An Epiblast Stem Cell derived multipotent progenitor population for axial extension

Shlomit Edri*, Penny Hayward, Peter Baillie-Johnson, Ben Steventon and Alfonso Martinez Arias*

Department of Genetics
Downing Site
University of Cambridge
Cambridge, CB2 3EH
UK

1. *Authors for correspondence*
Shlomit Edri: se349@cam.ac.uk
Alfonso Martinez Arias: ama11@hermes.cam.ac.uk

Abstract

During vertebrate embryogenesis, the body axis elongates posteriorly through a self renewing population of precursor cells in the tail bud.
The Caudal Lateral Epiblast (CLE) of the mammalian embryo harbours a stem cell/progenitor zone of bipotent progenitors that contribute to both the elongating spinal cord and the paraxial mesoderm of the embryo. These progenitors, called the Neural Mesodermal Progenitors (NMPs), are characterized by the coexpression of *Sox2* and *T*. A number of in vitro studies have been able to produce NMP-like (NMP-l) cells from Pluripotent Stem Cells (PSCs). However, in contrast with embryonic NMPs these in vitro derived ones do not self renew. Here we use different protocols for NMP-l cells and find that, in addition to NMPs, all produce cells with the potential to differentiate into Lateral Plate and Intermediate Mesoderm precursors. We show that Epiblast Stem Cells (EpiSCs) are a better starting source to produce NMPs and show that a specific differentiation protocol yields cells with the potential to self renew in vitro and to make large contributions in xenotransplant assay to axial elongation in both neural and mesodermal tissue. These cells are derived from a population that resembles the Caudal Epiblast (CE) of the embryo at the time of the appearance of the node and we show that a balance between Nodal and BMP signalling is important to define this population. Our study suggests that at the end of gastrulation and associated with the node, a multipotential progenitor population emerges that will give rise to the spinal cord and the mesodermal populations posterior to the brain.

Introduction

The spinal cord defines the path of the vertebrate Central Nervous System (CNS) posterior to the brain, plays a central role in the integration of sensory and motor information and acting as the gateway for this information to the brain. Its structure has been defined in great details and is well conserved across species, with neurons of different identities and functions. The spinal cord is generated from a set of embryonic progenitors distributed along the anteroposterior axis of the embryo. A widespread view posits that these progenitors arise during gastrulation within a neural plate induced by underlying mesoderm shortly after gastrulation and become 'posteriorized' by the activity of Wnt signalling (Niehrs, 2004; Stern, 2005). However, fate mapping and lineage tracing studies suggest that while this process may apply to the brain and the hindbrain, the spinal cord has a different origin (Henrique et al., 2015; Stern, 2005; Steventon and Martinez Arias, 2017).

Anatomically and physiologically it is possible to distinguish between two different tracts in the spinal cord: the thoracic and tail regions. In both amniote and anamniotes the tail tract is generated from cells located in the Chordoneural hinge (CNH), a structure derived from the primitive streak that also gives rise to the more posterior somatic mesoderm (Cambray and Wilson, 2002; Cambray and Wilson, 2007; Davis and Kirschner, 2000; Gont et al., 1993; Steventon et al., 2016). The thoracic tract has different origins in different organisms: in anamniotes e.g. fish and frogs, it is inferred to arise during gastrulation from a pool of pre-existing cells within the epiblast, while in amniotes e.g. chickens and mice, it arises from the expansion of the CLE, a proliferative region located at the caudal end of the embryo, where the primitive streak persists, and acts as source for paraxial, intermediate and lateral plate mesoderm (Henrique et al., 2015; Stern, 2005; Steventon and Martinez Arias, 2017; Sweetman et al., 2008; Wilson et al., 2009). Lineage tracing studies have shown that the CLE harbours a population of bipotential progenitors for neural and mesodermal progenitors located behind the Node, at the Node Streak Border. These cells have been called NMPs, are often characterized by simultaneous expression of *T* (also known as *Bra*) and *Sox2* (Cambray and Wilson, 2007; Wymeersch et al., 2016) and are able of self renewal (Cambray and Wilson, 2002; McGrew et al., 2008; Tzouanacou et al., 2009).

Recently a few studies have claimed the generation of NMP-l cells in adherent cultures of mouse and human embryonic PSCs (Gouti et al., 2014; Lippmann et al., 2015; Turner et al., 2014). In these studies, Embryonic Stem Cells (ESCs) are coaxed into a transient *T* and *Sox2* coexpressing state that, depending on the culture conditions, can be differentiated into either paraxial mesoderm or spinal cord progenitors and their derivatives. However, there is no evidence that these presumptive NMPs are propagated in vitro as they are in the embryo (Tsakiridis and Wilson, 2015). While it has been shown that they can contribute to both somites and spinal cord in a xenotransplant assay (Gouti et al., 2014), these contributions are not extensive and lack the sustained integration in the NMP region of the host embryo (Baillie Johnson et al., 2018).

Coexpression of *T* and *Sox2* might not be a unique characteristic of NMPs as it is a signature of EpiSC populations (Kojima et al., 2014), and this does not imply that EpiSCs, a PSC population, are NMPs. While other markers have been used to identify NMPs in vitro e.g. *Nkx1-2*, *Cdx2*, *Cdh1* and *Oct4*, these are also expressed in the epiblast and in the primitive streak during gastrulation (see Appendix supplementary information file S1), emphasizing the notion that such gene expression signatures are not a unique feature of NMPs. In terms of the ability of *T* - *Sox2* coexpressing cells to differentiate into neural and mesodermal progenitors, it has been shown that exposure of

ESCs to Wnt and FGF signalling can lead to thoracic neural progenitors, perhaps bypassing the NMP state (Mazzoni et al., 2013; Nordstrom et al., 2002; Nordstrom et al., 2006). Altogether these observations raise questions about the identity of the *T* - *Sox2* coexpressing cells derived from ESCs.

Here, we show that *T* - *Sox2* coexpressing cells derived from ESC and EpiSCs based differentiation protocols display differences at the level of gene expression and represent different developmental stages of the transition between naïve, primed pluripotency and neuro-mesodermal fate choices. We find that, in adherent culture, all protocols generate a multipotent population in which NMPs are mixed with progenitors of the Lateral Plate and Intermediate Mesoderm (LPM and IM) as well as the allantois. Moreover, we discover that a new protocol based on EpiSCs leads to an NMP population with many of the attributes of the embryonic NMPs, in particular their ability to self renew and to make long contributions in xenotransplants assays. Our study leads us to propose an explanation for the emergence, amplification and differentiation of this important population in the mammalian embryo.

Results

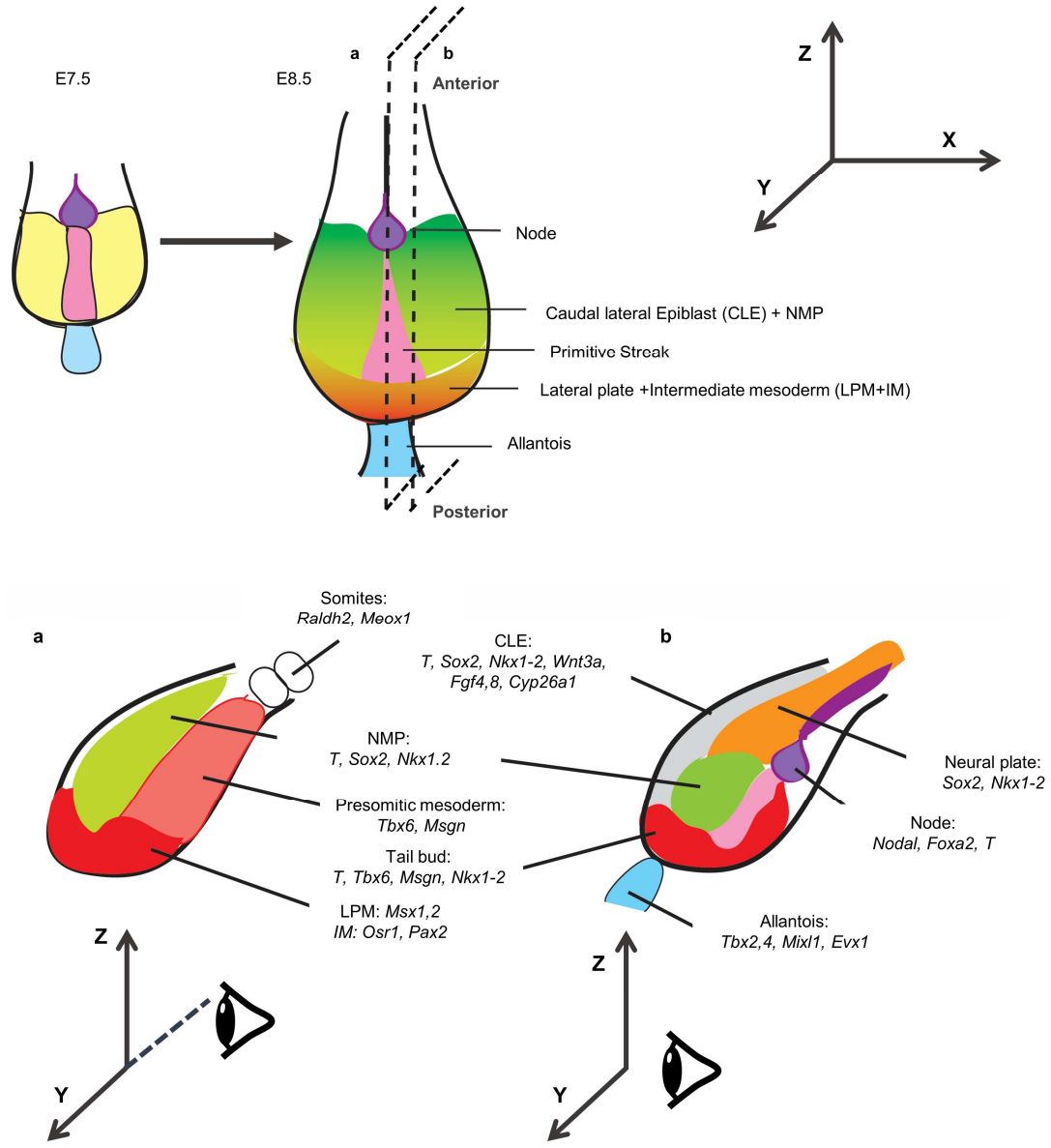

Fig 1. Organization and gene expression patterns in the caudal region of the mouse embryo at E8.5; top, ventral views; bottom: lateral (**a**) and medial (**b**) views. The caudal region of the embryo is derived from the posterior epiblast of E7.5 (yellow) when the primitive streak (pink) reaches the most distal region of the embryo and the node (purple) appears. This region proliferates and undergoes several morphogenetic events which lead to the organization visible at E8.5 and indicated in the figure. The sources for the outlines shown here can be found in Supplementary information file S1.

In order to evaluate whether the NMP-l cells derived in culture correspond to the NMPs in the embryo, we need to identify genetic and functional hallmarks for the comparison. As a first step, we have used data available in the literature (Supplementary information file S1) to create a coarse grained reference map of the CLE at E8.5 (Fig. 1), as this is

the first time at which self renewing NMPs can be identified and tested functionally. These progenitors do not change until the emergence of the CNH at E9.5, and thus are a good reference for our study.

The E8.5 caudal region of the mouse embryo emerges from the posterior epiblast of E7.5 and it is defined as a structure posterior and lateral to the node (Fig. 1). This structure has a clear organization in terms of gene expression and developmental potential (Steventon and Martinez Arias, 2017; Wymeersch et al., 2016), and we have used this to divide the caudal epiblast into a several domains, each with a specific genes expression signature, as can be seen in Fig.1 (see Supplementary information file S1 for details). Two very clear domains can be observed: an anterior to posterior gradient of *Sox2* and *Nkx1-2* expression and a posterior to anterior gradient of *T* expression. The CLE, as mentioned above, harbours the NMPs (Cambray and Wilson, 2007; Wymeersch et al., 2016) and in the most posterior region of the embryo, lie the progenitors of the LPM, IM and the allantois, a mesodermal derivative that will give rise to fetal blood and the umbilical cord. The CLE is identified as much by a gene expression signature as well as by what genes expresses earlier, at E7.0, that are missing at E8.5 e.g. *Cdh1*, *Oct4*, *Otx2* and *Fgf5* (Supplementary information file S1).

EpiSCs yield a postimplantation epiblast population that resembles the CLE

There are several protocols that allow the differentiation of ESCs into NMPs, defined as cells that coexpress *T* and *Sox2*. In all cases, the resulting population can be further differentiated into neural and mesodermal progenitors (summarized in (Henrique et al., 2015)). However, it is not clear whether these NMP-l cells are similar to each other and, importantly, how each relates to the NMPs in the embryo. To begin to answer these questions we compared NMPs obtained from the following three different protocols: two published that start from ESCs, we call them ES-NMPs (Turner et al., 2014) and ES-NMPFs, as in the later the differentiation medium contains FGF (Gouti et al., 2014); and a new one that we have developed and starts from EpiSCs; which we call Epi-NMPs (Fig. 2a; see Methods). In all cases, including the EpiSCs conditions (with or without the Wnt inhibitor, XAV), there are cells coexpressing *T* and *Sox2* both in the mRNA and protein levels (Fig. 2b and Supplementary Fig.S1), however exposure to Chiron (Wnt pathway activator) on the third day in the different NMP protocols posteriorize the cells in the context of the mouse embryo.

