First posted online on 25 April 2019 as 10.1242/dev.168187
Access the most recent version at http://dev.biologists.org/lookup/doi/10.1242/dev.168187

An Epiblast Stem Cell derived multipotent progenitor population for axial extension

Shlomit Edri[1*], Penny Hayward[1], Peter Baillie-Johnson[2], Benjamin Steventon[1] and Alfonso Martinez Arias[1*]

1. Department of Genetics
   Downing Site
   University of Cambridge
   Cambridge, CB2 3EH
   UK
2. Current Address: Wellcome-MRC Cambridge Stem Cell Institute
   Tennis Court Road
   Cambridge, CB2 1QR
   UK

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

Abstract

The mammalian embryos Caudal Lateral Epiblast harbours bipotent progenitors that contribute to the spinal cord and the paraxial mesoderm in concert with the body axis elongation. These progenitors, called Neural Mesodermal Progenitors (NMPs) are identified as cells coexpressing *Sox2* and *T/Brachyury*, a criterion used to derive NMP-like cells from embryonic stem cells *in vitro*. However, these progenitors do not self renew, as embryonic NMPs do. Here we find that protocols that yield NMP-like cells *in vitro* first produce a multipotent population that, in addition to NMPs, generate progenitors for the lateral plate and intermediate mesoderm. We show that Epiblast Stem Cells (EpiSCs) are an effective source for these multipotent progenitors that are further differentiated by a balance between BMP and Nodal signalling. Importantly, we show that NMP-like cells derived from EpiSCs can be maintained *in vitro* and exhibit a gene expression signature like their embryonic counterparts.

Development • Accepted manuscript

© 2019. Published by The Company of Biologists Ltd.

Introduction

The anteroposterior axis of a vertebrate can be subdivided into three anatomically distinct regions: the head, the trunk and the tail. The trunk starts at the end of the hindbrain, runs to the anus and comprises derivatives of the mesoderm and the ectoderm such as the thoracic cage, muscles, kidneys and spinal cord. 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 ectoderm, while in amniotes *e.g.* chickens and mice, it is derived from the expansion of the Caudal Epiblast (CE), a proliferative region located at the caudal end of the embryo, where the primitive streak persists, that acts as a source for paraxial, intermediate and lateral plate mesoderm as well as for the spinal cord (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 CE harbours a population of bipotential progenitors located behind the node, at the Node Streak Border (NSB), and extending laterally into the Caudal Lateral Epiblast (CLE), that give rise to neural and mesodermal precursors. These cells have been called Neural Mesodermal Progenitors (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 capable of limited self renewal (Cambray and Wilson, 2002; McGrew et al., 2008; Tzouanacou et al., 2009).

Recently a few studies have claimed the generation of NMP-like cells in adherent cultures of mouse and human embryonic Pluripotent Stem Cells (PSCs) (Denham et al., 2015; 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 (PXM) or spinal cord progenitors and their derivatives. However, there is no evidence that these NMP-like cells are propagated *in vitro* as they are in the embryo (Tsakiridis and Wilson, 2015). Furthermore, coexpression of *T* and *Sox2* might not be a unique characteristic of NMPs as it is also a signature of EpiSCs (Kojima et al., 2014) which are pluripotent and this does not imply that EpiSCs are NMPs. While other markers have been used to refine the molecular identity of 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 Supplementary section 1-2 and Fig. S1), emphasizing the notion that these gene expression signatures are not uniquely associated with NMPs. Altogether, these observations raise questions about the identity of the *T* - *Sox2* coexpressing cells derived from ESCs and about the signature of the NMPs.

Here, we show that *T* - *Sox2* coexpressing cells derived from ESCs and EpiSCs based differentiation protocols display differences at the level of gene expression and represent collections of different developmental stages of the transition between naïve, primed pluripotency and neuro-mesodermal fate choices. Furthermore, we find that in adherent culture, all available protocols generate a multipotent population where in addition to an NMP signature we find also signatures for Lateral Plate and Intermediate Mesoderm (LPM and IM) as well as the allantois. We report on a new protocol, based on EpiSCs, that sequentially generates, at a high frequency, the multipotent population and an NMP-like population with many of the attributes of the embryonic NMPs. Particularly, these cells can be maintained *in vitro* for a limited period of time and contribute to posterior neural and mesodermal regions of the embryonic body in a xenotransplantation assay. Our study leads us to propose that, *in vitro* and *in vivo*, NMPs are derived from a multipotent population that emerges in the epiblast at the end of gastrulation and gives rise not only to the elements of the spinal cord and PXM but all elements of the trunk mesoderm.

Development • Accepted manuscript

Results

EpiSCs yield a postimplantation epiblast population that resembles the CLE

Several protocols allow the differentiation of ESCs into an NMP-like population, defined as cells that coexpress *T* and *Sox2* that can be further differentiated into neural and mesodermal progenitors (summarized in (Henrique et al., 2015)). However, it is not clear whether these NMP-like cells derived through different protocols are similar to each other and, importantly, how each relates to the NMPs in the embryo. To begin to answer these questions we compared NMP-like cells obtained from three different protocols: ES-NMPs (Turner et al., 2014) and ES-NMPFs (Gouti et al., 2014), derived from ESCs, as well as Epi-NMPs, derived from a new protocol that we have developed from EpiSCs (Fig. 1A-B and Methods). All protocols yield cells coexpressing *T* and *Sox2* at the level of both mRNA and protein (Fig. 1C-D and Supplementary Fig. S2A), but differ in the numbers of cells with this signature as well as in the levels and degree of correlated expression of the two genes (Supplementary Fig. S2). At the protein level, all the conditions exhibit high percentage of cells coexpressing *Sox2* and *T* (Fig. 1D) and a significant positive correlation between the two genes is observed only in the ES-NMP condition whereas a negative one in the EpiSCs population. Across all the conditions *Sox2* shows the same degree of variability both in the protein and mRNA levels, *T* however exhibits greater variability at the protein level in the EpiSCs and EpiSCs derived populations in comparison to the ESCs derived NMP-like populations, which is not the case at the mRNA level (Fig. 1D and Supplementary Fig. S2B). At the later, the Epi-NMP population has the highest number of *T* - *Sox2* positive cells with low variance in comparison to the other conditions (Supplementary Fig. S2B).

To characterize the different NMP-like populations further, we investigated the expression of a total 27 genes associated with the epiblast, the CE and the NSB/CLE region, where NMPs are thought to reside, as well as of genes associated with neural and mesodermal differentiation between stages E7.0 and E8.5 (see Supplementary sections 1-2 and Fig. S1 for the criteria we followed to select these genes). Both of the ESCs derived NMP-like populations exhibit expression of *Cdh1*, *Oct4*, *Fgf5* and *Otx2*, an epiblast signature (Fig. 1C-D, Fig. 2A and Fig. 3). Surprisingly ES-NMPF cells also display high levels of genes associated with mesendoderm differentiation *e.g. Mixl1* (endoderm), *Tbx6* (paraxial mesoderm) and *Evx1* (extraembryonic mesoderm) (Fig. 2A, Fig. S3 and (Gouti et al., 2014)); this suggests that ES-NMP and ES-NMPF are overlapping populations in various stages of differentiation including cells early epiblast/gastrula like stages. In contrast, Epi-NMPs are in a different state: in addition to the accepted NMP signature (*T*, *Sox2* and *Nkx1-2*) these cells express significant levels of *Nodal*, *Fgf8*, *Fgf5*, *Foxa2, Otx2* and *Oct4* together with *Cyp26a1* (Fig. 1C-D, Fig. 2A, Fig. S3 and Fig. 3). This is a profile associated with the late epiblast (about E7.5), around the time of the appearance of the CE, before NMPs can be detected (Supplementary sections 1-2 and Fig. S1).

