiled images were stitched. The backgrounds of individual images were corrected, and then the images were stitched using ImageJ Grid/Collection stitching so that their backgrounds matched each other.

**EdU labeling.** EdU staining was performed using the Click-iT EdU Cell Proliferation Kit (Invitrogen, C10337). Somitoids were collected with wide-bore tips into a 3.5 cm dish and incubated with 10  $\mu$ M EdU in the N2B27 medium for 2 hrs in a cell culture incubator. The samples were collected with wide-bore tips into a 2 ml Eppendorf tube. After 2 washes with PBS(-), they were stained for SOX2, TBX6, and DAPI according to the IHC protocol described above. After the final step of PBST wash, they were incubated with the Click-iT reaction cocktail at room temperature for 30 min, protected from light. The samples were washed 3 times with PBST for 15 min each, and images were taken with an Opera Phenix HSC system.

**Live imaging of somitoids.** Bright-field images of live somitoids were taken with an Opera Phenix HSC system in the wide-field mode using a 10 $\times$  air objective (NA 0.3). The focus was manually adjusted. For time-lapse imaging, the incubator module was set at 37 °C and 5% CO<sub>2</sub>, and images were taken every 10 min.

**Time-lapse imaging and quantification of HES7 oscillation.** The HES7 promoter-luciferase reporter line was used<sup>7</sup>. A glass-bottom 3.5 cm dish was coated with 2% w/v lipidure (Amsbio, CM5206) in ethanol twice at room temperature for 5 min each to prevent somitoids from attaching to the dish. After the second coating, the remaining lipidure solution was removed, and the dish was dried at room temperature until the residual ethanol evaporated. To monitor HES7

oscillations under the microscope (Fig. 3; Supplementary Fig. 12), we put day 4 somitoids on the lipidure-coated dish with 2 ml of the fresh N2B27 medium containing 0.5 mM luciferin and 10% Matrigel. Live-luminescence imaging was performed with an LV200 luminescent microscope (Olympus, cellSens Dimension software) using a 10x objective lens (UPLXAPO10X). Images were taken every 5 or 10 min with 100 ms exposure for the bright field and 1.5 min exposure for the bioluminescence signal. To monitor a collective luciferase signal from day 1 to day 4 (Supplementary Fig. 11a), we collected 8 somitoids and put them on a non-coated 3.5 cm dish with 2 ml of the fresh N2B27 medium containing 0.5 mM luciferin. The luciferase activity was recorded using Kronos Dio Luminometer (Atto, Kronos control software v2.3) every 10 min with 1 min exposure.

For image analyses, the ImageJ software and a Python script were used as previously described<sup>10</sup>. Briefly, kymographs were generated from time-lapse images filtered by Median filter and Tissue alignment filter to keep somitoids at the same position in a frame<sup>61</sup>. Resliced stack images were arranged from top to bottom in a temporal order. The detrended intensity and phase dynamics were calculated by pyBOAT<sup>62</sup> with a 50-frame window. The peak-to-peak period of HES7 oscillation was measured with Python functions detecting peaks and bottoms of HES7 oscillation. The period of somite formation was manually calculated from time-lapse bright-field images.

**Somite morphometry.** DAPI images taken with an FV3000 confocal microscope or an Opera Phenix HSC system were used. Somite morphometric analysis was performed in a semi-automated manner using custom-written Python scripts which is available via GitHub. First, images were loaded and resized to obtain a pixel size of 1  $\mu\text{m}$  and subsequently binarized, choosing an algorithm that could accurately segment the somitoid. The algorithm chosen depends on the pixel intensity histogram, and either the triangle<sup>63</sup> or the yen<sup>64</sup> algorithm worked well for all images. Next, holes and debris smaller than 10000 pixels in the binarized image were removed, and the mask was smoothed using the morphological opening operation with a disk-shaped footprint of 5 pixels radius. Then the smoothed mask was used to compute the euclidean distance transform and find its local maxima, which resulted in a disordered collection of 2D points roughly located along the midline of the somitoid. To order the points and find a smooth curve along the longitudinal axis of the somitoid, we manually selected the posterior position and the anterior position of the somitoid. The ordered points were used to generate a B-spline representation with a smoothing value of 10000, and the width of the somitoid at every point along the spline curve was computed. This approach allowed us to generate a characteristic width profile of each somitoid, in which minima and maxima correspond to the somite-to-somite edges and the maximum somite width, respectively (Fig. 1f, top). Somite morphometric parameters were then computed using the minima-to-minima distance (inter-somite distance) and the maxima of the width profile (width) (Fig. 1f, bottom). Somite area and circularity were calculated by using the somite-to-somite distance and the somite width (Supplementary Fig. 3). Only somitoids with single somites were used for this measurement. The graphs were plotted with Python's boxplot function with the default setting. The middle line of the boxes indicates the median, and the box edges are the 25th and 75th percentiles. The maximum and minimum values are indicated by whiskers.