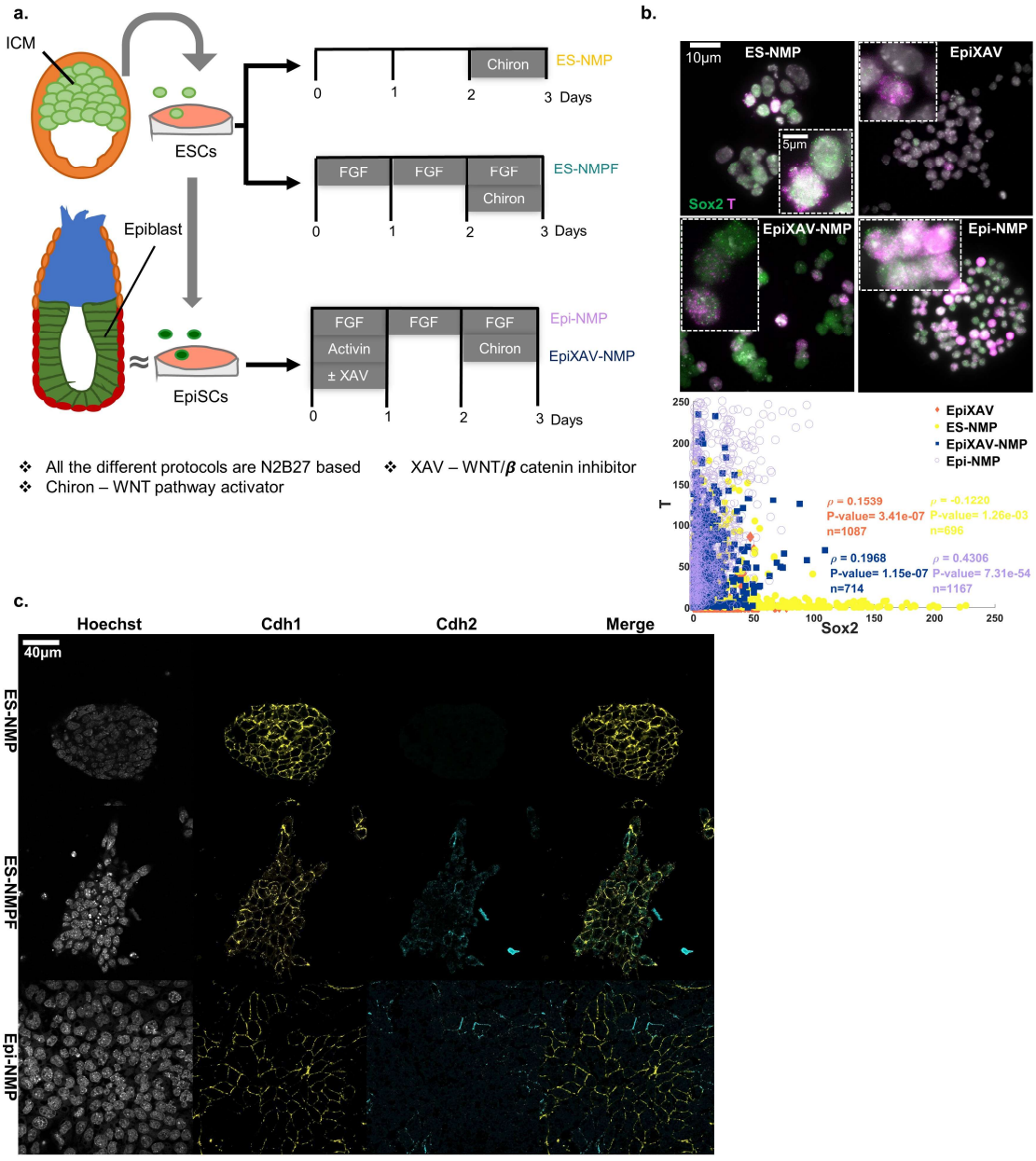

Fig 2. Comparison between in vitro protocols that produce NMP like (NMP-l) cells a. Diagram of the protocols: ES-NMP (Turner et al., 2014) ES-NMPF (Gouti et al., 2014) and Epi-NMP (see Methods). **b.** images, obtained by using single molecule fluorescence in-situ hybridization (sm-FISH), of cells expressing *Sox2* (in green) and *T* (in magenta) mRNA in different conditions: ES-NMP, EpiXAV, EpiXAV-NMP and Epi-NMP. The insets are zoom in on cells coexpressing *Sox2* and *T*. Quantification plot of the number of mRNA molecules in a cell in the different conditions can be found underneath the images. Each dot represents a cell where the x-axis and y-axis represent the number of *Sox2* and *T* molecules respectively in a cell. The Spearman coefficient correlation between *Sox2* and *T*, the P-value and the total number of cells, noted as ρ, P-value and n respectively, can be found as an inset in the quantification plot. **c.** Immunostaining of ES-NMP, ES-NMPF and Epi-NMP on their 3rd day were imaged by a confocal microscopy. Hoechst (dapi) in grey, *Cdh1* in yellow and *Cdh2* in cyan. The composite image of Cdh1 (yellow) and Cdh2 (cyan) is presented on the right-hand side column. All the images were taken using a LSM700 on a Zeiss Axiovert 200 M with a 63× EC Plan-NeoFluar 1.3 NA DIC oil-immersion objective (see Methods).

Measuring the number of transcripts of *Sox2* and *T* in the different populations reveals that they differ in the degree of correlated expression of the two genes (Fig. 2b and Supplementary Fig.S1b). ES-NMP exhibit a negative correlation whereas Epi-NMP exhibit a very high correlation; as a reference, the correlation in EpiSCs treated with XAV (see Methods) is lower. EpiSCs and NMPs derived from EpiSCs populations (EpiXAV, EpiXAV-NMP, and Epi-NMP) exhibit low levels of *Sox2* expression, which is a hallmark of posterior epiblast (Corsinotti et al., 2017). Analysis of the mean and the variability values of *T* expression, highlights Epi-NMP as the population with a higher number of *T-Sox2* positive cells and that they might be found in a transcriptional stable state (low variance of *T* and *Sox2*) in comparison to the other conditions (Supplementary Fig.S1b).

To characterize the different putative NMP populations further, we investigated the expression levels of 37 genes in total throughout this work, out of which we tested 27 genes in the comparison of the three different protocols to produce NMP that we have identified as reflecting the CE (Supplementary Fig. S3). Both of the ES derived NMPs exhibit clear expression of *Cdh1* and *Oct4* and low levels of *Fgf5* and *Otx2*, with ES-NMPF cells expressing high levels of genes associated with mesendoderm differentiation: *Mixl1* (endoderm), *Tbx6* (paraxial mesoderm) and extraembryonic mesoderm (*Evx1*) (Fig. 3b, Supplementary Fig. S3 and see also (Gouti et al., 2014)). This suggests that ES-NMPF contains high numbers of differentiating cells. On the other hand, Epi-NMPs are in a very different state: in addition to *T*, *Sox2* and *Nkx1-2* these cells express significant levels of *Nodal*, *Fgf8*, *Fgf5*, *Foxa2* and *Oct4* together with *Cyp26a1* (see Fig. 3b and Supplementary Fig. S3). This profile suggests that, on average, these cells are in a state that resembles the late epiblast (about E7.5), when the node appears. To better comprehend the gene expression profiles of the different NMP conditions we calculated the NMP index (Fig. 3c, see Methods) and the Epiblast index (Fig. 3d, see Methods), which in a simple manner allow us to understand the potential of each state. The more these indices are close to the diagonal line (Fig. 3c-d) it reflects the duality of the cell population on average. Furthermore, for an ideal NMP population the NMP index should be close to the diagonal line but also in the lower region of both the neural and the mesodermal potentials, avoiding a clear bias to any of the two fates. Looking on Fig 3c-d we can understand that Epi-NMP has the maximum epiblast potential with a slight differentiation towards the mesoderm. The ES-NMP exhibit low epiblast potential with a slight differentiation towards the neural. In contrast, the ES-NMPF shows both high epiblast and mesodermal potential. The differences between the three populations are further emphasized by examination of the protein levels of some of these markers (Fig 2c, Supplementary Fig. S1a and Fig. S2). NMPs derived from ESCs exhibit high levels of *Sox2*, *Oct4* and *Cdh1* expression with some cells expressing *Otx2*. In the case of ES-NMP there is no expression of *Cdh2*, whereas for ES-NMPF we observe a combination at the level of single cells of *Cdh1* and *Cdh2* (Fig. 2c). In contrast, the Epi-NMPs exhibit lower level of *Sox2* and *Oct4* (Supplementary Fig. S1a) and a mutual exclusive expression of *Cdh1* together with *Cdh2* (Fig. 2c), which is a characteristic of the late Epiblast.

All the above highlights that the Epi-NMP are in the late epiblast state, the ES-NMP are in earlier state, most likely still pluripotent, and the ES-NMPF has a bias towards mesodermal differentiation.

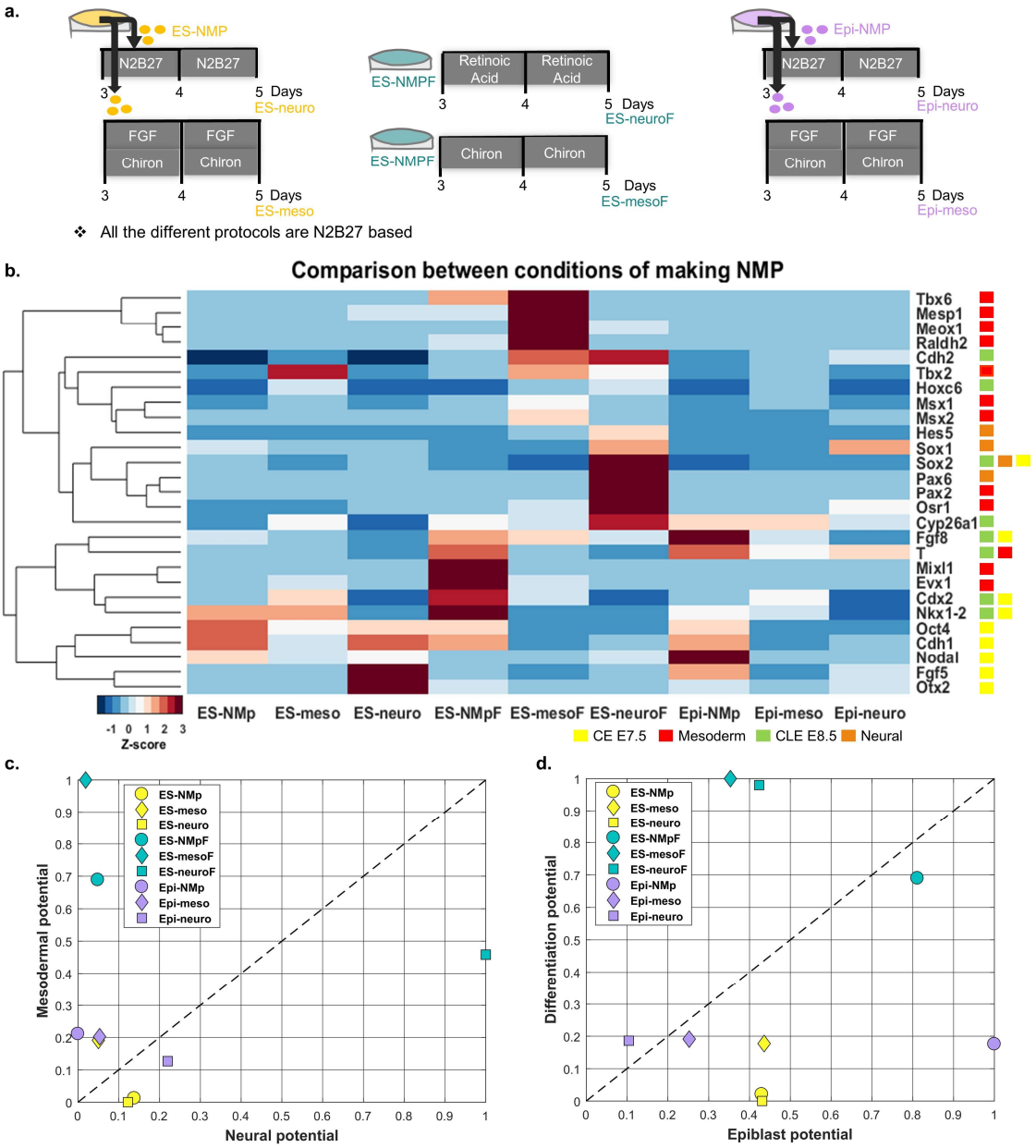

Fig 3. Differentiation NMP-l cells into neural and mesodermal precursors. a. ES cells and EpiSC were grown for 3 days accordingly to the specific protocols (Fig.2a; see Methods), the base medium in all conditions is N2B27. After 3 days, the ES-NMP (yellow) and the Epi-NMP (purple) were split into 2 flasks and cultured for 2 days in a medium that allows differentiation to either neural or mesodermal cells: N2B27 alone for neural or FGF + Chiron for mesoderm (see Methods). In the case of the ES-NMPF (turquoise) we followed the published protocol (Gouti et al., 2014) and did not split/passage the cells, which were grown for 5 days in the same flask in the neural or mesodermal conditions: FGF – FGF - FGF + Chiron - Retinoic acid - Retinoic acid for neural differentiation; FGF – FGF – FGF + Chiron- Chiron - Chiron for mesodermal differentiation. **b.** Expression heatmap of 27 genes, obtained by RT-qPCR, in cells grown in the different conditions, as indicated in Fig. 3a. The normalized expression of each gene across the different conditions was scaled via calculating the Z-score (see Methods and Supplementary Fig. S3). Each gene got a label according to Fig. 1 and Supplementary file S1: CE E7.5 in yellow, mesoderm in red, CLE E8.5 in green and neural in orange. **c.** calculation of the NMP index. In all

the conditions the average expression Z-score value of the mesodermal genes (marked in red) and the neural genes (marked in orange) were calculated and scaled between 0 – 1 across conditions. Those are the Neural potential in the x-axis and Mesodermal potential in the y-axis (see Methods), highlighting the neural-mesodermal state of each condition. **d.** Calculation of the epiblast index. In all the conditions the average expression Z-score value of the differentiating genes (marked in red and orange) and the epiblast genes (marked in yellow and green) were calculated and scaled between 0 – 1 across conditions. Those are the Epiblast potential in the x-axis and Differentiation potential in the y-axis (see Methods), highlighting the epiblast state of each condition

To test the ability of the NMP-l populations to differentiate into mesodermal and neural precursors we set them into specific differentiation conditions (Fig. 3a; see Methods). We named the resulting populations ES-neuro/ES-neuroF and ES-meso/ES-mesoF for those with an ES-NMP/ES-NMPF origin, and Epi-neuro and Epi-meso for those with an Epi-NMP origin. The resulting populations were tested for the expression of our 37/27 test genes (see Fig. 3a-b and Fig. S3). In all cases the cells differentiated but did that in a different manner depending on their origin (see Fig. 3b-d and Supplementary Fig. S4-6). ES-NMPF cells in contrast to the other NMP protocols, express high levels of mesendodermal associated genes (*Tbx6*, *Mixl1* and *Evx1*) and the profile does not change much when cells are differentiated into mesoderm (ES-mesoF) confirming the mesodermal bias of this population (Fig. 3b-c). Moreover, when these cells are placed in neural differentiation conditions, the NMP index of the ES-neuroF still has high mesodermal potential together with the neural one (Fig. 3c) and we observe the emergence of a mixed set of markers associated with LPM and IM (Pax2, Osr1) and neural fates (Pax6 and Sox1), demonstrating the mesodermal bias that ES-NMPF population has (Fig. 3b). The ES-neuro seems to resemble the ES-NMP by showing very similar epiblast index with a slight bias towards neural potential, whereas the ES-meso exhibit the same epiblast potential with a slight shift to the mesodermal potential (Fig. 3c-d). Epi-meso and Epi-neuro exhibit lower epiblast index in comparison to Epi-NMP, with a small bias towards mesodermal and neural fate respectively, however close to the NMP diagonal in the lower region of the neural and mesodermal potential (Fig. 3c-d). Epi-meso still shows epiblast potential (not entirely differentiated) along with the NMP index we are looking for.

Altogether these results suggest that different protocols yield related but different NMP-l populations which might have different functional properties.

Developmental staging of in vitro derived NMP populations

The differences between the candidate NMP populations derived in vitro suggest that they might represent different stages of the transition between the postimplantation epiblast and the CLE. To test this, we created a developmental stage reference using a microarray study of the epiblast at different embryonic stages between early postimplantation (E5.5) and pre-CLE (E8.25) (Kojima et al., 2014), and mapped the NMP populations, as well as their differentiated cells, onto it. A Principal Component Analysis (PCA) of these data using the 27 genes selected for our study (Supplementary Fig. S7c) enabled us to recapitulate, mainly along PC1, the trajectory obtained using the complete microarray data, thus validating the use of our reduced set of genes for mapping states (Supplementary Fig. S7c; see Methods). With this as a reference, we observe that in PC1-3 space, ES-NMP and Epi-NMP and their differentiation protocols mapped closely to the different embryonic stages, whereas the ES-NMPF and its neural and mesodermal differentiation populations are separate from these conditions and from the embryonic stages (Supplementary Fig. S7c). The Epi-NMP and its derivatives projected closely to

each other within the embryo trajectory between the LMS and early bud (EB) stages. The ES-neuro population seems to be comparable to the early stages of the embryo epiblast (projected closely to the cavity (CAV) and PS stages), again demonstrating the early staging of the ES-NMP.