To correlate the gene expression profiles with the state of the cells in the different culture conditions, we defined two measures of the degree of differentiation based on the average expression Z-score values for mesodermal and neural genes (Methods). These measures are indexes that define a global value of the state of the cells in a particular condition. The ‘Epiblast index’ reflects the degree of differentiation of the population based on the expression of the chosen 27 genes and is defined by the ratio between the non-differentiated (epiblast) and differentiated stages. On the other hand, the ‘NMP index’, is defined as the degree of neural or mesodermal differentiation *i.e.*, whether the population exhibits any bias towards either fate (Fig. 2B-C, Methods). In both cases, the distance of the cells from the diagonal and from the origin of the plot reflects the average phenotype of the cell population. The closer to both, the more the population is in a progenitor, uncommitted epiblast state as in the case of the Epiblast index, this means that the cells are in low differentiation state. In the case of the NMP index, it means that they do not exhibit a differentiation bias (Methods). The indexes show that Epi-NMPs have a high undifferentiated epiblast identity with some

Development • Accepted manuscript

degree of differentiation towards mesoderm whereas ES-NMPs exhibit low epiblast identity and a degree of differentiation towards the neural fate. In contrast, ES-NMPFs exhibit high epiblast identity but also a strong mesodermal differentiation bias. The differences between the three populations are further emphasized by an examination of the protein levels of some of these markers (Fig. 1C-D and Fig. 3): NMP-like populations derived from ESCs exhibit high levels of *Sox2*, *Oct4* and *Cdh1* expression with some cells expressing *Otx2*, but little or no expression of *Cdh2*, a signature characteristic of early undifferentiated epiblast (Morgani et al., 2018). ES-NMPF exhibits a combination of *Cdh1* and *Cdh2* at the level of single cells (Fig. 3), a situation rarely seen *in vivo*. In contrast, the Epi-NMP exhibit lower level of *Sox2* and *Oct4* (Fig. 1C-D) and a mutual exclusive expression of *Cdh1* and *Cdh2* (Fig. 3), which is a characteristic of the late Epiblast (Corsinotti et al., 2017; Morgani et al., 2018).

Exposure of the different NMP-like populations to neural and mesodermal differentiation environments reveals their potential (Fig. 2A, Fig. S3 and Methods). In all cases the cells differentiated into neural and mesodermal progenitors but with different biases depending on their origin (Fig. 2). ES-NMPFs and its differentiated progeny exhibit a mesodermal bias, while ES-NMPs exhibit a slight bias towards the neural fate, both in agreement with their indexes. In contrast, Epi-NMPs which has a high epiblast index, differentiate equally into neural and mesodermal cell types (Fig. 2C).

Altogether these results suggest that different protocols yield related, but different, NMP-like populations which might have different functional properties. Furthermore, these populations are differently biased in their differentiation potential. The NMP-like population derived from EpiSCs, appears to be the closest to an uncommitted epiblast state and to harbour the most unbiased state.

Developmental staging of in vitro derived NMP populations

The differences between the candidate NMP-like populations derived *in vitro* suggest that they might represent different stages of the transition between the early 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 early-CLE (E7.5 (Kojima et al., 2014), and mapped the NMP-like populations, as well as their differentiated derivatives, onto it (Supplementary section 3 and Fig. S4). Using this as a reference, we observe that in a three-dimensional Principal Component Analysis space, ES-NMP, Epi-NMP and their derivatives mapped closely to the different embryonic stages, whereas the ES-NMPF and its differentiated populations lie separate from these trajectories and from the embryonic stages (Fig. S4C). Furthermore, the Epi-NMP and its derivatives projected closely to each other within the embryo trajectory between the LMS and EB stages.

We also used our developmental reference to explore the proximity of the *in vitro* derived populations to specific epiblast states *in vivo*. To do this, we used the microarray epiblast analysis of Kojima et al. 2014 as a reference and calculated the cosine similarity between the *in vitro* population and the different stages of the embryo as a metric for the proximity of each *in vitro* population to a particular embryonic stage (Fig. 4A, Fig. S4B and Methods). Using this measure, and keeping with the indexes discussed above, we find that ES-NMPs correlate with an early epiblast state, whereas ES-NMPFs appear to be a broad population associated with several late differentiating epiblast stages, mostly late epiblast, confirming that the later represents heterogeneous populations of differentiating cells. On the other hand, the Epi-NMP exhibits two peaks, associated with the early epiblast and the primitive streak. This analysis also reveals that Epi-meso, a population derived from the Epi-NMPs, resembles the LB stages, which correspond to E7.5/8.25 the stage where the CLE can be seen to harbour the NMP for the first time (Wymeersch et al., 2016).

Development • Accepted manuscript

Altogether these results support the notion that different starting conditions and differentiation protocols lead to populations with different identities. ES-NMP seems to resembles early epiblast state and ES-NMPF represents heterogeneous populations with representations of differentiated cells. On the other hand, Epi-NMPs appear to represent a tighter population resembling an epiblast stage.

Multiple tail bud fates emerge from differentiating ESCs and EpiSCs in culture

In the course of our survey of markers distinguishing the embryonic regions of the CE in the different populations, we noticed that all protocols lead to coexpression of *T* and *Sox2* together with the expression of genes that are not associated with NMPs, *e.g. Mesp1*, *Evx1*, *Mixl1*, *Gata6*, *Bmp4*, *Msx1*, *Msx2*, *Osr1*, *Pax2* and *Tbx2* (Fig. 2, Fig. S3 and (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 *in vitro* cells, these genes are expressed in the posterior domain of the CE, in the progenitors of the allantois (*Tbx2*, *Tbx4*, *Mixl1* and *Evx1*), the LPM (*Msx1* and *Msx2*) and the IM (*Pax2*, *Osr1*) (Supplementary sections 1-2 and Fig. S1). The *in vitro* derived populations can be mapped to this stage interval, thus suggesting that they are not restricted to harbour NMPs only, but rather that they represent a multi-potential population which includes progenitors of LPM, IM and allantois.

In the embryo, the further 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 ones (NMPs) (Wymeersch et al., 2016). To test this, we altered the levels of BMP in the Epi-meso population (Fig. 4B-C and Fig. S5A, C), which appears to be the closest to the embryo CLE, and applied the NMP index to ascertain the differentiation bias of the resulting population. In our cultures, inhibition of BMP signalling elevates the expression of NMP markers *e.g. T*, *Sox2* and *Cdx2* (Epi-meso versus Epi-mesoFCD in Fig. S5C) and thus increases its NMP identity (Fig. 4C). On the other hand, addition of BMP to the derivatives of Epi-meso population (EM2-FCB) elevates dramatically its mesodermal state (Fig. 4C) and specifically increases the expression of genes associated with posterior fates: *Bmp4*, *Msx1*, *Msx2* and *Tbx2* together with *Cdx2* and *Snail*, (Fig. S5A, C). Similarly to Epi-mesoFCD, inhibition of BMP in the Epi-meso2 population sample (EM2-FCD Fig. 4B-C), slightly improves its NMP index in comparison to Epi-meso2.

When Epi-meso cells are grown in N2B27 supplemented with Chiron alone (Fig. 4B-C and Fig. S5B-C), we observe an increase in the levels of expression of neural markers (*Sox1*, *Sox2* and *Hes5*) with a concomitant shift of its NMP index to neural and a loss of mesodermal identity. Furthermore, inhibition of Wnt in the Epi-meso2 state (EM2-FP, Fig. 4B-C and Fig. S5B-C) leads to a reduction in the expression of neural progenitor markers and an elevation in the expression of mesodermal ones (*Gata6* and *Snail1*), which appropriately shift its NMP index to the mesodermal side with low neural averaged value in comparison to Epi-meso2. This is surprising as it is often thought that Wnt signalling suppresses neural development during the early stages of gastrulation (E6.0-E7.0). However, while this is the case for anterior neural fates early in development, the expansion of neural progenitors requires Wnt signalling (Garriock et al., 2015; Zechner et al., 2003). Therefore, these observations support the suggestion that the Epi-meso populations are related to the NMPs rather than to an early epiblast population.