**Comparison of somite morphometry.** Confocal images of somitoids labeled with the fluorescent nuclear reporter were used. The newest somites were manually segmented, and the morphometric parameters were calculated using the ImageJ software. For the somites of human embryos, the images were obtained from the Virtual Human Embryo project<sup>20</sup> and the Kyoto collection, and the first 4–5 rows of somites were measured. For the somites of mouse trunk-like structures (TLSs) and CHIR99021- and LDN193189-treated TLSs (TLS-CLs)<sup>15</sup>, the DAPI images were provided by the authors. The somites were randomly selected and measured, as the newest somites were difficult to define in TLSs and TLS-CLs. The graphs were plotted with Python's boxplot function with the default setting. The middle line of the boxes indicates the median, and the box edges are the 25th and 75th percentiles. Values located 1.5 times outside the quartile range were defined as outliers and plotted with dots. The maximum and minimum values, excluding outliers, are indicated by whiskers (Supplementary Fig. 2a).

**Cell morphometry.** Phalloidin and DAPI staining image of a day 6 somitoid was taken by a confocal microscope. Nine cells were randomly chosen in 3 regions (Area 1, 2, and 3) of the image, and the cells were manually segmented. The morphometric parameters were calculated using the ImageJ software. The graphs were plotted with Python's boxplot function with the default setting. The middle line of the boxes indicates the median, and the box edges are the 25th and 75th percentiles. Values located 1.5 times outside the quartile range were defined as outliers and plotted with dots. The maximum and minimum values, excluding outliers, are indicated by whiskers (Fig. 1k).

**Size measurements of somitoids with different initial cell numbers.** Somitoids were created from 350 to 1000 human iPSCs that were labeled with the fluorescent nuclear reporter or DAPI, and images were taken by a confocal microscope. The body, which includes NMP and PSM regions, and the newest somites in somitoids

were manually segmented. When somitoids showed multiple axes, the longest string of somites was chosen. The morphometric parameters were measured using the ImageJ software. The graphs were plotted with Python's boxplot function with the default setting. The middle line of the boxes indicates the median, and the box edges are the 25th and 75th percentiles. Values located 1.5 times outside the quartile range were defined as outliers and plotted with dots. The maximum and minimum values, excluding outliers, are indicated by whiskers (Fig. 4f).

**Measurement of first somite formation.** Time-lapse bright-field images of somitoids were analyzed using MORGAna<sup>65</sup>. Briefly, approximately 3% of the images were manually annotated and were fed into MORGAna to generate a classifier neural network. After the application of the network to the time-lapse images, a segmentation mask, together with morphological parameters, such as area, length, and midline, of the somitoids were obtained for every image frame in the movie (Supplementary Fig. 13a). Subsequently, the left-right (LR) position at which the first somite is formed was annotated manually in the movie. The annotated image frame determined the timing of the first somite formation (Supplementary Fig. 13b). The intersection point P between the LR segment and the midline of the somitoid was used to obtain the relative position of somite formation (Supplementary Fig. 13c). Finally, the angle of the first somite formation was defined as the angle between the vector orthogonal to the LR segment and the tangent vector in the intersecting position P along the midline (Supplementary Fig. 13d).