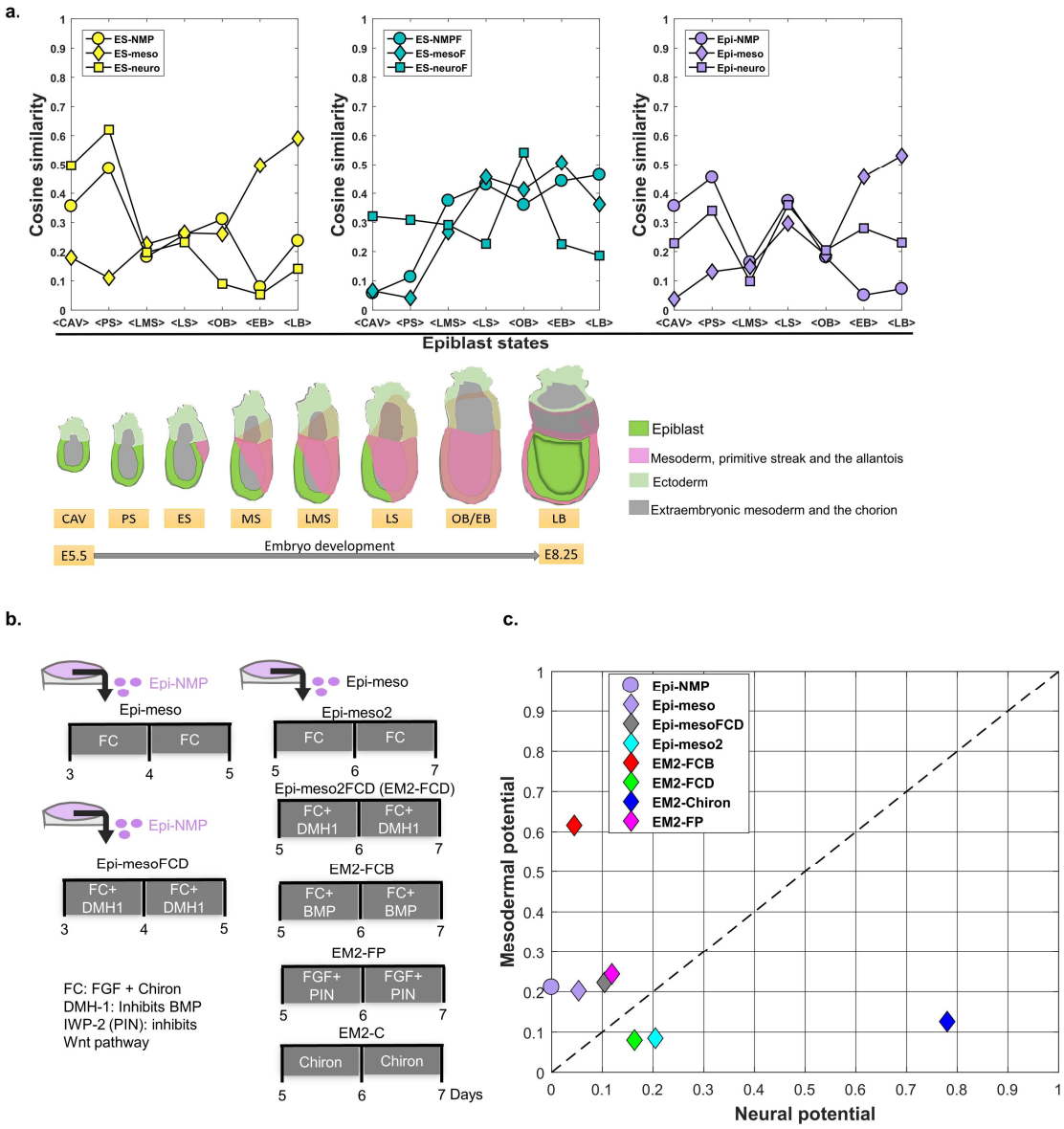

Fig 4. Comparison of the in vitro protocols to different epiblast stages and the effect of Wnt, FGF and BMP on mesodermal differentiation of the Epi-NMP population. a. Microarray gene expression data of the epiblast/ectoderm (excluding the primitive streak) from different stages of the mouse embryo (Kojima et al., 2014) was used as an anchor to compare the different protocols to the embryo; staging is shown underneath the similarity plots (after Kojima et al. 2014), where green indicates the epibast/ectoderm; pink the mesoderm, the primitive streak, and the allantois and grey the extraembryonic mesoderm and the chorion: CAV, cavity; PS, pre-streak; ES, early-streak; MS, mid-streak; LMS, late mid-streak; LS, late streak; OB/EB, no bud/early bud; LB, late bud. The pairwise cosine similarity measure was calculated based on the expression of the 27 genes shown in Fig. 3b between the NMP in vitro protocols or their

differentiation and the different stages of the epiblast mouse embryo (see Supplement Fig. S7a and Methods). Here y-axis represent the average cosine similarity across the same epiblast mouse stages (x-axis) as presented in Supplement Fig. S7a. Value of 0 indicates dissimilarity and value of 1 indicates maximal similarity (see Supplement Fig. S7a and Methods). **b.** Differentiation protocols for Epi-NMP into Epi-meso and Epi-meso2 with modulation of BMP signalling (DMH-1 as an inhibitor and BMP4 as an agonist), Wnt signalling (IWP-2 an inhibitor of Wnt secretion as a Wnt signalling antagonist) and without the addition of exogenous FGF (see Methods). **c.** NMP index of the Epi-NMP and its differentiation protocols was calculated based on the Z-score of the normalized expression profile obtained by RT-qPCR shown in Supplementary Fig. S8c.

We also used our developmental reference to explore the proximity of the in vitro derived populations to specific epiblast states in vivo by finding a gene expression metric. To do this we calculated the cosine similarity between the different cells populations and the different epiblast stages as shown in Fig. 4a and Supplementary Fig. S7b (see Methods). This similarity score suggests that the ES-NMP and ES-NMPF correlate with different stages: the ES-NMPF population appears to spread over several late stages of epiblast development and, in agreement with the expression analysis, resemble to ES-mesoF. ES-NMP, on the other hand, and together with ES-neuro, shows a similarity to earlier stages than the ES-NMPF. The Epi-NMP population appears not to show a clear similarity to any stages. ES-meso exhibit high similarity to the EB and late bud (LB) stages, which range between E7.75-E8.25. Epi-meso, interestingly resembles only the LB stages, which correspond to E8.25, the stage where the CLE harbours the NMP.

Altogether these results show that different starting conditions and differentiation protocols lead to populations with different identities and representations. An interpretation of these is that ES-NMP represents a heterogeneous early population, while ES-NMPF is a heterogeneous population dispersed over several stages with much differentiation. Epi-NMPs appear to be a tighter population resembling a later epiblast. The NMP-l population derived from Epi-NMP, here stated under the name of Epi-meso, is the one closer to the CLE in the embryo, where the NMPs are established. Moreover, in the protein level the Epi-meso exhibits low level of *Sox2* and no *Oct4* (Supplementary Fig. S1a) and a mutual exclusive expression of *Cdh1* together with *Cdh2*, where most of the cells express *Cdh2* rather than *Cdh1* (Supplementary Fig. S2). All this is a clear characteristic of the transition from late Epiblast, represented by Epi-NMP, to CLE.

Multiple tailbud fates emerge from differentiating ESCs and EpiSCs in culture

In the course of our survey of cell type markers in the different populations, we noticed that all protocols that lead to coexpression of *T* and *Sox2* also lead to the expression of genes not associated with NMPs e.g. *Mesp1*, *Evx1*, *Mixl1*, *Gata6*, *Bmp4*, *Msx1*, *Msx2*, *Osr1*, *Pax2* and *Tbx2* (Fig. 3b and Supplementary Fig. S3 and see (Amin et al., 2016; Gouti et al., 2014)). A survey of the literature shows that in the embryo between E7.0 and E8.5, roughly the stage of the differentiating cells, these genes are expressed in the CE and in the posterior region of the primitive streak in the progenitors of the allantois (*Tbx2*, *Tbx4*, *Mixl1* and *Evx1*), the LPM (*Msx1* and *Msx2*) and the IM (*Pax2*, *Osr1*) (Fig. 1 and Supplementary information file S1). This suggests that the three populations are not restricted to harbour merely NMPs and to differentiate to spinal cord and paraxial mesoderm progenitors, but rather that they represent multi-potential populations with different degrees of bias.

In the embryo, the differentiation of the CE is under the control of BMP signalling, that favours more posterior fates (LPM, IM and allantois progenitors) at the expense of more anterior, i.e. NMPs. To test this, we altered the levels of BMP on the Epi-meso population (Fig 4b-c, Supplementary Fig. S8). Inhibition of BMP signalling at the stage of

Epi-meso elevates the expression of NMP markers *T*, *Sox2* and *Cdx2* (Epi-meso versus Epi-mesoFCD in Supplementary Fig. S8c) and slightly its NMP potential (the NMP index moves towards the diagonal of the neuro-mesodermal space as can be seen in Fig. 4c). When passaging Epi-meso to make Epi-meso2 to maintain the NMP signature in culture (see Fig. 4b), addition of BMP to this population (EM2-FCB) elevates dramatically the mesodermal potential (Fig. 4c) and specifically increases the levels of the genes on the posterior fates: *Bmp4*, *Msx1*, *Msx2* and *Tbx2* together with *Cdx2* and *Snail*, compared to Epi-meso2 and EM2-FCD (Epi-meso2 with BMP signalling inhibition, Supplementary Fig. S8a, c). EM2-FCD, similarly to Epi-mesoFCD, slightly improves its NMP index in comparison to Epi-meso2.

These results suggest that in all cases, differentiation of PSCs into a caudal population does not result in the specification of NMPs only, but rather of multipotent populations for all axial derivatives. The differences between the different protocols might not only result in different stages of development but also in different proportions of the different mesodermal populations.

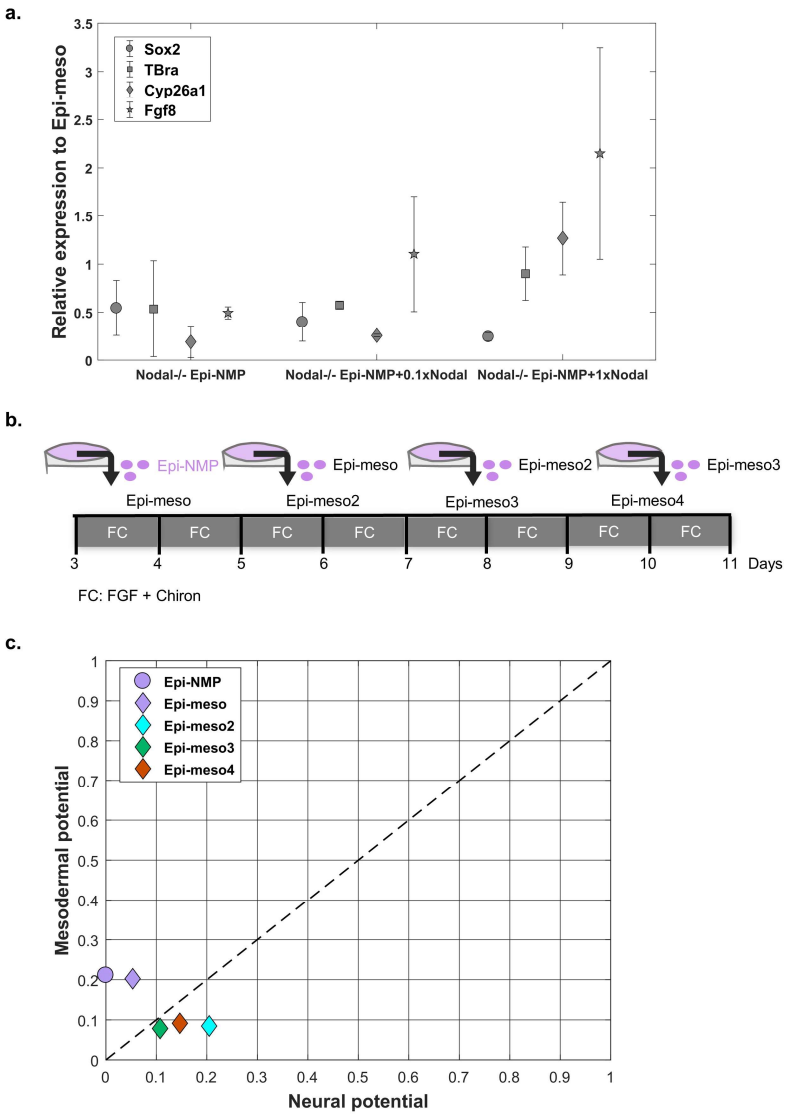

Fig 5. The effect of Nodal signalling and the maintenance of Epi-NMP in culture. **a.** Nodal null cells were cultured under the Epi-NMP protocol (Nodal -/- Epi-NMP). This population of cells was compared to the same population just with the addition of 2 doses of Nodal concentration to the growing medium of the Epi-NMPs on their 3rd day: FGF, Chiron and either 100ng/ml of Nodal (Nodal-/- Epi-NMP+0.1xNodal) or 1µg/ml of Nodal (Nodal-/- Epi-NMP+1xNodal). The different shaped dots represent the genes average expression across biological replicas obtained by RT-qPCR and the error bars indicate the standard error between those replicas. The gene expression across the different conditions was normalized to Epi-meso condition (see Methods). **b.** Differentiation protocol for Epi-NMP into mesodermal precursors; cells were passaged and cultured in FGF and Chiron at every passage to generate the different generation of Epi-meso cells (Epi-meso, Epi-meso2, Epi-meso3, etc. see Methods). **c.** The NMP index of the Epi-NMP and its precursors was calculated based on the Z-score of the normalized expression profile obtained by RT-qPCR shown in Supplementary Fig. S9b.

In the embryo, the anterior region of the CLE is dominated by the node and we observe the expression of *Nodal* and *Foxa2* in the Epi-NMP population (see Fig. 3b and Supplementary Fig. S3a). We have observed before that suppressing Nodal signalling

reduces the *T* and *Sox2* coexpressing cells (Turner et al., 2014) and wondered whether Nodal signalling might influence the NMP-l signature of this population. To test this, we cultured Epi-NMP from Nodal mutant EpiSCs (Nodal -/- Epi-NMP see Methods and Fig. 5a) and compared it to Nodal mutant Epi-NMPs supplemented with 2 different doses of Nodal in the culturing medium on the third day: FGF, Chiron and either 100ng/ml of Nodal (Nodal-/- Epi-NMP+0.1xNodal) or 1µg/ml of Nodal (Nodal-/- Epi-NMP+1xNodal). Loss of Nodal leads to an increment of *Sox2* and a decrease of *Cyp26a1* and *Fgf8* expression. The higher concentration of Nodal seems to rise the levels of *T*. These results suggest that Nodal signalling is necessary to maintain the relative levels of *Sox2* and *T*, mainly low *Sox2* and high levels of *Fgf8* and *Cyp26a1*, which are characteristic of the CE.

Epi-NMPs create a population that can be propagated in vitro

In the embryo, the initial NMP population needs to be amplified if it is to account for the cellular mass along the length of the region posterior to the brain (Steventon and Martinez Arias, 2017; Wymeersch et al., 2016) and this amplification should be a criterion to identify the NMPs in vitro. Earlier studies have shown that ESC derived NMPs are not able to maintain the *T*, *Sox2* coexpressing cells when passaged into the conditions in which they were generated: FGF and Chiron or Chiron alone (Gouti et al., 2014; Turner et al., 2014) and unpublished observations). We were surprised to observe that when Epi-NMPs are induced to differentiate into mesoderm by exposure to FGF and Chiron, they maintain *T* and *Sox2* expression for at least two passages (see Fig. 5b-c and Supplementary Fig S9). We call these populations Epi-meso, Epi-meso2, Epi-meso3 (EM1, EM2, EM3…) and have explored their identity following the expression of our collection of markers. When analysing the NMP index (Fig. 5c), there is a shift from the mesodermal to the neural side when comparing Epi-meso to its derivatives (EM2, EM3 and EM4) however still close to the neural-mesodermal diagonal and in the lower region of both the neural and mesodermal potential, suggesting that the NMP potential can be kept in culture with a slight bias towards neural.