In the embryo, as the posterior region of the CE is dominated by BMP signalling, the differentiation of the anterior domain is dependent on its proximity to the NSB. We observe that the Epi-NMP population expresses *Nodal* and *Foxa2* genes (Fig. 2A and Fig. S3A) which are associated with the NSB. Inhibition of Nodal signalling reduces *T -Sox2* coexpressing cells (Turner et al., 2014) and this led us to test whether Nodal signalling influences the NMP signature and differentiation potential of the NMP-like cells. To do this, we cultured Epi-NMP from Nodal mutant EpiSCs (Nodal -/- Epi-NMP, Methods and Fig. 5A) and compared them to Nodal mutant Epi-NMPs supplemented with 2 different doses of

Development • Accepted manuscript

Nodal in the presence of FGF and Chiron: 100ng/ml of Nodal (Nodal-/- Epi-NMP+0.1xNodal) or 1µg/ml of Nodal (Nodal-/- Epi-NMP+1xNodal). Addition of Nodal to Nodal -/- Epi-NMP cells lifts their levels of *T* and lowers their levels of *Sox2* together with an increase of *Cyp26a1* and *Fgf8* expression. These results suggest that Nodal signalling is necessary to maintain the relative levels of *Sox2* and *T* and significant levels of *Fgf8* and *Cyp26a1*, which are characteristic of the CE.

In summary, our results indicate that in all protocols tested, differentiation of PSCs towards a caudal population does not result in the specification of NMPs only, but rather of a multipotent population for all axial derivatives; different protocols appear to exhibit different representations of this population which is further differentiated by BMP and Nodal. 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.

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

In the embryo, the initial NMP population needs to be amplified, together with the progenitors of the LPM and IM, 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); if this were not the case the initial population would be exhausted before completing axial elongation. We suggest that this amplification should be an additional criterion to identify NMPs *in vitro*.

Earlier studies have shown that ESCs derived NMPs are not able to maintain the *T - Sox2* coexpressing cells when they are passaged in the conditions in which they were generated, namely FGF and Chiron or Chiron alone ((Gouti et al., 2014; Turner et al., 2014) and unpublished observations). Surprisingly, we noticed 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 (Epi-NMP to Epi-meso and Epi-meso to Epi-meso2) with a low differentiation index (Fig. 5B-C and Fig. S6). By passaging Epi-meso, the levels of *T* decrease, but unlike other situations, do not disappear (the levels of *T* are detectable by RT-qPCR, see supplementary Fig. S6). Furthermore, in the transition from Epi-NMP to Epi-meso, cells lose the expression of epiblast markers *e.g. Fgf5*, *Nodal*, *Otx2*, *Oct4* and *Cdh1* (Fig. 2A and Fig. S3A and protein expression of *Oct4*, *Otx2*, *Cdh1* and *Cdh2* in Fig. 1C-D and Fig. 3). This suggests that Epi-meso population is a state distinct from the epiblast and containing many features of the NMPs that are a subset of the CLE, however different from the epiblast characteristic of what we have called Epi-NMP.

During the passages of Epi-meso, we observe a progressive decrease but not an extinction in the expression of NMP markers **(***Cyp26a1*, *Fgf8* and *Nkx1-2*), similarly to the case of *T*. This decrease is accompanied with a slow increase in the expression of differentiation genes associated with neural fate: *Cdh2*, *Sox2*, and *Hes5* (Fig. S6). This expression pattern is likely to reflect a decrease in the NMP state of the population and mirrors a similar decrease in the embryo (Wymeersch et al., 2016). Another explanation for this magnitude of loss in the context of culturing cells *in vitro*, is the lack of mechanical support in the *in vitro* system and the progressive differentiation to neural fates as a default differentiation programme. It is important to emphasize that our current understanding of the NMPs does not determine the levels of *T* or *Sox2* that are required and that low levels are likely to still be sufficient to maintain this state.

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

During the elongation of the posterior body axis, NMPs progressively exit from their niche in the epiblast adjacent to the node and enter either the primitive streak, where they ingress and integrate with the presomitic mesoderm, or are retained in the epiblast and enter the posterior region of the emerging spinal cord. Depending on the time of exit from the niche,

Development • Accepted manuscript

they will contribute to different anteroposterior axial levels. With this in mind, we tested the ability of the *in vitro* derived cells to display these behaviours by transplanting differentiating ESCs and EpiSCs into the elongating region of chicken embryos. Previous experiments have demonstrated that these embryos are good hosts for these experiments and that transplanted mammalian cells integrate with the host and produce functional neural and mesodermal derivatives (Fontaine-Perus et al., 1997; Fontaine-Perus et al., 1995; Gouti et al., 2014). However, as pluripotent stem cells can also contribute to spinal cord and somitic mesoderm upon transplantation, we have focussed our assessment of NMP behaviour on the length of time that the transplanted cells remain within the NMP niche and continually generate these progenitor populations. This is in turn reflected by the length of labelled cells distributed along the anteroposterior axis (Baillie-Johnson et al., 2018).

We focused our experiments on the EpiSC-derived NMP-like populations as the different tests discussed above suggest that they are the closest to the embryonic NMPs. In our experiments, we transplanted cells from the EpiSC, Epi-NMP and Epi-meso conditions into a small region caudal and lateral to the node of the developing chick embryo. This region corresponds to a region that has been shown to contribute to spinal cord and paraxial mesoderm progenitors by fate mapping (Baillie-Johnson et al., 2018) and is outlined by dotted boxes in Fig. 6 (for details see Fig. S7). After around 15 hours of incubation, we measured the grafted cells' contributions to the somites and the developing neural tube, but most importantly the axial length to which they could contribute. As shown in Fig. 6 A-D, EpiSCs contributed only to short axial extensions and their descendants were mainly located in the mesodermal compartments. This result suggests that the EpiSCs exit early from the NMP domain and at that stage, their most likely fate is mesodermal. This is further reflected in labelled EpiSCs' contributions to the anterior axial levels shown in Fig. 6d. (*viz.* at the level of somites 1-6 see also Supplementary section 4 and Fig. S8). On the other hand, Epi-NMP and Epi-meso populations contributed to increasingly more posterior regions (contrast Fig. 6H and L to D; see also Fig. S8) with mixed neural and mesodermal contributions; in particular, we noticed that the Epi-meso grafts made more frequent dual neural and mesodermal contributions than the Epi-NMPs and that they showed a significant bias towards more posterior positions (Fig. 6L and Fig. S8 and Supplementary section 4). Since the Epi-meso population is derived from Epi-NMPs, these results suggest that their temporal sequence *in vitro* results in cells with an ability to colonise more posterior axial levels after transplantation. Perhaps this reflects the fact that Epi-meso cells express more posterior Hox genes than Epi-NMPs (Fig. 2A and Fig. S3) and this might contribute to their ability to colonize more posterior regions of the embryo (see (Denans et al., 2015)). This may additionally explain why the Epi-meso population less frequently produced long contributions to the embryonic axis compared to the Epi-NMP population, if it has a bias towards these more posterior axial levels. On comparing the lengths of the contributions from all three populations, those from Epi-NMP and Epi-meso were found to be significantly longer than those from the EpiSCs condition (Supplementary section 4).

Taking the above result together with our gene expression analyses, we conclude that the continued propagation of the Epi-NMP population in culture can produce of a population that closely resembles the newly arisen embryonic NMPs at E8.25.

Discussion

We have used and compared three PSCs based differentiation protocols to study the emergence *in vitro* of a population of bipotential progenitors, NMPs, that in the mammalian embryo, give rise to the paraxial mesoderm and spinal cord of the thoracic tract. Our results show that each of these protocols produces populations of cells with different gene expression signatures and ability to contribute to axial elongation but with two common denominators: coexpression of *T* and *Sox2* as well as of genes associated with LPM, IM and allantois. These results suggest that coexpression of *T* and *Sox2* is not a univocal criterion to