### Single-cell RNA sequencing (scRNA-seq)

**Cell preparation.** Somitoids were created with individual conditions (Control 10% Matrigel, W/O Matrigel, 3 different concentrations of CHIR) and collected with wide-bore tips into 2 ml Eppendorf tubes on day 7 (Fig. 2; Supplementary Figs. 8–9, 15, 20–22). After 2 washes with PBS(-), the samples were trypsinized into single cells through the treatment with the TrypLE solution and 0.5 mM EDTA in PBS (1:1 mixture) at 37 °C for 5 min. The cells were mechanically dissociated by pipetting and transferred into PBS containing 0.1% bovine serum albumin (BSA). The cell suspension was centrifuged at 200  $\times g$  for 3 min and resuspended into 0.1% BSA in PBS.

**MULTI-seq sample preparation.** The somitoids created with different CHIR concentrations were labeled with MULTI-seq barcode oligonucleotides for sample multiplexing as described (Supplementary Figs. 20–22)<sup>46</sup>. Briefly,  $5 \times 10^5$  cells from the prepared single-cell suspension were resuspended in Cell Prep Buffer (PBS(-) containing 0.1% poly(vinyl alcohol) (PVA) and 1 mM EDTA). First, a 1:1 mixture of the cholesterol-conjugated Anchor-oligonucleotides (Anchor CMO, synthesized by Integrated DNA Technologies) and Barcode oligonucleotides with a distinct barcode for each sample (final concentration, 0.2  $\mu\text{M}$ ) was added and incubated on ice for 5 min. Then, the same concentration of Co-Anchor CMO (synthesized by Integrated DNA Technologies) was added and incubated for another 5 min, followed by rigorous washes with PBS containing 1% BSA 3 times. After washing, we counted the number of cells in each sample and combined them so that the multiplexed suspension contained the same numbers of cells from each sample. The combined sample was filtered through a 35  $\mu\text{m}$  cell strainer and counted again before barcoding below. The MULTI-seq barcode sequences used in this study are listed in Supplementary Data 3.

**Barcoding and sequencing.** Transcripts of each cell in the single-cell suspension were barcoded with Chromium Controller (10 $\times$  Genomics, firmware version 4.00). The reagent system used in this study was Chromium Single Cell 3' GEM, Library & Gel Bead Kit v3 (10 $\times$  Genomics) and a Chromium Single Cell B Chip Kit (10 $\times$  Genomics). Barcoding and cDNA library construction were performed according to the manufacturer's instructions. After the cDNA amplification step, the barcode fraction was collected, amplified, and single-indexed with KAPA HiFi HotStart ReadyMix (Roche, #KK2601) for sequencing. Both finished cDNA and MULTI-seq-barcode libraries were sequenced with NextSeq500 (Illumina). We read 8 base pairs (bp) for TruSeq Indices, 28 bp for 10 $\times$  barcodes and unique molecular identifiers (UMIs), and 56 bp for both fragmented cDNA and MULTI-seq barcodes.

### scRNA-seq data analyses

**Read alignment.** Sequenced reads were aligned to the human genome (GRCh38) and counted to generate the feature-barcode matrices with the CellRanger pipeline (v.6.1.1, 10 $\times$  Genomics). The reads having the same UMI were collapsed as a single count. The basic statistics of sequencing results are shown in Supplementary Data 3.

**Quality control (QC).** The quality control of the single-cell transcriptome data was performed with R version 4.1.2<sup>66</sup> with the package *Seurat* (v.4.0.5)<sup>67</sup> as follows: First, we excluded the droplet barcodes that had less than 200 UMIs and the undetected genes that were found less than 3 times in an entire sample. Then, we identified cell barcodes based on scatter plots of detected gene counts against the proportion of mitochondrial gene expression in each cell (Supplementary Fig. 24). The areas demarcated by the red polygons were filtered as cells and used for further analyses. After this initial filtering, potential cell doublets were removed with *Scrublet* (v.0.2.3)<sup>68</sup> except the MULTI-seq sample. Doublets in the MULTI-seq sample were removed based on the barcode information. The basic statistics of the filtered cells are shown in Supplementary Data 3.