During the passages, we observe a progressive decrease in the expression of NMP markers (*T*, *Cyp26a1*, *Fgf8* and *Nkx1-2*) and a slow increase in the expression of differentiation genes associated with neural fates: *Cdh2*, *Sox2*, and *Hes5* (Supplementary Fig. S9). This is surprising since, while Epi-NMP cells differentiate well into neural progenitors when placed in N2B27, we expected them to differentiate into mesoderm when exposed to Wnt and FGF signalling and not exhibit a bias towards a neural fate. When Epi-meso is grown in N2B27 supplemented with Chiron alone (EM2-Chiron, Fig. 4b-c and Supplementary Fig. S8b-c), we observe high expression of neural markers (*Sox1*, *Sox2* and *Hes5*) and its NMP index is dramatically shifted to the neural side with low mesodermal potential. Furthermore, inhibition of Wnt in the Epi-meso2 state (EM2-FP, Fig.4b-c and Supplementary Fig. S8b-c) reduces the expression of neural progenitor markers and elevates the expression of mesodermal ones (*Gata6* and *Snail1*), which appropriately shift its NMP index to the mesodermal side with low neural potential in comparison to Epi-meso2. Wnt and FGF create a signalling environment comparable with the CLE and our observations suggest that, when applied to Epi-NMPs, they induce a population that exhibit short term self renewal progenitors (EM1, EM2…) that resembles the NMPs. These observations are consistent with the report that, in the context of FGF, Wnt signalling promotes proliferation of the CLE/NMP population (Wymeersch et al., 2016) and does not affect neural specification nor *Sox2* expression (Garriock et al., 2015). Furthermore, in the transition from Epi-NMP to Epi-meso, cells lose expression of epiblast markers e.g. *Fgf5*, *Nodal*, *Otx2*, *Oct4* and *Cdh1* (see Fig. 3b, Fig. S3A and protein expression of *Oct4*, *Otx2* and *Cdh1* in Fig. 2c and Supplementary

Fig. S1a and Fig.S2), hence this suggest that Epi-meso population is a refined state containing many features of the NMPs that are a subset of the CLE.

Epi-NMPs and the Epi-meso contribute to axial extension

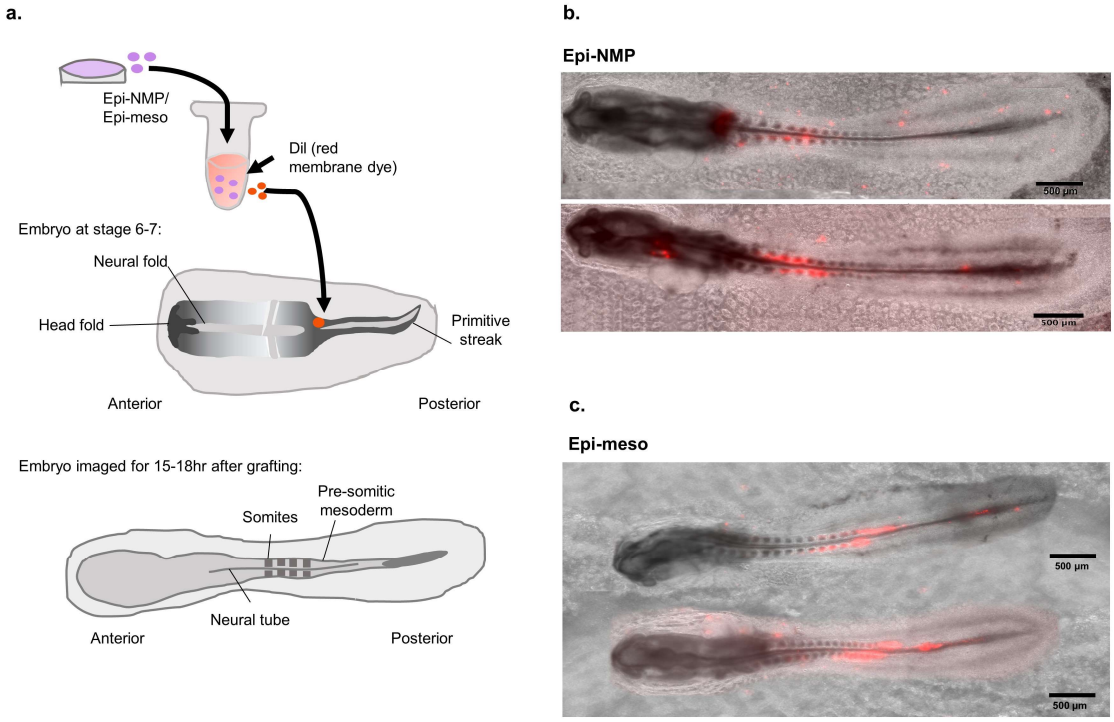

Fig 6. Epi-NMP and Epi-meso are competent to produce NMP-l contributions on transplantation in a chicken embryo. **a.** Scheme of grafting cells into chicken embryos: colonies from Epi-NMP and Epi-meso were labelled with membrane dye (DiI). The labelled tissues were grafted into the region of the caudal lateral epiblast of a chicken embryo around stage 6-7. Embryo cultures were imaged as single time points and as time-lapses for 15-18 hours after grafting (see Methods). **b-c.** Labelled Epi-NMP (b.) and Epi-meso (c.) (marked in red) were transplanted as explained in Fig. 6a. The red cells represent the contribution of the Epi-NMP/Epi-meso grafts in the chicken embryo after 15-18hr since grafting.

A most important feature of a cell type is its ability to fulfil a function in a developing embryo. In the case of mouse NMPs, in addition to their ability to differentiate into neural and mesodermal progenitors, they contribute to axial elongation (Amin et al., 2016; Cambray and Wilson, 2002; Tzouanacou et al., 2009). Therefore, NMP-l populations should be tested for this ability in vivo. While the most stringent test would be the transplantation of these cells into mouse embryos (Gouti et al., 2014), chicken embryos have been shown to be reliable hosts for the functionality of neural and mesodermal progenitors (Fontaine-Perus et al., 1997; Fontaine-Perus et al., 1995; Gouti et al., 2014). Application of this technique to ESCs and different ESC derived NMPs confirms that it is a reliable assay of developmental potential and shows that ESC derived NMPs have some ability to differentiate into neural and mesodermal precursors but do not make long contributions to axial elongation (Baillie Johnson et al., 2018; Gouti et al., 2014). When we use Epi-NMP and Epi-meso in this assay, we observe significant axial contributions of both populations in the chicken embryo that contrast with the limited and discontinuous of the ESC derived NMPs. Significantly, in the case of the Epi-meso we

observe contributions to presomitic, somatic, preneural and neural tissues (Fig. 6). We also observe an association of the transplanted cells with the node region of the host, a characteristic of the NMPs in the embryo and as likewise can be observed in the positive control: transplant of the chicken node in a host embryo (Baillie Johnson et al., 2018).

These results suggest that, in this assay, Epi-NMP and Epi-meso behave more like the NMPs of the embryo than the ESC based NMP-like cells.

Discussion

A large number of studies over the last few years have shown that PSCs represent excellent models for the study of cell fate decisions in development as their differentiation follows paths and temporal schedules similar to those in embryos (Keller, 2005; McCauley and Wells, 2017; Turner et al., 2016). In particular, the experimental versatility of the culture system allows an exploration of the processes in a manner that is not possible in the embryo and these studies have shown that while there are events and features of the differentiation process that occur in embryos which are not represented in culture, most if not all that happens in culture has a correlation with the embryo. However, for these studies to be significant it is important to establish the correlation between events in vitro and the events in the embryo. Here we have used ESCs and EpiSCs to study, in vitro, the emergence of a population of bipotential progenitors, NMPs, that, in the embryo, give rise to the paraxial mesoderm and spinal cord of the thoracic tract. Our results show that each of these protocols produce populations of cells with different gene expression signatures, functional ability and with two common denominators: coexpression of *T* and *Sox2* as well as of a number of genes associated with LPM, IM and allantois i.e. all protocols produce multipotent populations. Therefore, the coexpression of *T* and *Sox2* is not a univocal criterion to identify NMPs. We believe that the identification of these cells requires additional criteria, in particular self renewal and the ability to make long contributions to axial extension in association with the node (Gouti et al., 2015; Henrique et al., 2015; Steventon and Martinez Arias, 2017; Wilson et al., 2009). Applying these to differentiating PSCs populations we identify a specific EpiSC based protocol that yields NMPs like those of the embryo in terms of cellular function and gene expression. This population emerges from a late epiblast like state and can also give rise to LPM, IM and extraembryonic mesoderm in a signalling dependent manner. ESC based protocols yield similar cells that can be differentiated into mesodermal and neural progenitors but lack several features characteristic of NMPs. Furthermore, as we have shown here, these cells appear to be spread throughout several, and different, epiblast stages and might represent different populations. It has been suggested recently that differentiating cells can reach a final state through a variety of paths (Briggs et al., 2017) and this might be also the case for the ESC derived NMPs. However, these studies are based, mainly, on patterns of gene expression and whether the differentiated cells behave as bona fide neural and mesodermal progenitors posterior to the brain, will require additional functional tests.

In the case of NMPs there are many studies in which Wnt signalling can caudalized epiblast like populations (Amin et al., 2016; Mazzoni et al., 2013; Neijts et al., 2016; Nordstrom et al., 2002; Nordstrom et al., 2006) but in these cases, which are mostly ESC based, the NMP-l cells fail to self renew as they do in vivo. In contrast to the ESCs based protocols, here we have shown that exposure to FGF and Chiron of pre-treated EpiSCs generates a population with a gene expression signature characteristic of a late CE, around the time of the appearance of the node. Upon further exposure to Wnt and

FGF signalling, this population evolves and generates cells with many of the hallmarks of the NMPs, including limited but robust self renewal as well as ability to differentiate into neural and mesodermal progenitors in a Wnt dependent manner and long contributions to axial extension in a xenotransplant assay. For these reasons, we shall call the Epi-NMP population Epi-CE, and the Epi-meso Epi-NMP. A significant feature of the Epi-CE, that is in common with the ESC derived populations, and, also of the Epi-NMPs is the expression of markers for LPM, IM and allantois, suggesting that the NMPs are derived from a multipotent population that is likely to also exist in the embryo. We find that the fate of the Epi-CE cells is dependent on a balance between BMP and Nodal signalling and has a strict requirement for Wnt signalling in both neural and mesodermal lineages. The dependence on BMP and Nodal mirrors events in the embryo where BMP signalling promotes posterior (LPM, IM and allantois (Sharma et al., 2017)) and Nodal anterior (NMPs) fates. The importance of Nodal in the establishment of the multipotent population and perhaps also in the definition of the NMP domain is underpinned by our studies with mutants. Consistent with this, embryos mutant for *Foxa2* which lack a node, exhibit deficiencies in the organization of the CLE and axial elongation (Ang and Rossant, 1994; Weinstein et al., 1994) and the same can be observed in embryos mutant for *Smad2* and *Smad3* (Vincent et al., 2003).
It is interesting that the protocol that we use to derive Epi-CE and Epi-NMP is based on a protocol used to derived these progenitors from human ESCs which have been shown to partially self renew (see Methods, Fig. 5b and (Lippmann et al., 2015)). EpiSCs are closer to human ESCs than mouse ESCs and therefore this observation emphasizes the possible importance of the initial state of the population for the paths and outcomes of differentiation

The transition between the Epi-CE and Epi-NMP is linked to the loss of expression of several genes that are associated with the epiblast e.g. *Fgf5*, *Otx2* and, specially, *Oct4*, a POU domain transcription factor that, together with *Sox2*, maintains pluripotency. A similar transition can be observed in the embryo where *Oct4* expression ceases at around E8.5/9.0 (Downs, 2008; Osorno et al., 2012), the time at which cells start differentiating. It is possible that the combination of *Oct4* and *Sox2* promotes multipotency and that only when *Oct4* expression ceases then *Sox2* starts playing a pro-neural role. A function for *Oct4* in axial elongation can be gauged from the severe axial truncations that follow loss of *Oct4* activity from E7.0/7.5 (DeVeale et al., 2013) and the extended axial elongations associate with overexpression of *Oct4* (Aires et al., 2016) which may reflect an increase in the initial size of the multipotent CE pool rather than a specific alteration in the NMP population.

During the passage of the Epi-NMP population in the presence of Wnt and FGF signalling, we notice that the cells progressively lose *T* expression and increase *Sox2*. This is surprising since Wnt signalling is thought to suppress neural differentiation and promote mesodermal. However, while in the embryo this is true during the first phase of gastrulation, before E7.5, it might not be during the development of the caudal region of the embryo. When the trunk develops, there is a clear evidence that Wnt/β-catenin signalling is required for the expansion of the neural progenitors in the spinal cord (Zechner et al., 2003) and an increase in Wnt/β-catenin signalling do not supress neural development (Garriock et al., 2015). Furthermore, the fact that *Sox2* gene has a Tcf response element (Takemoto et al., 2006) and responds to Wnt signalling, suggests that the dependence of *Sox2* that we observe is associated with the development of neural structures. Altogether these observations can account for our results and contrast with previous reports in which Wnt signalling suppresses neural fates (Gouti et al., 2017; Gouti et al., 2014; Turner et al., 2014). We suggest that the antagonism of Wnt signalling

and *Sox2* expression does not reflect an effect on neural fate but the assignation of posterior identity in the early epiblast (Corsinotti et al., 2017). This would be consistent with the observation that NMP-l cells derived from ESCs reflect early stages of development and that their differentiation bypasses the embryo path. On the other hand, there is ample evidence for a requirement for Wnt/β-catenin signalling in axial elongation (Cunningham et al., 2015; Dunty et al., 2008; Yamaguchi et al., 1999). This requirement might reflect a necessity for the proliferation of the CLE which has been hinted in some studies (Wymeersch et al., 2016).

In summary, we have identified a specific experimental protocol that, starting with EpiSCs yields a multipotent progenitor population for all the structures that are posterior to the brain (Epi-CE) and can be further differentiated in vitro into NMPs (Epi-NMP) or LPM/IM, depending on the signalling environment. Significantly the Epi-NMP population exhibits a limited but robust self renewing ability which, together with its gene expression signature, lead us to believe that it is an in vitro correlation of the in vivo NMPs. Our results suggest that there might not be NMP specific protocols that bypass the multipotent state observed in this work; we have found the signature of this state in the data of all published protocols.

Acknowledgements
This work was supported by Cambridge Trust and Cambridge Philosophical Society scholarships and AJA Karten trust award to S. Edri, a Henry Wellcome Fellowship to Ben Steventon and an EPSRC studentship to P.Baillie-Johnson and BBSRC project grants (No. BB/M023370/1 and BB/P003184/1 to AMA. We are grateful to J Collignon for the Nodal mutant cells and to James Briscoe and Valerie Wilson for discussions.

Materials and Methods

Cell culture

E14-Tg2A were grown in tissue-culture plastic flasks coated with either 0.1% gelatine (Sigma-Aldrich, G1890-100G) in PBS (with Calcium and Magnesium, Sigma-Aldrich, D8662) or 0.5% Plasma Fibronectin (FCOLO, 1mg/ml, Temecula) in PBS (with Calcium and Magnesium). The differentiation protocols are as the following:

ES-NMP

Cells were plated at a density of $4.44x10^3$ cells/cm^2 in a 0.1% gelatine coated flask with a base medium of N2B27 (NDiff 227, Takara Bio) for 2 days. After 48hr N2B27 is supplemented with 3μM of CHIR99021 (Chiron 10mM, Tocris Biosciences) for additional 24hr, which are in total 72hr.