Development • Accepted manuscript

identify NMPs, that the populations generated *in vitro* are not restricted to NMPs and that, therefore, the identification of these progenitors *in vitro* requires additional criteria, in particular an ability to self renew and to make long contributions to axial extension, as well as an association with the node (Gouti et al., 2015; Henrique et al., 2015; Steventon and Martinez Arias, 2017; Wilson et al., 2009). Applying these criteria to differentiating PSC populations, we identify a specific protocol that starting from EpiSCs yields a population, Epi-meso, similar to the NMPs in the embryo in terms of cellular function, gene expression, limited maintenance over time, long axial contributions and the exit timing of the progenitors from the caudal domain of the embryo (see also (Edri et al., 2018). We surmise that this population emerges from a late epiblast like state, Epi-NMP, that can also give rise to LPM, IM and extraembryonic mesoderm in a signalling dependent manner. Our observations are in agreement with recent descriptions of the development of the tail bud (Wymeersch et al., 2019) and suggest that such a multipotent population might be an obligatory intermediate for the emergence of the NMPs *in vitro*. ESCs based protocols yield similar populations that can be differentiated into mesodermal and neural progenitors but lack several features characteristic of NMPs, in particular their ability to self renew and to contribute significantly to axial extension (Baillie-Johnson et al., 2018; Gouti et al., 2014; Turner et al., 2014). Furthermore, as we have shown here, these populations represent highly heterogeneous populations with a low representation of NMPs. A similar multipotent population is likely to exist in the embryo; analysis of lineage tracing data at the single cell level reveals the existence of clones that span the spinal cord and, at least, two mesodermal derivatives (see Figure 4 in (Tzouanacou et al., 2009).
Upon further exposure to Wnt and FGF signalling *in vitro*, this multipotent population, Epi-NMP, evolves and generates cells with many of the hallmarks of the NMPs, including a limited ability to maintain in culture coexpression of *Sox2* and *T* over few passages, ability to differentiate into neural and mesodermal progenitors in a Wnt dependent manner and to make long and more posterior contributions to axial extension in a xenotransplantation assay. For these reasons, we propose to call the Epi-NMP population Epi-CE, and the Epi-meso, Epi-NMP. In the embryo, the emergence of the NMPs from the multipotent population is likely to respond to a regionalization of signalling with BMP and Wnt signalling in the posterior domain, favouring LPM and IM (Sharma et al., 2017) and Nodal and Wnt signalling in the anterior region favouring NMPs. In agreement with this, 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 (see also (Edri et al., 2018).

There are many studies in which Wnt signalling can caudalize epiblast like populations (Amin et al., 2016; Mazzoni et al., 2013; Neijts et al., 2016; Nordstrom et al., 2002; Nordstrom et al., 2006) and in these cases, which are mostly ESCs based, the NMP-like cells fail to self renew as they do *in vivo*. In contrast to these ESCs based protocols, here we have shown that exposure of pre-treated EpiSCs to FGF and Chiron generates a population with a gene expression signature characteristic of a late CE, around the time of the appearance of the node *i.e.* the multipotent population. The importance of Nodal in the establishment of the multipotent population, and perhaps also in the definition of the NMP domain, is underscored by our studies with Nodal mutant cells in which the rescue of a population with a disrupted relative level between *Sox2* and *T*, is crucially dependent on the levels of Nodal signalling. Consistent with a role of the node and of Nodal in this population, embryos mutant for *Foxa2* which lack a node, exhibit deficiencies in the organization of the CE 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).

In vitro, 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

Development • Accepted manuscript

(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 *Sox2* can implement 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 associated with overexpression of *Oct4* (Aires et al., 2016). This 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 noticed that cells progressively lose *T* expression and increased *Sox2* expression. This is surprising since a widespread notion suggests that Wnt signalling suppresses neural differentiation and promotes mesoderm. However, while in the embryo this is true during the first phase of gastrulation, before the appearance of the node at E7.5, and reflects the maintenance of an anterior neural fate away from mesendoderm (Arkell and Tam, 2012). This might not be the case during the development of the caudal region of the embryo as there is a clear evidence that during this period Wnt/β-catenin signalling is required for the expression of *Sox2* (Takemoto et al., 2006; Takemoto et al., 2011) and for the expansion of the neural progenitors in the spinal cord (Zechner et al., 2003). Furthermore, during this period, increases in Wnt/β-catenin signalling do not supress neural development (Garriock et al., 2015). A requirement for Wnt signalling in the development of the spinal cord is further emphasized by the observation that the *Sox2* gene has a Tcf response element and responds to Wnt signalling (Takemoto et al., 2006). Thus, we suggest that the response of neurally specified cells to Wnt signalling is a measure of the stage and position of the cells generated by the *in vitro* protocols.

In summary, using a specific experimental protocol we have shed light on the origin of the NMP population *in vivo* and *in vitro*. Our work highlights the importance of the starting population in the differentiation of specific cell types as well as in considering the state of the *in vitro* produced cells to the embryo.

Acknowledgements
This work was supported by Cambridge Trust and Cambridge Philosophical Society scholarships and AJA Karten trust award to S. Edri, a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (GrantNumber 109408/Z/15/Z) to B. Steventon, 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, Meritxell Vinyoles, Vikas Trivedi and Valerie Wilson for discussions and anonymous reviewers for discussions and comments.

Development • Accepted manuscript

Materials and Methods

Cell culture

E14-Tg2A ESCs were grown in tissue-culture plastic flasks coated with 0.1% gelatine (Sigma-Aldrich, G1890-100G) in PBS (with Calcium and Magnesium, Sigma-Aldrich, D8662) filled with GMEM (Gibco, UK) supplemented with non-essential amino acids, sodium pyruvate, GlutaMAX^{TM}, β-mercaptoethanol, foetal bovine serum and LIF. Culture media were changed daily and cells passaged every other day. The differentiation protocols are as follows:

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 the N2B27 medium was supplemented with 3μM of CHIR99021 (Chiron 10mM, Tocris Biosciences) for additional 24hr, to a total of 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 density of 7.5x10^3 cells/cm^2. For ES-neuro and ES-meso differentiation, the cells were grown for 2 days in N2B27 or N2B27 supplemented with 20ng/ml FGF2 (R&D systems, 50μg/ml) and 3μM Chiron, respectively.

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 the N2B27 was supplemented with 10 ng/ml FGF2 and 5μM Chiron for additional 24hr, to a total of 72hr. To induce neural spinal cord identity (ES-neuroF) or mesodermal identity (ES-mesoF), cells were grown from day 3 – day 5 in either N2B27 supplemented with 100nM RA (Sigma) or N2B27 supplemented with 5μM Chiron, respectively.

Epi-NMP

E14-Tg2A ESCs were grown in tissue-culture plastic flasks coated with 0.5% Plasma Fibronectin (FCOLO, 1mg/ml, Temecula) in PBS (with Calcium and Magnesium). ESCs were grown in Epi-medium: N2B27 supplemented with 12ng/ml FGF2 and 25ng/ml Activin A (Stem Cells Institute 100μg/ml), with or without 20μM XAV939 (XAV Tocris Biosciences, 10mM) for at least 4 passages, To generate EpiSCs (or EpiXAV when the β-catenin inhibitor XAV is used). These cells were tested for EpiSCs character by seeding them at a colonal density (67 cells/cm^2) in restricted medium (2i: N2B27 supplemented with 3μM Chiron and 1μM PD0325901 (PD03, Tocris Biosciences, 10mM)). On observing no colony formation, it was concluded that the cells had exited naïve pluripotency and had entered the primed pluripotent state (data are not shown).

EpiSCs (treated with or without XAV) were plated at a density of 5x10^4 cells/cm^2 in a 0.5% Fibronectin pre-coated flask with Epi-media for the first day. The concentration of FGF2 was increased after 1 day to 20ng/ml in the base medium of N2B27 and Activin A or XAV (if used) were removed. On day 3, the N2B27 was supplemented with 3μM Chiron which was added to the 20ng/ml FGF2. After 72hr, the resulting population was known as Epi-NMP or EpiXAV-NMP, if XAV was used in the Epi-medium. This protocol is a variation of one previously used to derive NMP-like cells from human ESCs (Lippmann et al., 2015)

Epi- meso and Epi-neuro

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

Development • Accepted manuscript

Epi-meso differentiation

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

EM2-FP

Epi-meso (cultured without XAV) cells were detached from the culture flask using Accutase and plated back to a 0.5% Fibronectin coated flask at a density 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

ESCs mutant for Nodal (Nodal-/-) were provided by J. Collignon following derivation from the 413.d mutant mouse line (Conlon et al., 1991). They were grown on a 0.5% Fibronectin coated culture flask with Epi-medium: N2B27 supplemented with 12ng/ml FGF2 and 25ng/ml Activin A for at least 4 passages. The Nodal null EpiSCs were plated at a density of $5x10^4$ cells/cm^2 on a 0.5% Fibronectin pre-coated flask with Epi-medium for the first day. The concentration of FGF2 was increased after 1 day to 20ng/ml in the base medium of N2B27 and Activin A was removed. On day 3, the N2B27 was supplemented with 3μM Chiron which was added to the 20ng/ml FGF2. After a total of 72hr, the resulting population was known as the Nodal-/- Epi-NMPs. 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 culture 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 the Superscript III system (Invitrogen). The quantification of double-stranded DNA obtained for primer-specific genes was achieved with 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