**Demultiplexing of the MULTI-seq sample.** We counted MUTLI-seq barcode reads associated with the filtered cell barcodes by utilizing the R package *deMULTIplex* (v.1.0.2)<sup>46</sup>, and assigned each cell to one of the following classes: a singlet of one of the original 3 samples, a doublet, or a label-negative cell with *Seurat*. The summary of the demultiplexing results is shown in Supplementary Data 3. We kept only singlets for further analysis.

**Data normalization.** The raw UMI counts of the QC-filtered cells were normalized in each sample according to a deconvolution approach based on pool-based size factors<sup>69</sup> with the R package *scran* (v.1.22.1)<sup>70</sup>. Briefly, cells were clustered with approximation and pooled as blocks, and the size factor to correct the library size of each cell was calculated with the *calculateSumFactors* function, taking into account the size factor of the block where the cell belonged. Finally, raw counts of each cell were normalized based on the size factor and Log<sub>2</sub>-transformed with the *logNormCounts* function. These scores appear as 'Log expression level'. To average the expression levels within cell populations, we used the R packages *scuttle* (v.1.4.0)<sup>71</sup> for heatmaps and *scater* (v.1.22.0)<sup>71</sup> for dot plots.

**Identification of highly variable genes (HVGs).** After normalization, highly variable genes were chosen with the *modelGeneVarByPoisson* function in the *scran* package, which decomposed the total variance of each gene's expressions into biological and technical components by assuming the Poisson distribution for the technical noise. Genes with the biological variance > 0.1 were chosen as HVGs for further dimension reduction and data integration. The number of HVGs were 1,618, 1,567, and 1,304 for 10% Matrigel, W/O Matrigel, and CHIR MULTI-seq samples, respectively.

**Data integration and visualization.** To integrate all three data sets, we first scaled the normalized data according to the whole library size of each sample with the *multiBatchNorm* function in the R package *batchelor* (v.1.10.0)<sup>72</sup>. HVGs across data sets were selected as the genes with mean biological variances > 0.1 ( $n = 1,486$ ) for principal component analysis (PCA). We utilized a mutual nearest neighbor approach by *fastMNN*<sup>72</sup> in the *batchelor* package for data integration considering the first 20 dimensions of PCA. The corrected results were used for visualization of all samples in a single plot of uniform manifold approximation and projection (UMAP)<sup>73</sup> with the *Seurat* package. This UMAP embedding was also used for all plots of individual samples simply by splitting the plot sample by sample. The gene expression levels shown in UMAP plots are Log<sub>2</sub>-transformed normalized values of individual samples or the scaled values with *themultiBatchNorm* function when multiple samples were involved.

**Clustering and marker gene detection.** To define the cell population clusters, a rank-weighted shared-nearest neighbor graph was constructed for the integrated data with the *R bluster* (v1.4.0)<sup>74</sup> package, and the Leiden algorithm<sup>75</sup> was applied for modularity optimization. Marker genes of each cluster were detected with the *scran* package by calculating an area under the curve (AUC) for each gene as an index of the performance to distinguish two clusters in a pairwise comparison, and the minimum AUC against other clusters was taken as a summarizing score of each gene. The genes with the highest minimum AUC were chosen as marker genes of each cluster, and we annotated each cluster based on them and also the expression of other known marker genes. The top 10 markers of each cluster were shown in Supplementary Fig. 8b. This annotation, based on the integrated data set across the three samples, was maintained in the analysis of individual samples to keep consistency in the paper.

**RNA velocity analysis.** Exon/intron read count matrix was generated with *Velocyto* (v.0.17.17)<sup>76</sup>, and the RNA velocity was calculated and visualized with *scVelo* (v.0.2.4)<sup>77</sup>.

**Gene ontology (GO) enrichment analysis.** We calculated AUC by comparing the CHIR 5  $\mu$ M and 10  $\mu$ M groups and chose 100 upregulated genes with the highest AUC from each group as inputs. The GO enrichment analysis and visualization of the results were performed with the *clusterProfiler* package (v.4.2.2)<sup>78</sup> in R.