ES- meso and ES-neuro

ES-NMP cells were detached from the culture flask using Accutase (BioLegend 0.5Mm) and divided into 2 flasks coated with 0.5% Fibronectin at a dense of $7.5x10^3$ cells/cm^2. To get ES-neuro the cells were grown for 2 days in N2B27 and for ES-meso the cells were grown in N2B27 supplemented with 20ng/ml FGF2 (R&D systems, 50μg/ml) and 3μM Chiron.

ES-NMPF, ES-neuroF, ES-mesoF (Gouti et al., 2014)

Cells were plated at a density of $5x10^3$ cells/cm^2 in a 0.1% gelatine coated CellBINDSurface dish (Corning) with a base medium of N2B27 supplemented with 10 ng/ml FGF2. After 48hr N2B27 is supplemented with 10 ng/ml FGF2 and 5μM Chiron for additional 24hr, which are in total 72hr. To induce neural SC identity (ES-neuroF) or mesodermal identity (ES-mesoF) day 3 – day 5 was followed by either N2B27

supplemented with 100nM RA (Sigma) or by N2B27 supplemented with 5μM Chiron respectively.

Epi-NMP

E14-Tg2A were grown on a 0.5% Fibronectin coated culture flask with N2B27 supplemented with 12ng/ml FGF2 and 25ng/ml Activin A (Stem Cells Institute 100μg/ml), this known as Epi-media with or without 20μM XAV939 (XAV Tocris Biosciences, 10mM) for at least 4 passages. Those cells known as EpiSCs (or EpiXAV when the β-catenin inhibitor XAV is used). Those cells can be tested to be EpiSC by seeding them in a colony assay density (67 cells/cm^2) in restricted medium (2i: N2B27 supplemented with 3μM Chiron and 1μM PD0325901 (PD03, Tocris Biosciences, 10mM)), resulting in no growth of cells, ensuring that the cells are no longer in the naïve pluripotent state and they moved on to the prime pluripotent state (data are not shown).

EpiSCs (treated with or without XAV) were plated at a 5x10^4 cells/cm^2 density in a 0.5% Fibronectin pre-coated flask with Epi-media for the first day. Day 2 is followed by increasing the concentration of FGF2 to 20ng/ml in the base medium of N2B27 and removing Activin A or XAV if originally was used to grow the EpiSCs. On day 3, N2B27 is supplemented with 3μM Chiron which is added to the 20ng/ml FGF2. After 72hr those cells known as the Epi-NMP or EpiXAV-NMP if XAV was used in the Epi-media. This protocol is a variation of one that has been used to derive NMP-l cells from human ESCs (Lippmann et al., 2015)

Epi- meso and Epi-neuro

Epi-NMP (without XAV) cells were detached from the culture flask using Accutase and divided into 2 flasks coated with 0.5% Fibronectin at a dense of 5x10^4 cells/cm^2. To get Epi-neuro the cells were grown for 2 days in N2B27 and for Epi-meso the cells were grown in N2B27 supplemented with 20ng/ml FGF2 and 3μM Chiron.

Epi-meso differentiation

Epi-meso (without XAV) cells were detached from the culture flask using Accutase and plated back to a 0.5% Fibronectin coated flask at a dense of 5x10^4 cells/cm^2 for 2 days in N2B27 supplemented with 20ng/ml FGF2 and 3μM Chiron. The first passage from Epi-meso is named Epi-meso2 (EM2), the second passage is named Epi-meso3 (EM3) and so forth.

BMP, FGF and Wnt signalling

Epi-mesoFCD

Epi-NMP (without XAV) cells were detached from the culture flask using Accutase and plated back to a 0.5% Fibronectin coated flask at a dense of 5x10^4 cells/cm^2 for 2 days in N2B27 supplemented with 20ng/ml FGF2, 3μM Chiron and 500nM dorsomorphin-H1 (DMH-1 5mM, Tocris Biosciences) which is a BMP inhibitor.

EM2-FCD

Epi-meso (without XAV) cells were detached from the culture flask using Accutase and plated back to a 0.5% Fibronectin coated flask at a dense of 5x10^4 cells/cm^2 for 2 days in N2B27 supplemented with 20ng/ml FGF2, 3μM Chiron and 500nM DMH-1.

EM2-FCB

Epi-meso (without XAV) cells were detached from the culture flask using Accutase and plated back to a 0.5% Fibronectin coated flask at a dense of 5x10^4 cells/cm^2 for 2 days in N2B27 supplemented with 20ng/ml FGF2, 3μM Chiron and 1ng/ml BMP4 (R&D Systems, 100μg/ml).

EM2-Chiron

Epi-meso (without XAV) cells were detached from the culture flask using Accutase and plated back to a 0.5% Fibronectin coated flask at a dense of $5x10^4$ cells/cm^2 for 2 days in N2B27 supplemented with 3μM Chiron alone.

EM2-FP

Epi-meso (without XAV) cells were detached from the culture flask using Accutase and plated back to a 0.5% Fibronectin coated flask at a dense of $5x10^4$ cells/cm^2 for 2 days in N2B27 supplemented with 20ng/ml FGF2 and 1μM IWP-2 (PIN 5mM, STEMGENT) which is a Wnt pathway inhibitor.

Nodal Null cells

Cells mutant for Nodal (Nodal-/-) were provided by J. Collignon and were derived from the 413.d mutant mouse line (Conlon et al., 1991), They were grown on a 0.5% Fibronectin coated culture flask with Epi-media: N2B27 supplemented with 12ng/ml FGF2 and 25ng/ml Activin A for at least 4 passages. The Nodal null EpiSCs were plated in $5x10^4$ cells/cm^2 density on a 0.5% Fibronectin pre-coated flask with Epi-media for the first day. Day 2 is followed by increasing the concentration of FGF2 to 20ng/ml, in the base medium of N2B27, and removing the Activin A. On day 3, N2B27 is supplemented with 3μM Chiron which is added to the 20ng/ml FGF2. After 72hr those cells known as the Nodal-/- Epi-NMP. In order to examine the role of Nodal in establishing the NMPs the Nodal mutant Epi-NMPs were supplemented with 2 different doses of Nodal in the culturing medium on the 3rd day: 20ng/ml FGF2, 3μM Chiron and either 100ng/ml of Nodal (R&D systems, sample name: Nodal-/- Epi-NMP+0.1xNodal) or 1μg/ml of Nodal (sample name: Nodal-/- Epi-NMP+1xNodal) in the base medium of N2B27.

Quantitative RT-PCR (qRT-PCR)

Total RNA was isolated from cells using Trizol. First strand cDNA synthesis was performed with Superscript III system (Invitrogen) and the quantification of double-stranded DNA obtained with specific genes designed primers, using QuantiFast SYBR Green PCR Master Mix (Qiagen) and the standard cycler program (Qiagen RotorGene Q). The qPCR was done in technical triplicates. The primers that have been used are available in Supplementary Material Table S1. All experiments were performed in biological duplicates or triplicates. Expression values were normalized against the housekeeping gene Ppia. To compare different experiments of qRT-PCR, each run of the qPCR included one common condition, in our case it was the Epi-meso. Each condition in every run was normalized to Epi-meso and averaged across biological replicas. Here are the steps to calculate the normalized gene expression values:

1. Identifying the C_t (threshold cycle) for each gene (technical triplicates) and calculating the expression values (2^{-Ct}).
2. Calculating the average and the standard deviation (std) for each gene from the triplicate expression values.
3. Dividing the average and the std of each gene in the expression value of Ppia.
4. The normalized gene expression values in condition x will be divided with the normalized gene expression values in the common condition in every qRT-PCR experiment (Epi-meso) as the following:
   F=x/y, where x denotes the expression of a gene at any condition and y denotes the expression of the same gene at Epi-meso condition. Both x and y have an error, the std that have been calculated in step 3: Dx and Dy, hence the total error is:

$$DF = \sqrt{\left(\frac{dF}{dx}Dx\right)^2 + \left(\frac{dF}{dy}Dy\right)^2} = \sqrt{\left(\frac{1}{y}Dx\right)^2 + \left(-\frac{x}{y^2}Dy\right)^2}$$

5. Averaging the biological replicas: F_1 and F_2 are biological replicas of the same gene in the same condition and their expression was normalized as the above steps. The average of the normalized expression and the error is calculated as the standard error:

$$\bar{F} = \frac{1}{N}\sum_{i=1}^{N} F_i$$

$$SE(F) = \frac{1}{\sqrt{N}(N-1)}\sqrt{\sum_{i=1}^{N}(F_i - \bar{F})^2}$$

where N is the number of replicas (between 2 and 3).

6. Standardizing the normalized expression values of a gene to Z-score values across conditions was done according the equation below:

$$z = \frac{\bar{F} - \mu}{\sigma}$$

Where μ and σ denote the average and the standard deviation of the normalized expression of a gene across all the conditions examined in this work respectively (the average and standard deviation of $\bar{F}$ for a specific gene across all the conditions).

NMP and Epiblast indices

Neural Genes	Mesodermal genes	Epiblast genes
Sox1, Pax6, Hes5, Sox2	*Bmp4, Evx1, Gata6, Meox1, Mesp1, Mixl1, Raldh2, Tbx2, Tbx6, Msx1, Msx1, Pax2, Osr1, Snai1, T*	*Cdh1, Fgf5, Oct4, Otx2, Cdx2, Fgf8, Nodal, Wnt3a, Cyp26a1, Nkx1-2, Hoxc6, Cdh2, Foxa2*

The NMP index was calculated according to the following: In all the conditions (17 in total) the average expression Z-score value of the neural genes and the mesodermal genes (noted in the table above) was obtained and scaled between 0 – 1 across the 17 conditions. This resulted in 2 values for each condition: the neural potential and the mesodermal potential. The epiblast index was calculated in a similar manner: the average of the Z-score expression values of the epiblast genes was calculated versus the differentiation genes (neural and mesodermal, listed in the table above) and scaled between 0 – 1 across the 17 conditions, resulting in an epiblast potential and a differentiation potential for each condition.

Single molecule fluorescence *in-situ* hybridization (sm-FISH)

Single molecule RNA FISH was carried out as described previously (Nair et al., 2015). Cells were dissociated using Accutase, washed in PBS, fixed in 37% formaldehyde at room temperature and permeabilized and stored in 70% ethanol at 4°C. All washes and hybridizations were carried out in suspension. Wash buffers included 0.1% Triton X-100

to minimize losses to sticking on the walls. Samples were mounted between a slide and #1 cover glass, in the glucose-oxidase-based 2 × SSC anti-fade buffer. The probes for the genes (supplementary Table S3) were designed using StellarisTM website and bought via StellarisTM FISH probes (Biosearch Technologies)(Raj et al., 2008). Additional information about how the probes were designed, prepared and used can be found in (Raj et al., 2008). Cells were imaged within 24 to 48 h on a Nikon Ti-E wide field microscope, using a 60X oil-immersion objective and a cooled camera (Orca flash 4.0, Hamamatsu). The cells in the images were segmented manually and the spot-detection was done semi-automatic using a MATLAB graphic user interface (GUI) developed by Marshall J. Levesque and Arjun Raj at the University of Pennsylvania or home-made protocols written in ICY(de Chaumont et al., 2012).

Principal Component Analysis

PCA (Nair et al., 2015) involves the assignment of data, in our case gene expression, to new coordinates named principal components or PCs. The variance of observed coordinates in each PC occurs in a decreasing order, observations (the samples) projected on PC1 have a greater variance than the same observations projected on PC2 and so on. The PCs were calculated according to the Z-scores expression values of the 27 genes expressed (see Fig. 3D and Fig. S1B-C) in the different stages of the mouse embryo epiblast/ectoderm and in the 3 in vitro protocols and their neural and mesodermal differentiation: ES-NMP, ES-NMPF and Epi-NMP.

Cosine similarity

Here we used cosine similarity as a measure of similarity between Z-scores expression values of list of genes in one condition versus other condition (i.e. Epi-NMp versus the mouse embryo epiblast stages (Kojima et al., 2014) per the same list of genes). The cosine similarity was calculated as the following:

$$similarity = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}}$$

Where A and B represents the list of genes with their values of Z-score gene expression in two conditions and A_i and B_i are the components of these two vectors. The similarity was constrained to the positive space, where 0 indicates that the two vectors are dramatically opposite and 1 indicates maximal similarity.

Bright field images

All bright field images were taken on a Nikon Ti-E wide field microscope, using a 20X objective and a cooled camera (Orca flash 4.0, Hamamatsu).

Confocal and immunostaining

In 4-well (Ibidi), plastic tissue-culture dishes the different samples were grown. Samples were washed in BBS + $CaCl_2$ (50 mM BES Sodium Salt, 280 mM NaCl, 1.5 mM Na2HPO4, 1 mM $CaCl_2$ adjusted to pH 6.96 with 1 M HCl) and fixed for 15 minutes in 4% paraformaldehyde. Samples were washed and permeabilised with BBT (BBS + $CaCl_2$ supplemented with 0.5% BSA and 0.5% Triton X-100) before overnight antibody staining, following a wash with BBT and then the samples were incubated for 2hr with the desired fluorescently-conjugated secondary antibody. Prior to imaging, samples were washed with BBS + $CaCl_2$ and covered in a mounting medium (80% spectrophotometric grade glycerol, 4% w/v n-propyl-gallatein in BBS + $CaCl_2$).

The following primary antibodies were used: Brachyury N19 (goat; Santa Cruz Biotechnologies, sc17743, dilution 1:100), Oct3/4 (mouse; Santa Cruz Biotechnologies, sc5279, dilution 1:100), Sox2 (rabbit; Millipore, AB5603, dilution 1:200), Otx2 (goat, R&D AF1979 dilution 1:200), Ncad (mouse, BD Bioscience 610920 dilution 1:200) and Ecad (rat, Takara M108, dilution 1:100). Secondary antibodies (Goat-A488, Rabbit-A633, Mouse-A568, Rat-A633) were from Molecular Probes, made in donkey and used in a 1 in 500 dilution with Hoechst 33342 (H3570, dilution 1 in 1000; Invitrogen ThermoFisher). Samples were imaged using an LSM700 on a Zeiss Axiovert 200 M with a 63× EC Plan-NeoFluar 1.3 NA DIC oil-immersion objective. Hoechst, Alexa488, -568 and -633 were sequentially excited with a 405, 488, 555 and 639 nm diode lasers, respectively. Data capture carried out using Zen2010 v6 (Zeiss).

Chicken Embryo Culture

Fertilised chicken eggs were stored in a humidified 10°C incubator for up to one week until required. Eggs were transferred to a humidified, rocking 37°C incubator for 24 hours prior to the preparation of embryo cultures, which was done according to a modified version of New Culture (New, 1955). Embryo cultures were incubated at 37°C prior to grafting and were fixed in 4% paraformaldehyde within 24 hours.

Graft Preparation and Transplantation: Cell cultures were prepared as described above. Adherent cell cultures were detached mechanically using a cell scraper in PBS (with calcium and magnesium) to lift intact colonies with minimal sample dissociation. The tissues were labelled by transferring them to a FBS-precoated FACS tube and were centrifuged at 170 x *g* for five minutes. The supernatant was discarded and the colonies washed by gentle resuspension in PBS (with calcium and magnesium), before the centrifugation step was repeated. The tissues were then resuspended gently in PBS (without calcium and magnesium) for labelling with DiI (Thermo Fisher Scientific Vybrant® V22885, 1% v/v) for 25 minutes in the dark, on ice. The labelled tissues were centrifuged at 170 x *g* for five minutes and the pellet was gently resuspended in PBS (with calcium and magnesium) for grafting.

Labelled tissues were grafted into the region of the chick caudal lateral epiblast as described in (Baillie Johnson et al., 2018), using an eyebrow knife tool. Embryo cultures were imaged as single time points and as time-lapses for 15-18 hours after grafting.