Development • Accepted manuscript

Material Table S1. All experiments were performed in biological duplicates or triplicates. Expression values were normalized against the housekeeping gene *Ppia*. To enable comparison between different qRT-PCR experiments, each run of the qPCR included one common condition (Epi-meso, in this case). Each condition in every run was normalized to Epi-meso and averaged across biological replicates. The steps to calculate the normalized gene expression values are as follows:

1. Identify the C_t (threshold cycle) for each gene (technical triplicates) and calculate the expression values (2^{-Ct}).
2. Calculate the average and the standard deviation (std) for each gene from the triplicate expression values.
3. Divide the average and the std of each gene by the expression value for *Ppia*.
4. The normalized gene expression values in condition x are divided by 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 is 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. Average the biological replicates: F_1 and F_2 are biological replicates of the same gene in the same condition and their expression was normalized as above. 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 biological replicates (between 2 and 3).
6. Standardize 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 as 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

Development • Accepted manuscript

values for each condition: the neural averaged value and the mesodermal averaged value. 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 averaged value and a differentiation averaged value 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, permeabilised 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 of cells sticking to the tube 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 S2) were designed using Stellaris™ website and bought via Stellaris™ 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 48h of fixation 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-automatically using a MATLAB graphic user interface (GUI) developed by Marshall J. Levesque and Arjun Raj at the University of Pennsylvania or with home-made protocols written in ICY (de Chaumont et al., 2012).

Principal Component Analysis

PCA 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-score expression values of the 27 genes measured (Fig. 2A and Fig. S3B) at different stages of 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

We used cosine similarity as a measure of similarity between Z-score expression values of a list of genes in one condition versus another 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, *i.e.* conditions are opposite and 1 indicates maximal similarity. Values of 0.5 and above indicates the degree of similarity between the two conditions.

Confocal and immunostaining

Samples of the different cell cultures were grown in 4-well (Ibidi) plastic tissue-culture dishes. Samples were washed in BBS + $CaCl_2$ (50 mM BES Sodium Salt, 280 mM NaCl, 1.5 mM Na2HPO4, 1mM $CaCl_2$ adjusted to pH 6.96 with 1M 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 incubation with primary antibodies. The following day, the samples were washed with BBT and

Development • Accepted manuscript

incubated for 2hr with the desired fluorescently-conjugated secondary antibodies. 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: *T* (*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), *Cdh2* (mouse; BD Bioscience 610920 dilution 1:200) and *Cdh1* (rat; Takara M108, dilution 1:100). Secondary antibodies (Goat-A488, Rabbit-A633, Mouse-A568, Rat-A633; Molecular Probes), were raised in donkey and used at a 1:500 dilution with Hoechst 33342 (H3570, dilution 1:1000; 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 was carried out using Zen2010 v6 (Zeiss). The cells from fluorescence microscopy images were segmented manually and the quantification of the cellular fluorescence level subtracting the background was done using the open source FIJI ImageJ platform (Schindelin et al., 2012).

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 Fig S7 and (Baillie-Johnson et al., 2018; Gouti et al., 2014), using an eyebrow knife tool. Embryo cultures were imaged at single time points and as time-lapses for 15-18 hours after grafting.

Development • Accepted manuscript

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.

MATLAB and the Statistics Toolbox (2018b release, MathWorks Inc.) were used to produce the histograms in Figure S8 and to compute the statistical tests used to compare the resulting distributions (Supplementary section 4). A two-tailed Wilcoxon Rank Sum test was used to compare the axial lengths of the labelled cells' contributions after grafting. The Kruskal-Wallis one-way analysis of variation was used to make an initial comparison and was followed by a test for unequal medians (the *multcompare* function).

Development • Accepted manuscript

References

Acampora, D., Di Giovannantonio, L. G., Di Salvio, M., Mancuso, P. and Simeone, A. (2009). Selective inactivation of Otx2 mRNA isoforms reveals isoform-specific requirement for visceral endoderm anteriorization and head morphogenesis and highlights cell diversity in the visceral endoderm. *Mech Dev* **126**, 882-897.

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.

Albano, R. M., Arkell, R., Beddington, R. S. and Smith, J. C. (1994). Expression of inhibin subunits and follistatin during postimplantation mouse development: decidual expression of activin and expression of follistatin in primitive streak, somites and hindbrain. *Development* **120**, 803-813.

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.

Arkell, R. M. and Tam, P. P. (2012). Initiating head development in mouse embryos: integrating signalling and transcriptional activity. *Open Biol* **2**, 120030.

Baillie-Johnson, P., Voiculescu, O., Hayward, P. and Steventon, B. (2018). The Chick Caudolateral Epiblast Acts as a Permissive Niche for Generating Neuromesodermal Progenitor Behaviours. *Cells Tissues Organs* **205**, 320-330.

Bondue, A. and Blanpain, C. (2010). Mesp1: a key regulator of cardiovascular lineage commitment. *Circ Res* **107**, 1414-1427.

Cajal, M., Lawson, K. A., Hill, B., Moreau, A., Rao, J., Ross, A., Collignon, J. and Camus, A. (2012). Clonal and molecular analysis of the prospective anterior neural boundary in the mouse embryo. *Development* **139**, 423-436.

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.

Chalamalasetty, R. B., Dunty, W. C., Jr., Biris, K. K., Ajima, R., Iacovino, M., Beisaw, A., Feigenbaum, L., Chapman, D. L., Yoon, J. K., Kyba, M., et al. (2011). The Wnt3a/beta-catenin target gene Mesogenin1 controls the segmentation clock by activating a Notch signalling program. *Nat Commun* **2**, 390.

Chalamalasetty, R. B., Garriock, R. J., Dunty, W. C., Jr., Kennedy, M. W., Jailwala, P., Si, H. and Yamaguchi, T. P. (2014). Mesogenin 1 is a master regulator of paraxial presomitic mesoderm differentiation. *Development* **141**, 4285-4297.

Chang, H., Huylebroeck, D., Verschueren, K., Guo, Q., Matzuk, M. M. and Zwijsen, A. (1999). Smad5 knockout mice die at mid-gestation due to multiple embryonic and extraembryonic defects. *Development* **126**, 1631-1642.

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., Wong, F. C., Tatar, T., Szczerbinska, I., Halbritter, F., Colby, D., Gogolok, S., Pantier, R., Liggat, K., Mirfazeli, E. S., et al. (2017). Distinct SoxB1 networks are required for naive and primed pluripotency. *Elife* **6**.

Cunningham, T. J., Colas, A. and Duester, G. (2016). Early molecular events during retinoic acid induced differentiation of neuromesodermal progenitors. *Biol Open* **5**, 1821-1833.

Development • Accepted manuscript

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.

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.

Denans, N., Iimura, T. and Pourquie, O. (2015). Hox genes control vertebrate body elongation by collinear Wnt repression. *Elife* **4**.

Denham, M., Hasegawa, K., Menheniott, T., Rollo, B., Zhang, D., Hough, S., Alshawaf, A., Febbraro, F., Ighaniyan, S., Leung, J., et al. (2015). Multipotent caudal neural progenitors derived from human pluripotent stem cells that give rise to lineages of the central and peripheral nervous system. *Stem Cells* **33**, 1759-1770.

Deschamps, J. and van Nes, J. (2005). Developmental regulation of the Hox genes during axial morphogenesis in the mouse. *Development* **132**, 2931-2942.

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.

Dunty, W. C., Jr., Kennedy, M. W., Chalamalasetty, R. B., Campbell, K. and Yamaguchi, T. P. (2014). Transcriptional profiling of Wnt3a mutants identifies Sp transcription factors as essential effectors of the Wnt/beta-catenin pathway in neuromesodermal stem cells. *PLoS One* **9**, e87018.

Edri, S., Hayward, P., Jawaid, W. and Martinez Arias, A. (2018). Emergence of a node-like population within an in vitro derived Neural Mesodermal Progenitors (NMPs) population. *bioRxiv* **https://doi.org/10.1101/326371**

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.