**Statistics and reproducibility.** The number of samples and independent experiments are given in the figure legends. No statistical method was used to pre-determine sample size. If cells did not aggregate on day 1 of the somitoid protocol due to cell culture conditions, the experiment was discontinued. The experiments were not randomized. The Investigators were not blinded to allocation during experiments and outcome assessment.

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

## Data availability

The scRNA-seq data used in this study are available in the ArrayExpress database under accession code [MTAB-11334](https://www.ebi.ac.uk/ena/arrayexpress/experiments/MTAB-11334). Source data are provided with this paper.

## Code availability

The custom Python scripts used for the somite morphometry and HCR volume analysis are available from Nikola86's Github [[github.com/Nikola86/2022\\_Somitoids\\_Analysis](https://github.com/Nikola86/2022_Somitoids_Analysis)].

Received: 22 July 2021; Accepted: 11 April 2022;

Published online: 28 April 2022

## References

- Chal, J. & Pourquié, O. Patterning and differentiation of the vertebrate spine. *Skelet. Syst.* 41–116 (2009).
- Hubaud, A. & Pourquié, O. Signalling dynamics in vertebrate segmentation. *Nat. Rev. Mol. Cell Biol.* 15, 709–721 (2014).
- Saga, Y. & Takeda, H. The making of the somite: molecular events in vertebrate segmentation. *Nat. Rev. Genet.* 2, 835–845 (2001).
- O'Rahilly, R. & Müller, F. Somites, spinal ganglia, and centra. *Cells Tissues Organs* 173, 75–92 (2003).
- Turnpenny, P. D. et al. Abnormal vertebral segmentation and the notch signaling pathway in man. *Dev. Dyn.* 236, 1456–1474 (2007).
- Chu, L.-F. et al. An in vitro human segmentation clock model derived from embryonic stem cells. *Cell Rep.* 28, 2247–2255.e5 (2019).
- Matsuda, M. et al. Recapitulating the human segmentation clock with pluripotent stem cells. *Nature* 580, 124–129 (2020).
- Matsuda, M. et al. Species-specific segmentation clock periods are due to differential biochemical reaction speeds. *Science* 369, 1450–1455 (2020).
- Diaz-Cuadros, M. et al. In vitro characterization of the human segmentation clock. *Nature* 580, 113–118 (2020).
- Matsumiya, M., Tomita, T., Kobayashi, K., Isomura, A. & Kageyama, R. ES cell-derived presomitic mesoderm-like tissues for analysis of synchronized oscillations in the segmentation clock. *Development* 145, dev156836 (2018).
- van den Brink, S. C. et al. Symmetry breaking, germ layer specification and axial organisation in aggregates of mouse embryonic stem cells. *Development* 141, 4231–4242 (2014).
- Beccari, L. et al. Multi-axial self-organization properties of mouse embryonic stem cells into gastruloids. *Nature* 562, 272–276 (2018).
- Moris, N. et al. An in vitro model of early anteroposterior organization during human development. *Nature* 582, 410–415 (2020).
- van den Brink, S. C. et al. Single-cell and spatial transcriptomics reveal somitogenesis in gastruloids. *Nature* 582, 405–409 (2020).
- Veenfliet, J. V. et al. Mouse embryonic stem cells self-organize into trunk-like structures with neural tube and somites. *Science* 370, (2020).
- Loh, K. M. et al. Mapping the pairwise choices leading from pluripotency to human bone, heart, and other mesoderm cell types. *Cell* 166, 451–467 (2016).
- Nakajima, T. et al. Modeling human somite development and fibrodysplasia ossificans progressiva with induced pluripotent stem cells. *Development* 145, dev165431 (2018).
- Xi, H. et al. In vivo human somitogenesis guides somite development from hPSCs. *Cell Rep.* 18, 1573–1585 (2017).
- Budjan, C. et al. Paraxial mesoderm organoids model development of human somites. *eLife* 11, e68925 (2022).
- Cork, R. J. & Gasser, R. F. The Virtual Human Embryo. <http://virtualhumanembryo.lsuhsu.edu> (2011).
- McCull, J. et al. 4D imaging reveals stage dependent random and directed cell motion during somite morphogenesis. *Sci. Rep.* 8, 12644 (2018).
- Naganathan, S. R., Popović, M. & Oates, A. C. Somite surface tension buffers imprecise segment lengths to ensure left-right symmetry. *bioRxiv* 2020.08.14.251645 <https://doi.org/10.1101/2020.08.14.251645> (2020).
- Shelton, E. R. et al. Stress-driven tissue fluidization physically segments vertebrate somites. *bioRxiv* 2021.03.27.437325 <https://doi.org/10.1101/2021.03.27.437325> (2021).
- Tzouanacou, E., Wegener, A., Wymeersch, F. J., Wilson, V. & Nicolas, J.-F. Redefining the progression of lineage segregations during mammalian embryogenesis by clonal analysis. *Dev. Cell* 17, 365–376 (2009).
- Gouti, M. et al. A gene regulatory network balances neural and mesoderm specification during vertebrate trunk development. *Dev. Cell* 41, 243–261.e7 (2017).
- Koch, F. et al. Antagonistic activities of Sox2 and brachyury control the fate choice of neuro-mesodermal progenitors. *Dev. Cell* 42, 514–526.e7 (2017).
- Bessho, Y. Periodic repression by the bHLH factor Hes7 is an essential mechanism for the somite segmentation clock. *Genes Dev.* 17, 1451–1456 (2003).
- Morimoto, M., Takahashi, Y., Endo, M. & Saga, Y. The Mesp2 transcription factor establishes segmental borders by suppressing Notch activity. *Nature* 435, 354–359 (2005).