Embryo Microscopy and Image Analysis

Widefield images were acquired with a Zeiss AxioObserver Z1 (Carl Zeiss, UK) using a 5x objective in a humidified 37°C incubator, with the embryo cultures positioned on the lid of a six-well plate. An LED white light illumination system (Laser2000, Kettering, UK) and a Filter Set 45 filter cube (Carl Zeiss, UK) was used to visualise red fluorescence. Emitted light was recorded using a back-illuminated iXon800 Ultra EMCCD (Andor, UK) and the open source Micro-Manager software (Vale Lab, UCSF, USA). The open source FIJI ImageJ platform (Schindelin et al., 2012) and the pairwise stitching plugin (Preibisch et al., 2009) were used for image reconstruction and analysis.

Supplementary figures

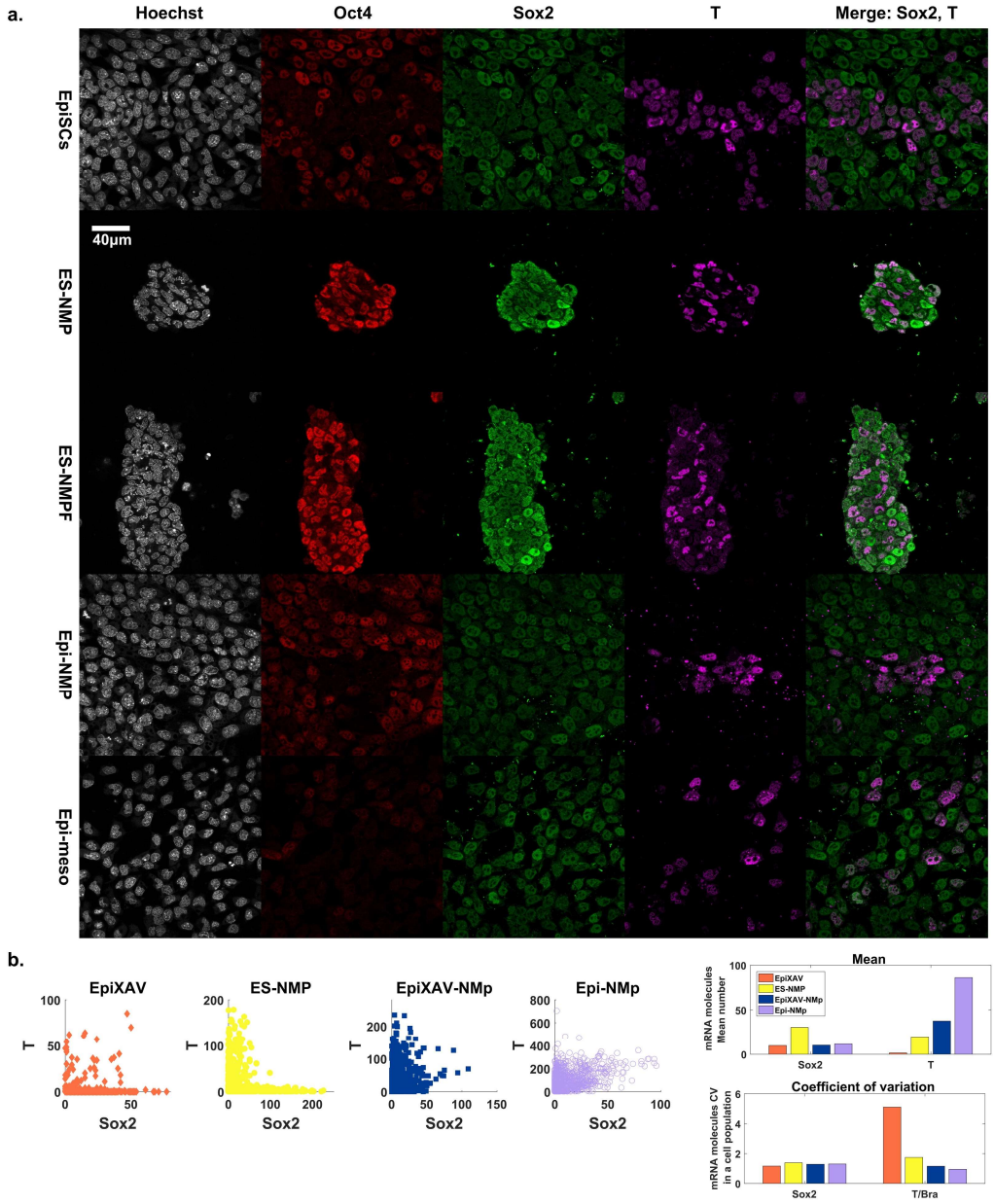

Fig S1. Coexpression of *Sox2* and *T* in the mRNA and protein levels in EpiSC and the different NMP protocols. a. Immunostaining of EpiSC, ES-NMP, ES-NMPF and Epi-NMP on their 3rd day and Epi-meso on its 2nd day were imaged by a confocal microscopy. Hoechst (dapi) in grey, *Oct4* in red, *Sox2* in green and *T* in magenta. The composite image of *Sox2* (green) and *T* (magenta) is presented on the right-hand side column. All the images were taken using a LSM700 on a Zeiss Axiovert 200 M with a 63× EC Plan-NeoFluar 1.3 NA DIC oil-immersion objective (see Methods). **b.** Quantification plots of the number of mRNA molecules in a cell obtained by sm-FISH for the different conditions (Fig. 2b, see Methods); each dot represents a cell where the x-axis and y-axis represent the number of *Sox2* and *T* molecules respectively in a cell. The plots on the right hand side display the mean and the coefficient of variation (CV) of the distribution of *Sox2* and *T* in the different condition: EpiXAV in orange, ES-NMP in yellow, EpiXAV-NMP in blue and Epi-NMP in purple.

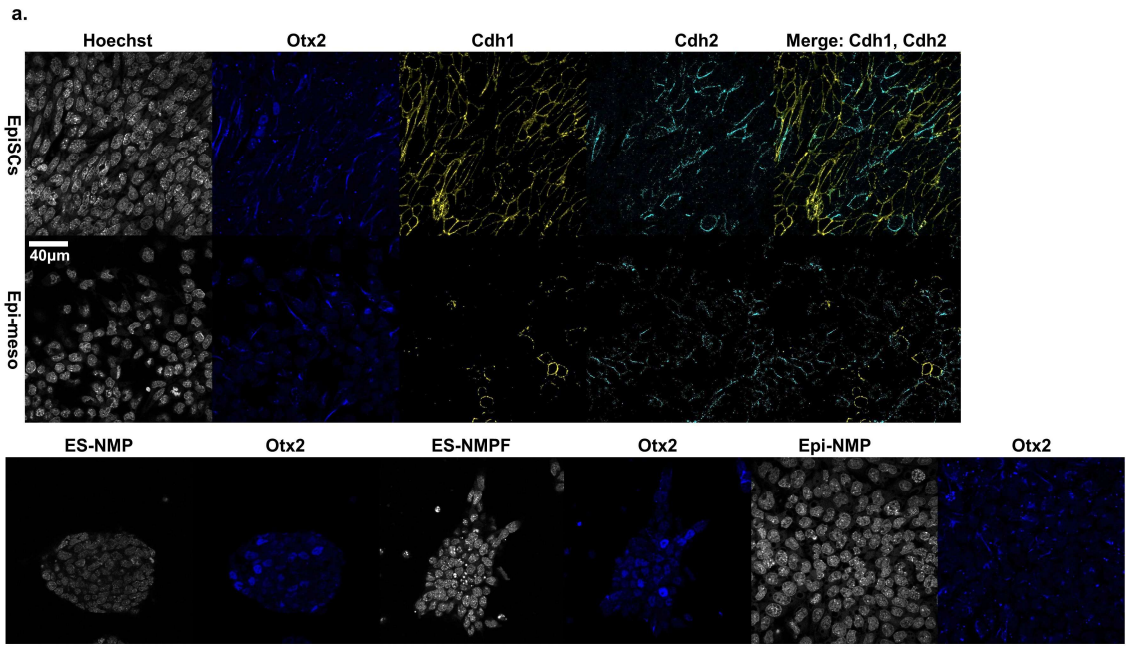

Fig. S2. Immunostaining of EpiSC, ES-NMP, ES-NMPF and Epi-NMP on their 3rd day and Epi-meso on its 2nd day. Hoechst (dapi) in grey, *Otx2* in blue, *Cdh1* in yellow and *Cdh2* in cyan were imaged by a confocal microscopy. The composite image of *Cdh1* (yellow) and *Cdh2* (cyan) is presented on the right-hand side column. All the images were taken using a LSM700 on a Zeiss Axiovert 200 M with a 63× EC Plan-NeoFluar 1.3 NA DIC oil-immersion objective (see Methods).

Fig S3. Comparison of expression of set of genes in the 3 in vitro protocols. **a.** Measuring the expression of 37 genes in ES-NMP and Epi-NMP. **b.** Measuring the expression of 27 genes in: ES-NMP, ES-NMPF and Epi-NMP. The bars represent the genes average expression across biological replicas obtained by RT-qPCR and the error bars indicate the standard error between those replicas. The gene expression across the different conditions was normalized to Epi-meso condition (see Methods).

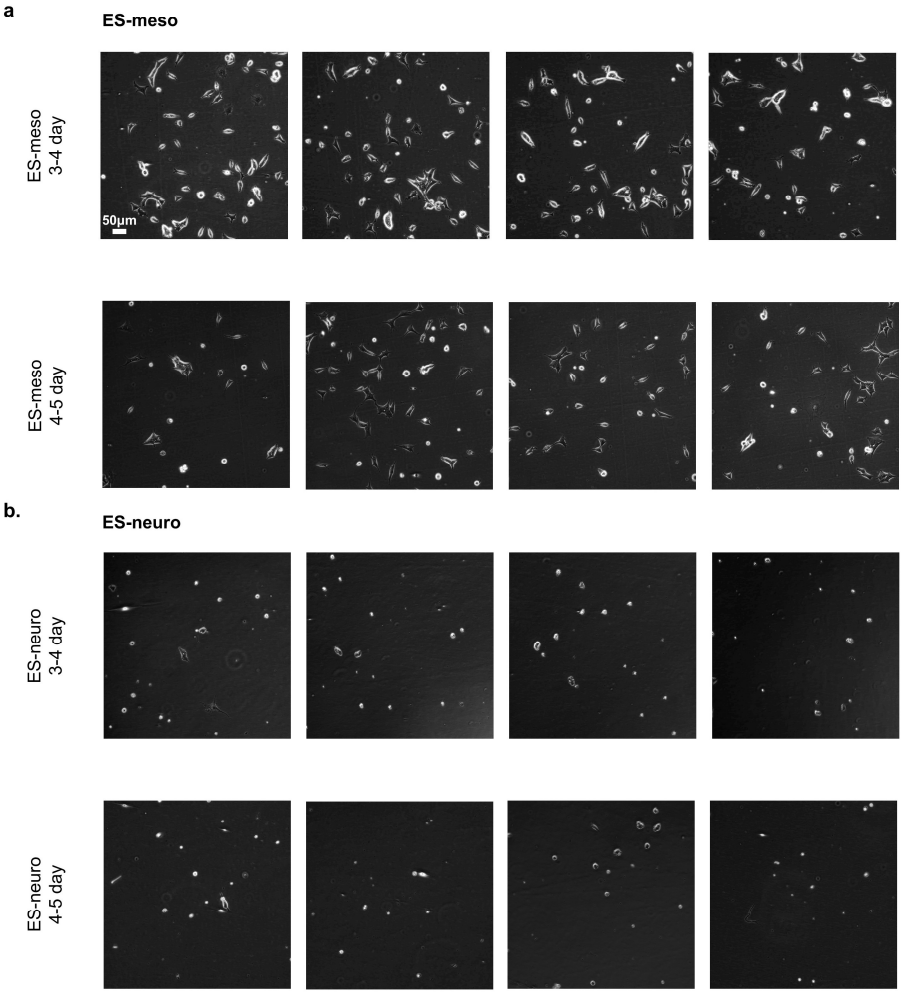

Fig S4. Brightfield images of ES-NMP differentiation. **a.** ES-meso cells were imaged between their 3rd and 4th day (1st row) and 4th and 5th day (2nd row), the cells were treated according to the ES-meso protocol (see Methods). **b.** ES-neuro cells were imaged between their 3rd and 4th day (1st row) and 4th and 5th day (2nd row), the cells were treated according to the ES-neuro protocol (see Methods). All the images were taken on a Nikon Ti-E wide field microscope, using a 20X objective and a cooled camera (Orca flash 4.0, Hamamatsu).

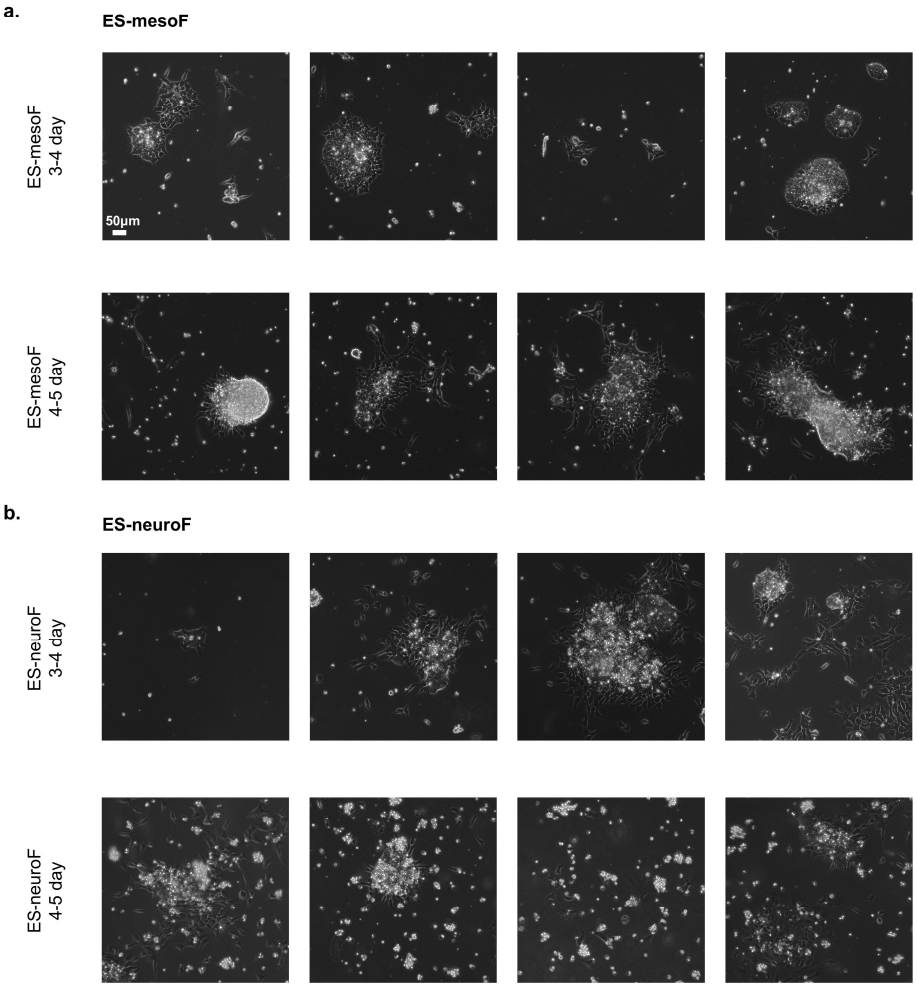

Fig S5. Brightfield images of ES-NMPF differentiation. **a.** ES-mesoF cells were imaged between their 3rd and 4th day (1st row) and 4th and 5th day (2nd row), the cells were treated according to the ES-mesoF protocol (see Methods). **b.** ES-neuroF cells were imaged between their 3rd and 4th day (1st row) and 4th and 5th day (2nd row), the cells were treated according to the ES-neuroF protocol (see Methods). All the images were taken on a Nikon Ti-E wide field microscope, using a 20X objective and a cooled camera (Orca flash 4.0, Hamamatsu).