Giros, A., Grgur, K., Gossler, A. and Costell, M. (2011). alpha5beta1 integrin-mediated adhesion to fibronectin is required for axis elongation and somitogenesis in mice. *PLoS One* **6**, e22002.

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.

Development • Accepted manuscript

Hebert, J. M., Boyle, M. and Martin, G. R. (1991). mRNA localization studies suggest that murine FGF-5 plays a role in gastrulation. *Development* **112**, 407-415.
Henrique, D., Abranches, E., Verrier, L. and Storey, K. G. (2015). Neuromesodermal progenitors and the making of the spinal cord. *Development* **142**, 2864-2875.
Hochgreb, T., Linhares, V. L., Menezes, D. C., Sampaio, A. C., Yan, C. Y., Cardoso, W. V., Rosenthal, N. and Xavier-Neto, J. (2003). A caudorostral wave of RALDH2 conveys anteroposterior information to the cardiac field. *Development* **130**, 5363-5374.
Khoa le, T. P., Azami, T., Tsukiyama, T., Matsushita, J., Tsukiyama-Fujii, S., Takahashi, S. and Ema, M. (2016). Visualization of the Epiblast and Visceral Endodermal Cells Using Fgf5-P2A-Venus BAC Transgenic Mice and Epiblast Stem Cells. *PLoS One* **11**, e0159246.
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.
Lawson, K. A., Dunn, N. R., Roelen, B. A., Zeinstra, L. M., Davis, A. M., Wright, C. V., Korving, J. P. and Hogan, B. L. (1999). Bmp4 is required for the generation of primordial germ cells in the mouse embryo. *Genes Dev* **13**, 424-436.
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.
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.
Morgani, S. M., Metzger, J. J., Nichols, J., Siggia, E. D. and Hadjantonakis, A. K. (2018). Micropattern differentiation of mouse pluripotent stem cells recapitulates embryo regionalized cell fate patterning. *Elife* **7**.
Naiche, L. A., Arora, R., Kania, A., Lewandoski, M. and Papaioannou, V. E. (2011). Identity and fate of Tbx4-expressing cells reveal developmental cell fate decisions in the allantois, limb, and external genitalia. *Dev Dyn* **240**, 2290-2300.
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.
Niswander, L. and Martin, G. R. (1992). Fgf-4 expression during gastrulation, myogenesis, limb and tooth development in the mouse. *Development* **114**, 755-768.
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.

Development • Accepted manuscript

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.

Papaioannou, V. E. (2014). The T-box gene family: emerging roles in development, stem cells and cancer. *Development* **141**, 3819-3833.

Parr, B. A., Shea, M. J., Vassileva, G. and McMahon, A. P. (1993). Mouse Wnt genes exhibit discrete domains of expression in the early embryonic CNS and limb buds. *Development* **119**, 247-261.

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.

Robb, L., Hartley, L., Begley, C. G., Brodnicki, T. C., Copeland, N. G., Gilbert, D. J., Jenkins, N. A. and Elefanty, A. G. (2000). Cloning, expression analysis, and chromosomal localization of murine and human homologues of a Xenopus mix gene. *Dev Dyn* **219**, 497-504.

Sakai, Y., Meno, C., Fujii, H., Nishino, J., Shiratori, H., Saijoh, Y., Rossant, J. and Hamada, H. (2001). The retinoic acid-inactivating enzyme CYP26 is essential for establishing an uneven distribution of retinoic acid along the anterio-posterior axis within the mouse embryo. *Genes Dev* **15**, 213-225.

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.

Schubert, F. R., Fainsod, A., Gruenbaum, Y. and Gruss, P. (1995). Expression of the novel murine homeobox gene Sax-1 in the developing nervous system. *Mech Dev* **51**, 99-114.

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.

Sirbu, I. O. and Duester, G. (2006). Retinoic-acid signalling in node ectoderm and posterior neural plate directs left-right patterning of somitic mesoderm. *Nat Cell Biol* **8**, 271-277.

Sirbu, I. O., Gresh, L., Barra, J. and Duester, G. (2005). Shifting boundaries of retinoic acid activity control hindbrain segmental gene expression. *Development* **132**, 2611-2622.

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

Steventon, B. and Martinez Arias, A. (2017). Evo-engineering and the cellular and molecular origins of the vertebrate spinal cord. *Dev Biol* **432**, 3-13.

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.

Takemoto, T., Uchikawa, M., Yoshida, M., Bell, D. M., Lovell-Badge, R., Papaioannou, V. E. and Kondoh, H. (2011). Tbx6-dependent Sox2 regulation determines neural or mesodermal fate in axial stem cells. *Nature* **470**, 394-398.

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

Development • Accepted manuscript

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.

Wolfe, A. D. and Downs, K. M. (2014). Mixl1 localizes to putative axial stem cell reservoirs and their posterior descendants in the mouse embryo. *Gene Expr Patterns* **15**, 8-20.

Wright, T. J., Hatch, E. P., Karabagli, H., Karabagli, P., Schoenwolf, G. C. and Mansour, S. L. (2003). Expression of mouse fibroblast growth factor and fibroblast growth factor receptor genes during early inner ear development. *Dev Dyn* **228**, 267-272.

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**.

Wymeersch, F. J., Skylaki, S., Huang, Y., Watson, J. A., Economou, C., Marek-Johnston, C., Tomlinson, S. R. and Wilson, V. (2019). Transcriptionally dynamic progenitor populations organised around a stable niche drive axial patterning. *Development* **146**.

Yamaguchi, T. P., Bradley, A., McMahon, A. P. and Jones, S. (1999). A Wnt5a pathway underlies outgrowth of multiple structures in the vertebrate embryo. *Development* **126**, 1211-1223.

Zakin, L. and De Robertis, E. M. (2004). Inactivation of mouse Twisted gastrulation reveals its role in promoting Bmp4 activity during forebrain development. *Development* **131**, 413-424.

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.