29. Trivedi, V., Choi, H. M. T., Fraser, S. E. & Pierce, N. A. Multidimensional quantitative analysis of mRNA expression within intact vertebrate embryos. *Development* **145**, dev156869 (2018).
30. Williams, D. R., Shifley, E. T., Braunreiter, K. M. & Cole, S. E. Disruption of somitogenesis by a novel dominant allele of *Lfng* suggests important roles for protein processing and secretion. *Development* **143**, Dev128538 (2016).
31. Nakaya, Y., Kuroda, S., Katagiri, Y. T., Kaibuchi, K. & Takahashi, Y. Mesenchymal-epithelial transition during somitic segmentation is regulated by differential roles of *Cdc42* and *Rac1*. *Dev. Cell* **7**, 425–438 (2004).
32. Correia, K. M. & Conlon, R. A. Surface ectoderm is necessary for the morphogenesis of somites. *Mech. Dev.* **91**, 19–30 (2000).
33. Dias, A. S., Almeida, I., de, Belmonte, J. M., Glazier, J. A. & Stern, C. D. Somites without a clock. *Science* **343**, 791–795 (2014).
34. Naiche, L. A., Holder, N. & Lewandoski, M. FGF4 and FGF8 comprise the wavefront activity that controls somitogenesis. *Proc. Natl Acad. Sci USA* **108**, 4018–4023 (2011).
35. Lewis, J. Autoinhibition with transcriptional delay. *Curr. Biol.* **13**, 1398–1408 (2003).
36. Ferjentsik, Z. et al. Notch is a critical component of the mouse somitogenesis oscillator and is essential for the formation of the somites. *PLoS Genet.* **5**, e1000662 (2009).
37. Duband, J. L. et al. Adhesion molecules during somitogenesis in the avian embryo. *J. Cell Biol.* **104**, 1361–1374 (1987).
38. Rifes, P. et al. Redefining the role of ectoderm in somitogenesis: a player in the formation of the fibronectin matrix of presomitic mesoderm. *Development* **134**, 3155–3165 (2007).
39. Martins, G. G. et al. Dynamic 3D cell rearrangements guided by a fibronectin matrix underlie somitogenesis. *PLoS ONE* **4**, e7429 (2009).
40. Hughes, C. S., Postovit, L. M. & Lajoie, G. A. Matrigel: a complex protein mixture required for optimal growth of cell culture. *PROTEOMICS* **10**, 1886–1890 (2010).
41. Veenvliet, J. V., Lenne, P.-F., Turner, D. A., Nachman, I. & Trivedi, V. Sculpting with stem cells: how models of embryo development take shape. *Development* **148**, dev192914 (2021).
42. Sasai, Y. Cytosystems dynamics in self-organization of tissue architecture. *Nature* **493**, 318–326 (2013).
43. Lancaster, M. A. et al. Guided self-organization and cortical plate formation in human brain organoids. *Nat. Biotechnol.* **35**, 659–666 (2017).
44. Harrison, S. E., Sozen, B. & Zernicka-Goetz, M. In vitro generation of mouse polarized embryo-like structures from embryonic and trophoblast stem cells. *Nat. Protoc.* **13**, 1586–1602 (2018).
45. Oginuma, M. et al. Intracellular pH controls WNT downstream of glycolysis in amniote embryos. *Nature* **584**, 98–101 (2020).
46. McGinnis, C. S. et al. MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. *Nat. Methods* **16**, 619–626 (2019).
47. Tam, P. P. The control of somitogenesis in mouse embryos. *J. Embryol. Exp. Morphol.* **65**, 103–128 (1981).
48. Lauschke, V. M., Tsiariris, C. D., François, P. & Aulehla, A. Scaling of embryonic patterning based on phase-gradient encoding. *Nature* **493**, 101–105 (2013).
49. Ishimatsu, K. et al. Size-reduced embryos reveal a gradient scaling based mechanism for zebrafish somite formation. *Development* **145**, dev161257 (2018).
50. Sonnen, K. F. et al. Modulation of phase shift between Wnt and notch signaling oscillations controls mesoderm segmentation. *Cell* **172**, 1079–1090.e12 (2018).
51. Rohde, L. A. et al. Cell-autonomous generation of the wave pattern within the vertebrate segmentation clock. *bioRxiv* 2021.05.29.446196 <https://doi.org/10.1101/2021.05.29.446196> (2021).