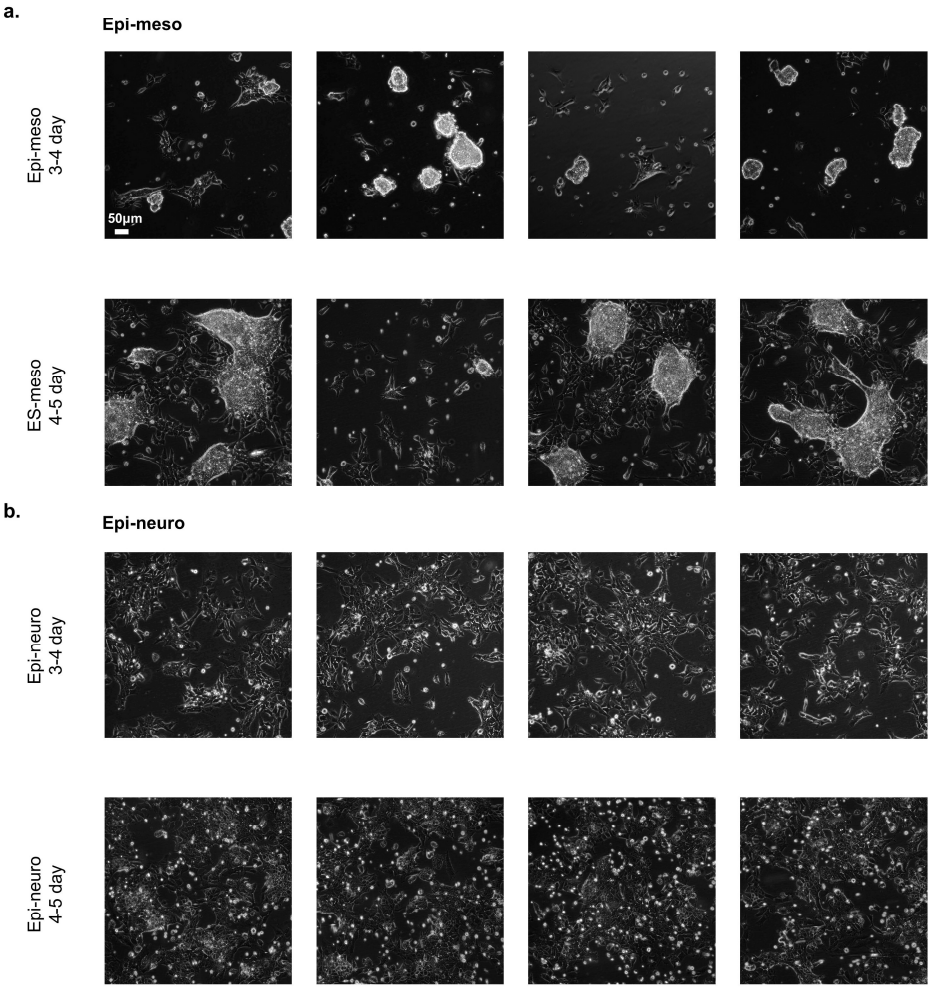

Fig S6. Brightfield images of Epi-NMP differentiation. a. Epi-meso cells were imaged between their 3rd and 4th day (1st row) and 4th and 5th day (2nd row), the cells were treated according to the Epi-meso protocol (see Materials and Methods). **b.** Epi-neuro cells were imaged between their 3rd and 4th day (1st row) and 4th and 5th day (2nd row), the cells were treated according to the Epi-neuro protocol (see Methods). All the images were taken on a Nikon Ti-E wide field microscope, using a 20X objective and a cooled camera (Orca flash 4.0, Hamamatsu).

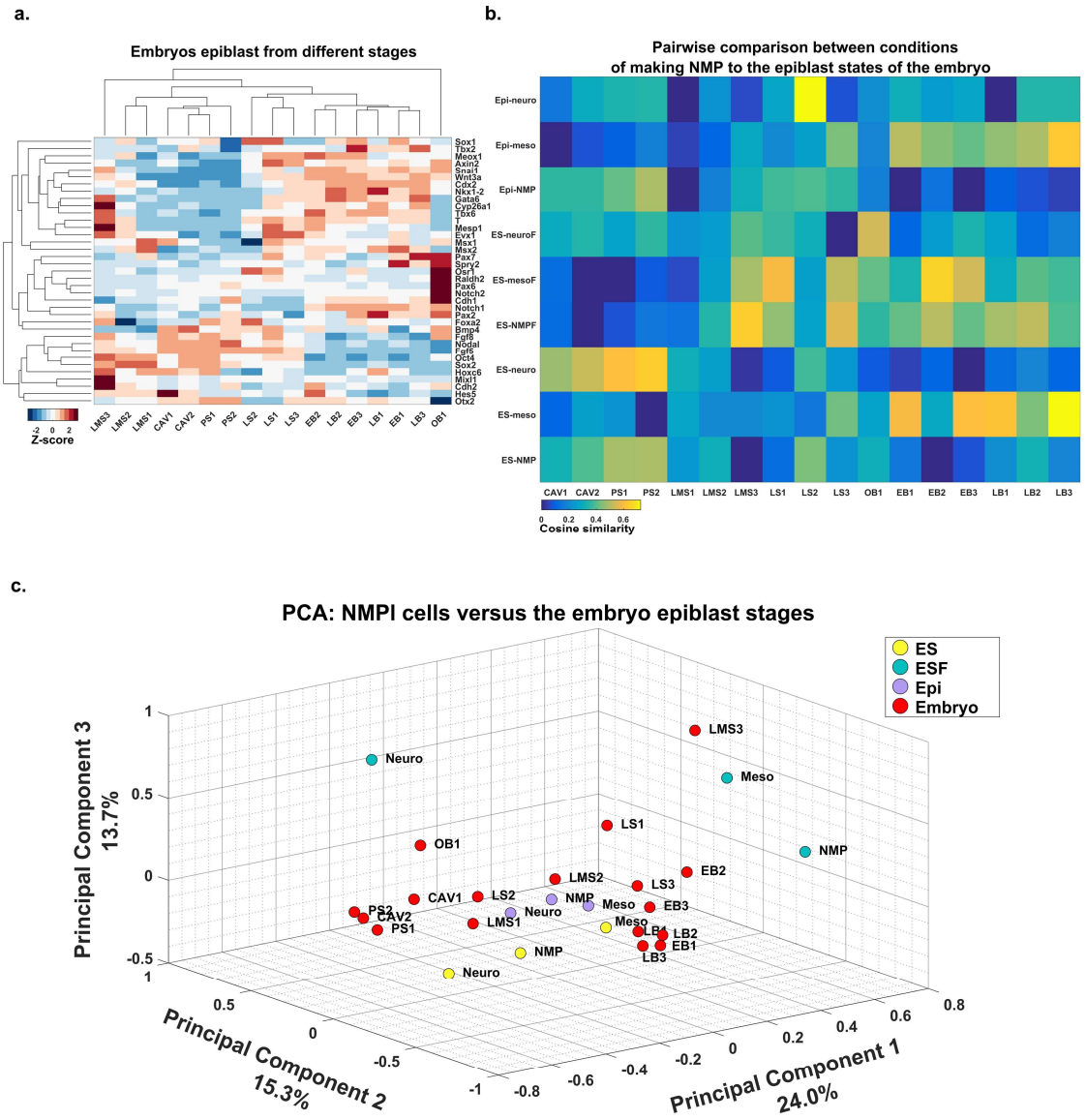

Fig S7. Comparison of expression of set of genes between the 3 in vitro protocols and the different embryo epiblast stages. **a.** Gene profile of the embryo epiblast in the different stages. Heatmap of the Z-score expression of the 27 genes (see Methods), was calculated from the microarray gene expression data of the epiblast/ectoderm (excluding the primitive streak) from different stages of the mouse embryo as published in (Kojima et al., 2014). **b.** The blue-yellow heatmap reflects the value of the pairwise cosine similarity measure that was calculated based on the expression of the 27 genes as in Fig S7a between the NMP in vitro protocols or their differentiation and the different stages of the epiblast mouse embryo. Dark blue (value of 0) indicates dissimilarity and bright yellow (value of 1) indicates maximal similarity (see Methods). **c**. PCA was performed on the expression of the 27 genes, indicated in Fig S7a, expressed in the NMP in vitro protocols, their differentiation and the different stages of the epiblast mouse embryo.

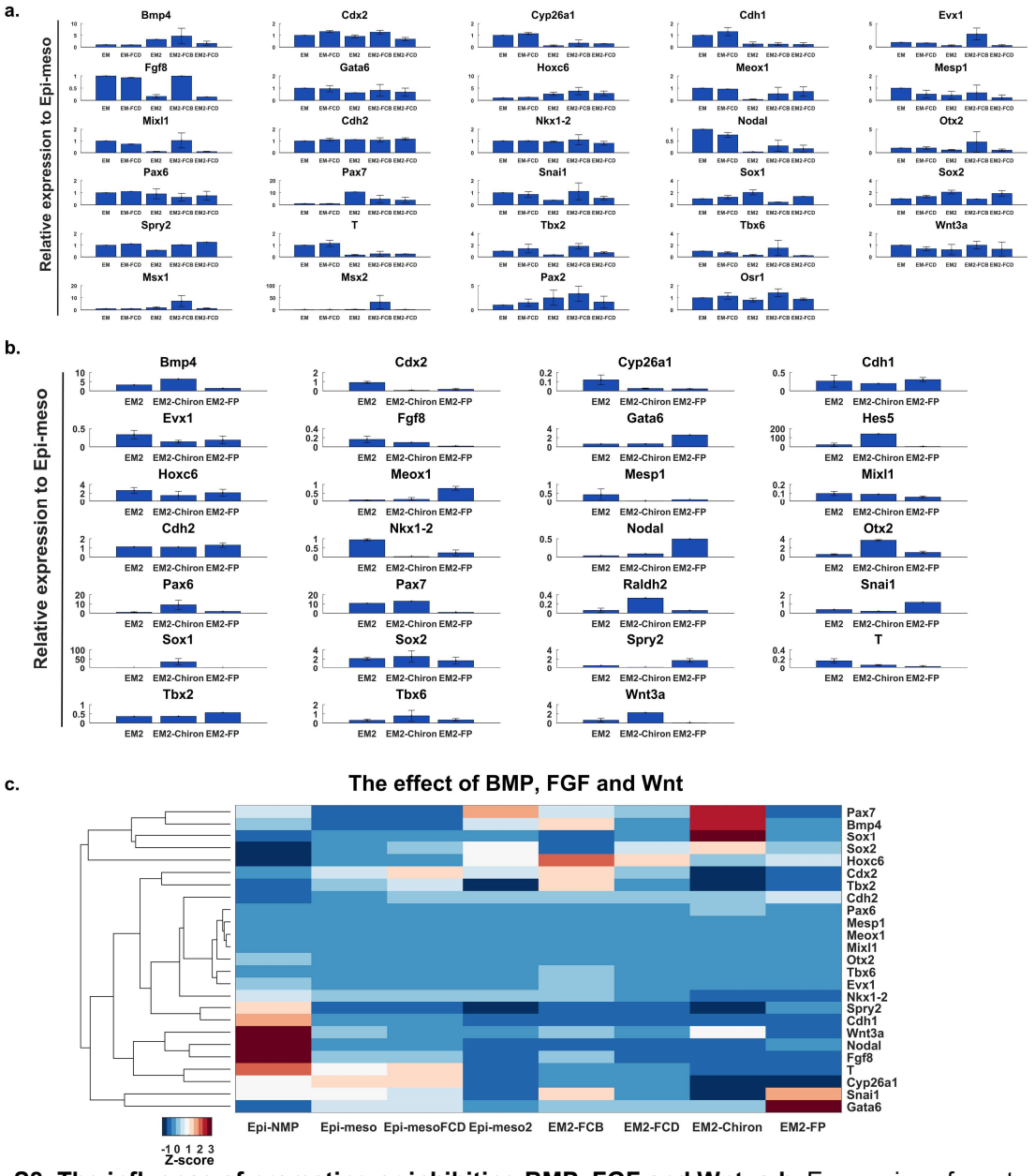

Fig S8. The influence of promoting or inhibiting BMP, FGF and Wnt. a-b. Expression of a set of chosen genes to monitor the differentiation of cells grown as indicated in Fig. 4b. The bars represent the genes average expression across biological replicas obtained by RT-qPCR and the error bars indicate the standard error between those replicas. The gene expression across the different conditions was normalized to Epi-meso condition (see Methods). **a.** the effect of BMP: measuring the expression of 29 genes in Epi-meso (EM) in comparison to Epi-meso condition supplement with BMP inhibitor, DMH-1 (EM-FCD) and Epi-meso2 (EM2) in comparison to Epi-meso2 condition supplement with either BMP (EM2-FCB) or DMH-1 (EM2-FCD). **b.** The effect of FGF and Wnt: measuring the expression of 27 genes in Epi-meso2 in comparison to Epi-meso2 condition with Chiron alone (no exogenous FGF, EM2-Chiron) or FGF with Wnt pathway inhibitor, IWP-2 (EM2-FP) (see Methods). **c.** Expression heatmap of Fig. S8a-b: all measurements were obtained by RT-qPCR and the normalized expression of each gene across the different conditions was scaled via calculating the Z-score (see Methods). The effect of BMP: Epi-NMP and Epi-meso are compared to Epi-mesoFCD. Epi-meso and Epi-meso2 are compared to Epi-

meso2 populations that their growing conditions either contained BMP (EM2-FCB) or DMH-1 (EM2-FCD). The effect of FGF and Wnt: Epi-meso2 is compared to EM2-Chiron and EM2-FP.