Development • Accepted manuscript

Figures

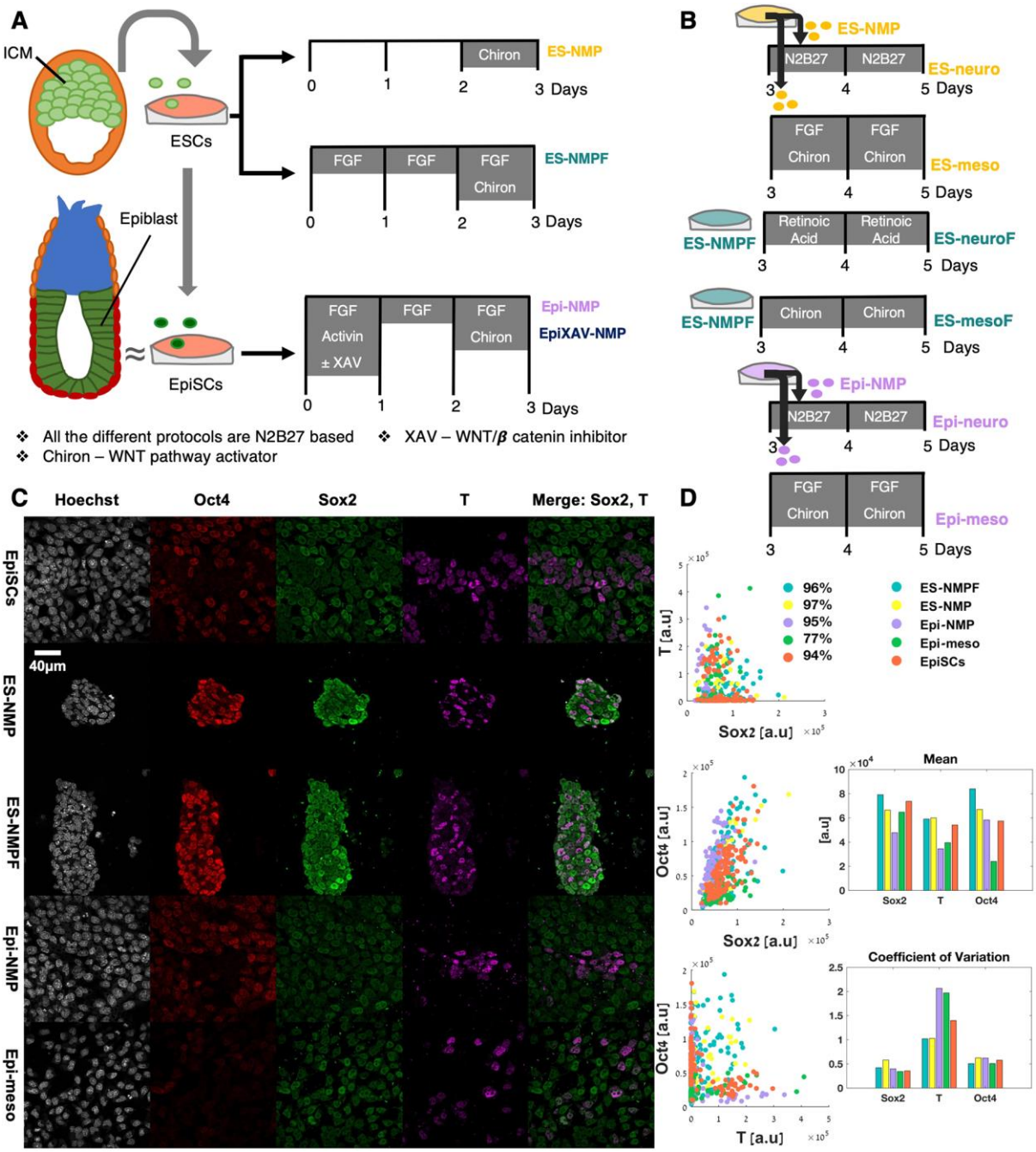

Fig 1. Comparison between *in vitro* protocols to produce NMP-like cells A. Diagram of the protocols: ES-NMP (Turner et al., 2014) ES-NMPF (Gouti et al., 2014) and Epi-NMP (Methods). **B.** Differentiation NMP-like cells into neural and mesodermal lineages**.** ES-NMP (yellow) and Epi-NMP (purple) were split into 2 flasks and cultured for 2 days in a medium that allows differentiation to either neural or mesodermal cells (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 (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. **C.** Confocal images of EpiSCs, ES-NMP, ES-NMPF and Epi-NMP on their 3rd day and Epi-meso on its 2nd day. Hoechst (Nuclei) in

Development • Accepted manuscript

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 (Methods). **D.** Quantification plots of the fluorescence intensity in arbitrary units (a.u) representing the protein levels in a cell. Each dot represents a cell where the x-axis and y-axis represent the fluorescence intensity of either *Sox2*, *T* and *Oct4*. The numbers next to the *T* versus *Sox2* plot indicate the percentage of cells that coexpress *Sox2* and *T* at the different conditions. The mean and the coefficient of variation (CV) of the distribution of *Sox2*, *T* and *Oct4* across the different conditions are presented on the right-hand side: EpiSCs in orange, ES-NMP in yellow, ES-NMPF in turquoise, Epi-NMP in purple and Epi-meso in green.

Development • Accepted manuscript

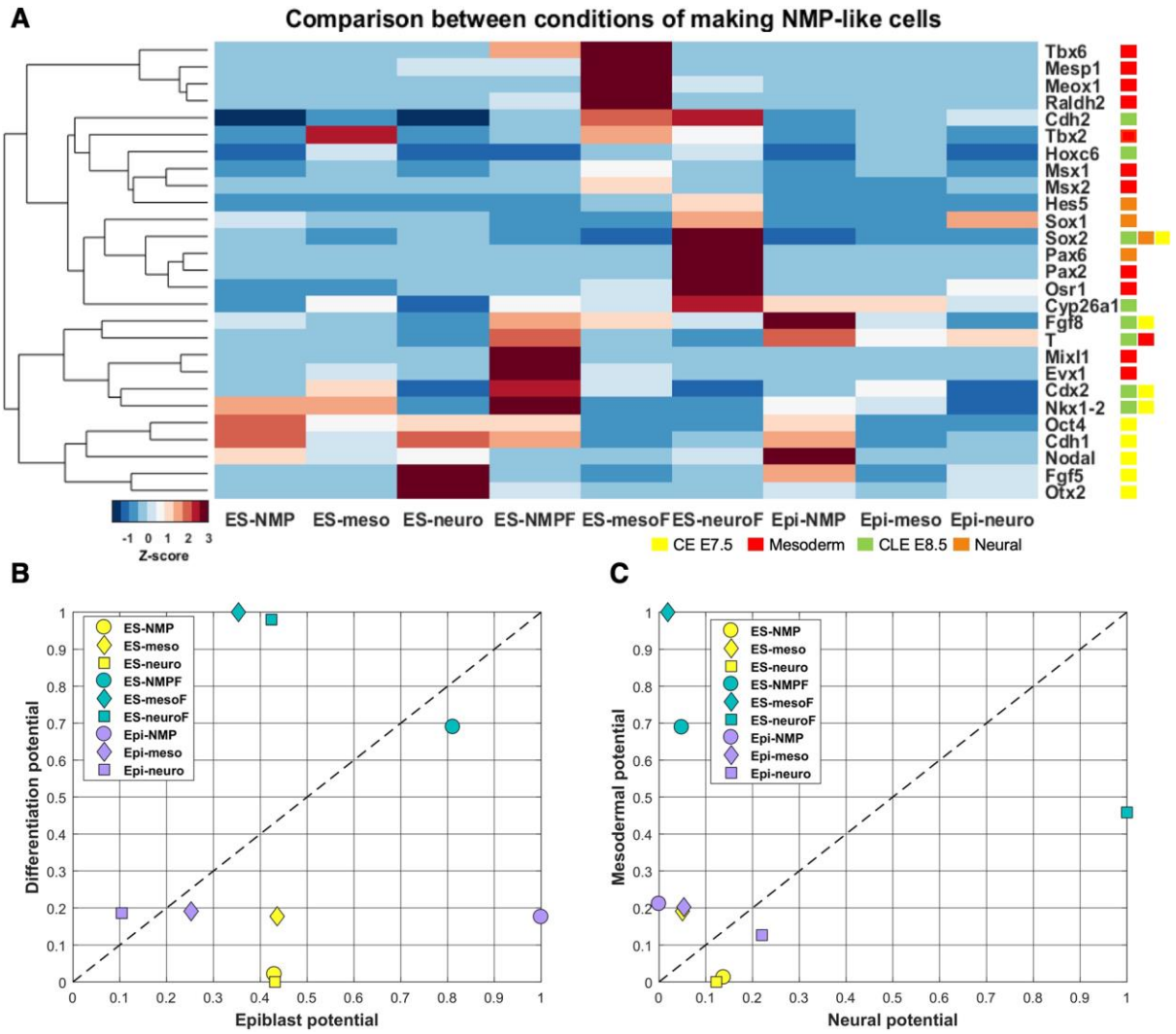

Fig 2. The neural and mesodermal index of the NMP-like cells and their derivatives. A. Expression heatmap of 27 genes, obtained by RT-qPCR, in cells grown in different conditions, as indicated in Fig. 1A-B. The expression of each gene was normalized to the expression of the Epi-meso condition and then was scaled across the different conditions through the calculation of Z-score (Methods and Fig. S3). Each gene was assigned to a label according to Fig. S1 and Supplementary sections 1-2: CE E7.5 in yellow, mesoderm in red, CLE E8.5 in green and neural in orange. **B.** 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 averaged values in the x-axis and Differentiation averaged values in the y-axis (Methods), highlighting the epiblast state of each condition. **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 averaged values in the x-axis and Mesodermal averaged values in the y-axis (Methods), highlighting the neural-mesodermal state of each condition.