52. Sanaki-Matsumiya, M. & Ebisuya, M. Human embryonic organoids to recapitulate periodic somitogenesis in vitro. *Protocol Exchange* <https://doi.org/10.21203/rs.3.pex-1852/v1> (2022).
53. Sparrow, D. B., Guillen-Navarro, E., Fatkin, D. & Dunwoodie, S. L. Mutation of HAIRY-AND-ENHANCER-OF-SPLIT-7 in humans causes spondylcostal dysostosis. *Hum. Mol. Genet.* **17**, 3761–3766 (2008).
54. Sparrow, D. B. et al. A mechanism for gene-environment interaction in the etiology of congenital scoliosis. *Cell* **149**, 295–306 (2012).
55. Ebisuya, M. & Briscoe, J. What does time mean in development? *Development* **145**, dev164368 (2018).
56. Rayon, T. & Briscoe, J. Cross-species comparisons and in vitro models to study tempo in development and homeostasis. *Interface Focus* **11**, 20200069 (2021). [rsfs.2020.0069](https://doi.org/10.1098/rsfs.2020.0069).
57. Takahashi, K. et al. Induction of pluripotent stem. *Cells Adult Hum. Fibroblasts Defined Factors Cell* **131**, 861–872 (2007).
58. Woltjen, K. et al. piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells. *Nature* **458**, 766–770 (2009).
59. Shimoda, M. et al. The T-box transcription factor Brachyury regulates epithelial-mesenchymal transition in association with cancer stem-like cells in adenoid cystic carcinoma cells. *BMC Cancer* **12**, 377 (2012).
60. Sofroniew, N. et al. napari/napari: 0.4.10rc0. (Zenodo, 2021). <https://doi.org/10.5281/ZENODO.3555620>.
61. Isomura, A., Ogushi, F., Kori, H. & Kageyama, R. Optogenetic perturbation and bioluminescence imaging to analyze cell-to-cell transfer of oscillatory information. *Genes Dev.* **31**, 524–535 (2017).
62. Mönke, G., Sorgenfrei, F. A., Schmal, C. & Granada, A. E. Optimal time frequency analysis for biological data - pyBOAT. *bioRxiv* 2020.04.29.067744 <https://doi.org/10.1101/2020.04.29.067744> (2020).
63. Zack, G. W., Rogers, W. E. & Latt, S. A. Automatic measurement of sister chromatid exchange frequency. *J. Histochem. Cytochem.* **25**, 741–753 (1977).
64. Jui-Cheng Yen, Fu-Juay Chang & Chang, Shyang A new criterion for automatic multilevel thresholding. *IEEE Trans. Image Process.* **4**, 370–378 (1995).
65. Gritti, N. et al. MORGAna: accessible quantitative analysis of organoids with machine learning. *Development* **148**, dev199611 (2021).
66. R Core Team, 2020. <http://www.r-project.org/index.html>.
67. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).
68. Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: computational identification of cell doublets in single-cell transcriptomic data. *Cell Syst.* **8**, 281–291.e9 (2019).
69. Lun, A. T. L., Bach, K. & Marioni, J. C. Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biol.* **17**, 75 (2016).
70. Lun, A. T. L., McCarthy, D. J. & Marioni, J. C. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. <https://doi.org/10.12688/f1000research.9501.2> (2016).
71. McCarthy, D. J., Campbell, K. R., Lun, A. T. L. & Wills, Q. F. Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* **33**, btw777 <https://doi.org/10.1093/bioinformatics/btw777> (2017).
72. Haghverdi, L., Lun, A. T. L., Morgan, M. D. & Marioni, J. C. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat. Biotechnol.* **36**, 421–427 (2018).
73. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *ArXiv* 1802.03426 (2020).
74. Lun, A. *bluster: Clustering Algorithms for Bioconductor.* (2021).
75. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233 (2019).
76. La Manno, G. et al. RNA velocity of single cells. *Nature* **560**, 494–498 (2018).
77. Bergen, V., Lange, M., Peidli, S., Wolf, F. A. & Theis, F. J. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat. Biotechnol.* **38**, 1408–1414 (2020).
78. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R Package for comparing biological themes among gene clusters. *OMICS J. Integr. Biol.* **16**, 284–287 (2012).