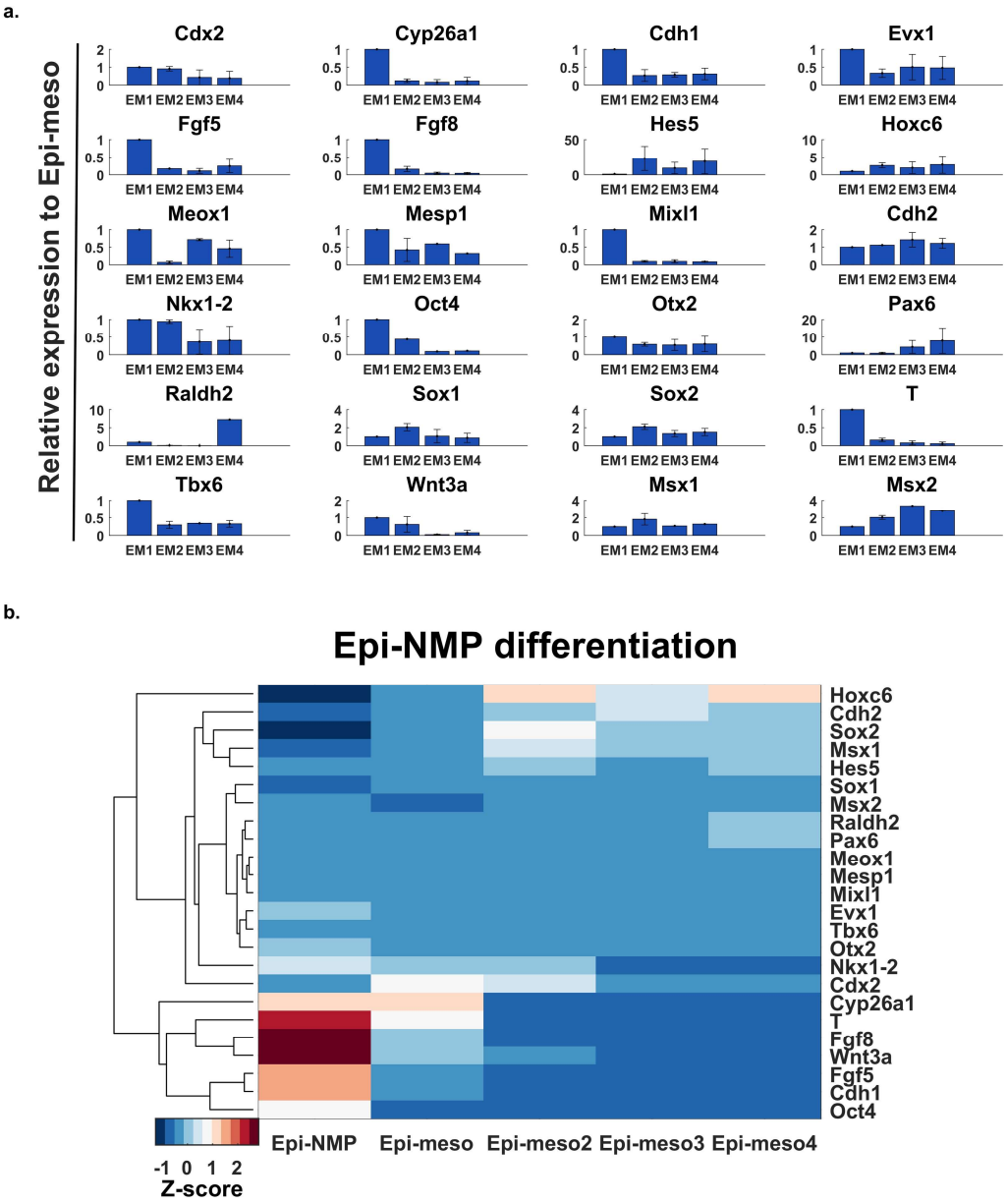

Fig S9. Cultering Epi-NMP over time. **a.** Measuring the expression of 24 genes in Epi-meso (EM1) and its derivatives: Epi-meso2, Epi-mes3 and Epi-mes4 (EM2, EM3 and EM4). The bars represent the genes average expression across biological replicas obtained by RT-qPCR and the error bars indicate the standard error between those replicas. The gene expression across the different conditions was normalized to Epi-meso condition (see Methods). **b.** Expression heatmap of Fig. S9a: all measurements were obtained by RT-qPCR and the normalized expression of each gene across the different conditions was scaled via calculating the Z-score (see Methods).

References

Aires, R., Jurberg, A. D., Leal, F., Novoa, A., Cohn, M. J. and Mallo, M. (2016). Oct4 Is a Key Regulator of Vertebrate Trunk Length Diversity. *Dev Cell* **38**, 262-274.

Amin, S., Neijts, R., Simmini, S., van Rooijen, C., Tan, S. C., Kester, L., van Oudenaarden, A., Creyghton, M. P. and Deschamps, J. (2016). Cdx and T Brachyury Co-activate Growth Signaling in the Embryonic Axial Progenitor Niche. *Cell Rep* **17**, 3165-3177.

Ang, S. L. and Rossant, J. (1994). HNF-3 beta is essential for node and notochord formation in mouse development. *Cell* **78**, 561-574.

Baillie Johnson, P., Voiculescu, O., Hayward, P. and Steventon, B. (2018). The chick caudo-lateral epiblast acts as a permissive niche for generating neuromesodermal progenitor behaviours. *bioRxiv* https://doi.org/10.1101/243980

Briggs, J. A., Li, V. C., Lee, S., Woolf, C. J., Klein, A. and Kirschner, M. W. (2017). Mouse embryonic stem cells can differentiate via multiple paths to the same state. *Elife* **6**.

Cambray, N. and Wilson, V. (2002). Axial progenitors with extensive potency are localised to the mouse chordoneural hinge. *Development* **129**, 4855-4866.

---- (2007). Two distinct sources for a population of maturing axial progenitors. *Development* **134**, 2829-2840.

Conlon, F. L., Barth, K. S. and Robertson, E. J. (1991). A novel retrovirally induced embryonic lethal mutation in the mouse: assessment of the developmental fate of embryonic stem cells homozygous for the 413.d proviral integration. *Development* **111**, 969-981.

Corsinotti, A., Wonk, F. C. K., Tatar, T., Szczerbinska, S., Halbritter, F., Colby, D., Gogolok, S., Pantier, R., Liggat, K., Mirfazeli, E. S., et al. (2017). Distinct SoxB1 networks are required for naïve and primed pluripotency. *vbioRxiv* **doi:** https://doi.org/10.1101/229716

Cunningham, T. J., Kumar, S., Yamaguchi, T. P. and Duester, G. (2015). Wnt8a and Wnt3a cooperate in the axial stem cell niche to promote mammalian body axis extension. *Dev Dyn* **244**, 797-807.

Davis, R. L. and Kirschner, M. W. (2000). The fate of cells in the tailbud of Xenopus laevis. *Development* **127**, 255-267.

de Chaumont, F., Dallongeville, S., Chenouard, N., Herve, N., Pop, S., Provoost, T., Meas-Yedid, V., Pankajakshan, P., Lecomte, T., Le Montagner, Y., et al. (2012). Icy: an open bioimage informatics platform for extended reproducible research. *Nat Methods* **9**, 690-696.

DeVeale, B., Brokhman, I., Mohseni, P., Babak, T., Yoon, C., Lin, A., Onishi, K., Tomilin, A., Pevny, L., Zandstra, P. W., et al. (2013). Oct4 is required ~E7.5 for proliferation in the primitive streak. *PLoS Genet* **9**, e1003957.

Downs, K. M. (2008). Systematic localization of Oct-3/4 to the gastrulating mouse conceptus suggests manifold roles in mammalian development. *Dev Dyn* **237**, 464-475.

Dunty, W. C., Jr., Biris, K. K., Chalamalasetty, R. B., Taketo, M. M., Lewandoski, M. and Yamaguchi, T. P. (2008). Wnt3a/beta-catenin signaling controls posterior body development by coordinating mesoderm formation and segmentation. *Development* **135**, 85-94.

Fontaine-Perus, J., Halgand, P., Cheraud, Y., Rouaud, T., Velasco, M. E., Cifuentes Diaz, C. and Rieger, F. (1997). Mouse-chick chimera: a developmental model of murine neurogenic cells. *Development* **124**, 3025-3036.

Fontaine-Perus, J., Jarno, V., Fournier le Ray, C., Li, Z. and Paulin, D. (1995). Mouse chick chimera: a new model to study the in ovo developmental potentialities of mammalian somites. *Development* **121**, 1705-1718.

Garriock, R. J., Chalamalasetty, R. B., Kennedy, M. W., Canizales, L. C., Lewandoski, M. and Yamaguchi, T. P. (2015). Lineage tracing of neuromesodermal progenitors reveals novel Wnt-dependent roles in trunk progenitor cell maintenance and differentiation. *Development* **142**, 1628-1638.

Gont, L. K., Steinbeisser, H., Blumberg, B. and de Robertis, E. M. (1993). Tail formation as a continuation of gastrulation: the multiple cell populations of the Xenopus tailbud derive from the late blastopore lip. *Development* **119**, 991-1004.

Gouti, M., Delile, J., Stamataki, D., Wymeersch, F. J., Huang, Y., Kleinjung, J., Wilson, V. and Briscoe, J. (2017). A Gene Regulatory Network Balances Neural and Mesoderm Specification during Vertebrate Trunk Development. *Dev Cell* **41**, 243-261 e247.

Gouti, M., Metzis, V. and Briscoe, J. (2015). The route to spinal cord cell types: a tale of signals and switches. *Trends Genet* **31**, 282-289.

Gouti, M., Tsakiridis, A., Wymeersch, F. J., Huang, Y., Kleinjung, J., Wilson, V. and Briscoe, J. (2014). In vitro generation of neuromesodermal progenitors reveals distinct roles for wnt signalling in the specification of spinal cord and paraxial mesoderm identity. *PLoS biology* **12**, e1001937.

Henrique, D., Abranches, E., Verrier, L. and Storey, K. G. (2015). Neuromesodermal progenitors and the making of the spinal cord. *Development* **142**, 2864-2875.

Keller, G. (2005). Embryonic stem cell differentiation: emergence of a new era in biology and medicine. *Genes Dev* **19**, 1129-1155.

Kojima, Y., Kaufman-Francis, K., Studdert, J. B., Steiner, K. A., Power, M. D., Loebel, D. A., Jones, V., Hor, A., de Alencastro, G., Logan, G. J., et al. (2014). The transcriptional and functional properties of mouse epiblast stem cells resemble the anterior primitive streak. *Cell Stem Cell* **14**, 107-120.

Lippmann, E. S., Williams, C. E., Ruhl, D. A., Estevez-Silva, M. C., Chapman, E. R., Coon, J. J. and Ashton, R. S. (2015). Deterministic HOX patterning in human pluripotent stem cell-derived neuroectoderm. *Stem cell reports* **4**, 632-644.

Mazzoni, E. O., Mahony, S., Peljto, M., Patel, T., Thornton, S. R., McCuine, S., Reeder, C., Boyer, L. A., Young, R. A., Gifford, D. K., et al. (2013). Saltatory remodeling of Hox chromatin in response to rostrocaudal patterning signals. *Nat Neurosci* **16**, 1191-1198.

McCauley, H. A. and Wells, J. M. (2017). Pluripotent stem cell-derived organoids: using principles of developmental biology to grow human tissues in a dish. *Development* **144**, 958-962.

McGrew, M. J., Sherman, A., Lillico, S. G., Ellard, F. M., Radcliffe, P. A., Gilhooley, H. J., Mitrophanous, K. A., Cambray, N., Wilson, V. and Sang, H. (2008). Localised axial progenitor cell populations in the avian tail bud are not committed to a posterior Hox identity. *Development* **135**, 2289-2299.

Nair, G., Abranches, E., Guedes, A. M., Henrique, D. and Raj, A. (2015). Heterogeneous lineage marker expression in naive embryonic stem cells is mostly due to spontaneous differentiation. *Sci Rep* **5**, 13339.

Neijts, R., Amin, S., van Rooijen, C., Tan, S., Creyghton, M. P., de Laat, W. and Deschamps, J. (2016). Polarized regulatory landscape and Wnt responsiveness underlie Hox activation in embryos. *Genes Dev* **30**, 1937-1942.

New, D. A. (1955). A New Technique for the Cultivation of the Chick Embryo in vitro. *Journal of embryology and experimental morphology* **3**, 320.

Niehrs, C. (2004). Regionally specific induction by the Spemann-Mangold organizer. *Nat Rev Genet* **5**, 425-434.

Nordstrom, U., Jessell, T. M. and Edlund, T. (2002). Progressive induction of caudal neural character by graded Wnt signaling. *Nat Neurosci* **5**, 525-532.

Nordstrom, U., Maier, E., Jessell, T. M. and Edlund, T. (2006). An early role for WNT signaling in specifying neural patterns of Cdx and Hox gene expression and motor neuron subtype identity. *PLoS Biol* **4**, e252.

Osorno, R., Tsakiridis, A., Wong, F., Cambray, N., Economou, C., Wilkie, R., Blin, G., Scotting, P. J., Chambers, I. and Wilson, V. (2012). The developmental dismantling of pluripotency is reversed by ectopic Oct4 expression. *Development* **139**, 2288-2298.

Preibisch, S., Saalfeld, S. and Tomancak, P. (2009). Globally optimal stitching of tiled 3D microscopic image acquisitions. *Bioinformatics* **25**, 1463-1465.

Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A. and Tyagi, S. (2008). Imaging individual mRNA molecules using multiple singly labeled probes. *Nat Methods* **5**, 877-879.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**, 676-682.

Sharma, R., Shafer, M. E. R., Bareke, E., Tremblay, M., Majewski, J. and Bouchard, M. (2017). Bmp signaling maintains a mesoderm progenitor cell state in the mouse tailbud. *Development* **144**, 2982-2993.

Stern, C. D. (2005). Neural induction: old problem, new findings, yet more questions. *Development* **132**, 2007-2021.

Steventon, B., Duarte, F., Lagadec, R., Mazan, S., Nicolas, J. F. and Hirsinger, E. (2016). Species-specific contribution of volumetric growth and tissue convergence to posterior body elongation in vertebrates. *Development* **143**, 1732-1741.

Steventon, B. and Martinez Arias, A. (2017). Evo-engineering and the cellular and molecular origins of the vertebrate spinal cord. *Dev Biol*.

Sweetman, D., Wagstaff, L., Cooper, O., Weijer, C. and Munsterberg, A. (2008). The migration of paraxial and lateral plate mesoderm cells emerging from the late primitive streak is controlled by different Wnt signals. *BMC Dev Biol* **8**, 63.

Takemoto, T., Uchikawa, M., Kamachi, Y. and Kondoh, H. (2006). Convergence of Wnt and FGF signals in the genesis of posterior neural plate through activation of the Sox2 enhancer N-1. *Development* **133**, 297-306.

Tsakiridis, A. and Wilson, V. (2015). Assessing the bipotency of in vitro-derived neuromesodermal progenitors *F1000Research* **4:100**, doi: 10.12688/f11000research.16345.12681.

Turner, D. A., Baillie-Johnson, P. and Martinez Arias, A. (2016). Organoids and the genetically encoded self-assembly of embryonic stem cells. *Bioessays* **38**, 181-191.

Turner, D. A., Hayward, P. C., Baillie-Johnson, P., Rue, P., Broome, R., Faunes, F. and Martinez Arias, A. (2014). Wnt/beta-catenin and FGF signalling direct the specification and maintenance of a neuromesodermal axial progenitor in ensembles of mouse embryonic stem cells. *Development* **141**, 4243-4253.

Tzouanacou, E., Wegener, A., Wymeersch, F. J., Wilson, V. and Nicolas, J. F. (2009). Redefining the progression of lineage segregations during mammalian embryogenesis by clonal analysis. *Dev Cell* **17**, 365-376.

Vincent, S. D., Dunn, N. R., Hayashi, S., Norris, D. P. and Robertson, E. J. (2003). Cell fate decisions within the mouse organizer are governed by graded Nodal signals. *Genes Dev* **17**, 1646-1662.

Weinstein, D. C., Ruiz i Altaba, A., Chen, W. S., Hoodless, P., Prezioso, V. R., Jessell, T. M. and Darnell, J. E., Jr. (1994). The winged-helix transcription factor HNF-3 beta is required for notochord development in the mouse embryo. *Cell* **78**, 575-588.

Wilson, V., Olivera-Martinez, I. and Storey, K. G. (2009). Stem cells, signals and vertebrate body axis extension. *Development* **136**, 1591-1604.

Wymeersch, F. J., Huang, Y., Blin, G., Cambray, N., Wilkie, R., Wong, F. C. and Wilson, V. (2016). Position-dependent plasticity of distinct progenitor types in the primitive streak. *Elife* **5**.

Yamaguchi, T. P., Takada, S., Yoshikawa, Y., Wu, N. and McMahon, A. P. (1999). T (Brachyury) is a direct target of Wnt3a during paraxial mesoderm specification. *Genes Dev* **13**, 3185-3190.

Zechner, D., Fujita, Y., Hulsken, J., Muller, T., Walther, I., Taketo, M. M., Crenshaw, E. B., 3rd, Birchmeier, W. and Birchmeier, C. (2003). beta-Catenin signals regulate cell growth and the balance between progenitor cell expansion and differentiation in the nervous system. *Dev Biol* **258**, 406-418.