## Acknowledgements

This work was supported by EMBL; Takeda Science Foundation (to M.E.); the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 101002564) (to M.E.); Project PGC2018-097872-A-I00 (MCIU/AEI/FEDER, UE) funded by the Spanish Ministry of Science, Innovation and Universities (MCIU) and co-funded by the European Regional Development Fund (ERDF, EU) (to M.E.); PRESTO (grant number JP20332265) from Japan Science and Technology Agency (JST) (to M.M.); and fellowships from the Japan Society for the Promotion of Science (JSPS) and Daiichi Sankyo Co., Ltd. (to M.S.-M.). We are grateful to K. Stapornwongkul, C. Villava, S. Xu, K. Anlas, M. Costanzo, and J. Cotterell for their comments, to the Mesoscopic Imaging Facility (MIF) at the European Molecular Biology Laboratory (EMBL), the Advanced Light Microscopy Unit (ALMU) at the Center for Genomic Regulation (CRG), Perkin Elmer, Olympus, and Luxendo/Bruker for imaging supports, to the genomics core facility at the Universitat Pompeu Fabra (UPF), the Genomics Core (GeneCore) facility at EMBL, and the Genome Biology Computational Support (GBCS) at EMBL for sequencing supports, to J.L.Lim, K. Anlas, and G. Martínez-Ara for experimental advice, to the Virtual Human Embryo (J. Cork) and the Kyoto Collection (C. Uwabe and S. Yamada at Congenital Anomaly Research Center, Kyoto University Graduate School of Medicine) for the images of human embryos, and to J. Veenvliet for the images of mouse T.L.Ss.

## Author contributions

M.S.-M., M.M., and M.E. designed the work. N.G. and V.T. contributed to image acquisition and analyses. F.N., M.S.-M., N.G., and J.S. performed scRNA-seq analyses. M.M. optimized the cell cultures. M.S.-M. performed most experiments and analyzed the data. M.E. supervised the work.

**Funding**

Open Access funding enabled and organized by Projekt DEAL.

**Competing interests**

The authors declare no competing interests.

**Additional information**

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-022-29967-1>.

**Correspondence** and requests for materials should be addressed to Marina Sanaki-Matsumiya or Miki Ebisuya.

**Peer review information** *Nature Communications* thanks Jesse Veenvliet and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022