Development • Accepted manuscript

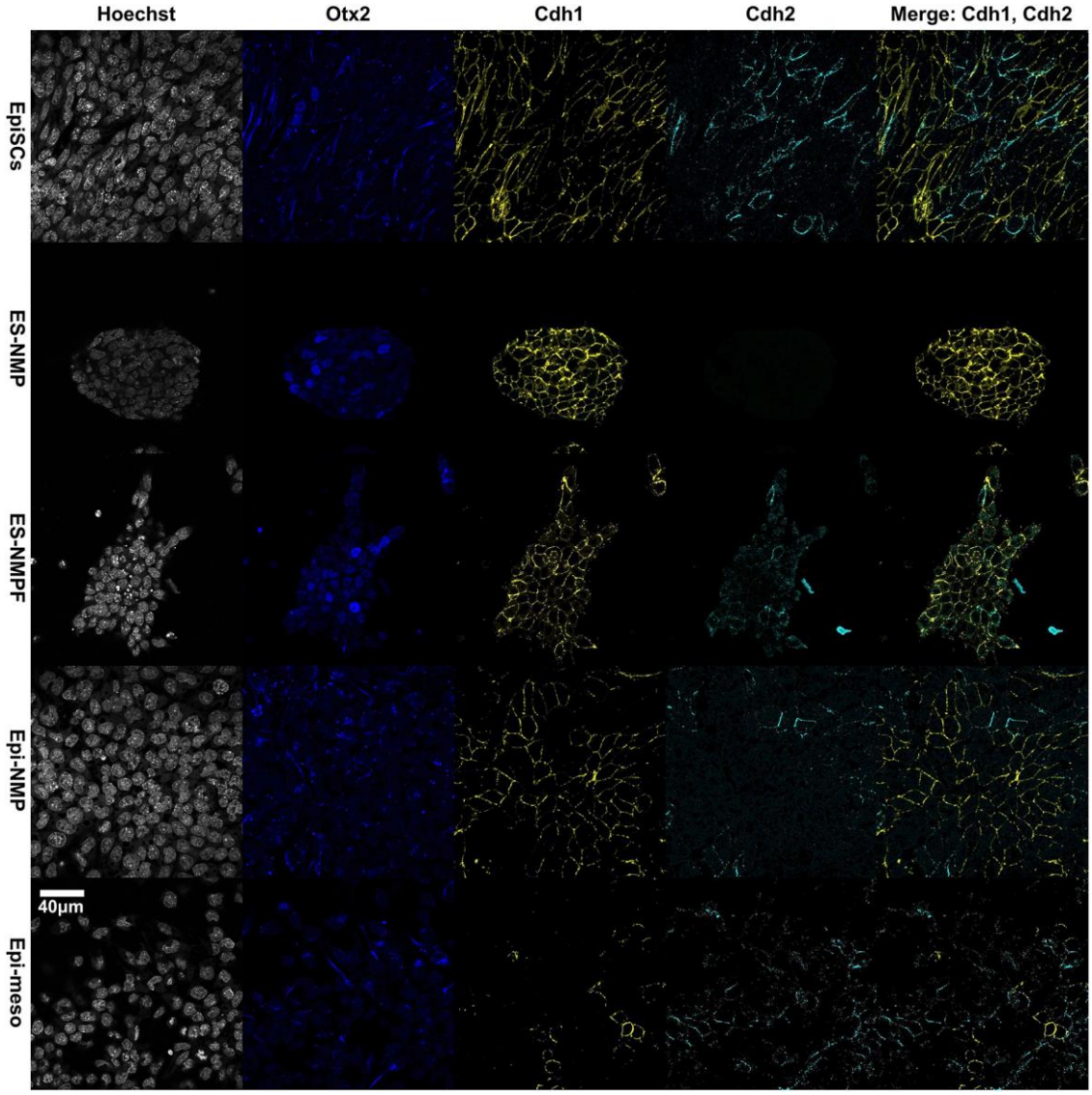

Fig 3. Mutual exclusive expression of *Cdh1* and *Cdh2* in the EpiSCs conditions. Confocal images of EpiSCs, ES-NMP, ES-NMPF and Epi-NMP on their 3rd day and Epi-meso on its 2nd day. Hoechst (nuclei) in grey, *Otx2* in blue, *Cdh1* in yellow and *Cdh2* in cyan were imaged by a confocal microscopy (Methods). The composite image of *Cdh1* (yellow) and *Cdh2* (cyan) is presented on the right-hand side column.

Development • Accepted manuscript

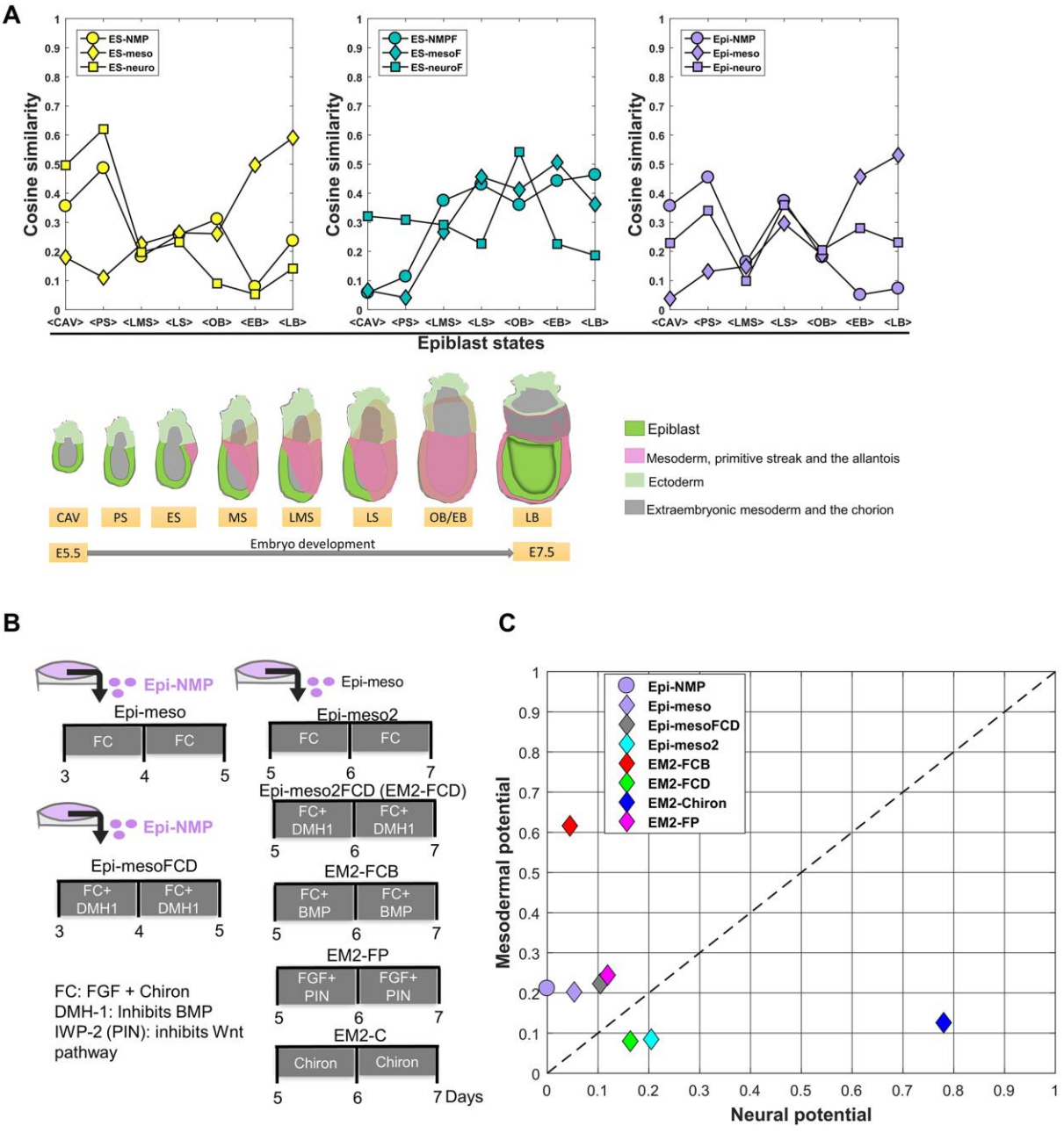

Fig 4. Comparison of the *in vitro* protocols to different epiblast stages and the effect of Wnt, FGF and BMP on mesodermal differentiation of the Epi-NMP population. A. Microarray gene expression data of the epiblast/ectoderm (excluding the primitive streak) from different stages of the mouse embryo (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 between the expression values of the 27 genes shown in Fig. 2A in the *in vitro* populations and the expression values of these genes in the different stages of the epiblast mouse embryo (Fig. S4A-B and Methods). The y-axis represents the average cosine similarity between *in vitro* populations to the same epiblast mouse stages (x-axis) as shown in Fig. S4B. Value of 0 indicates dissimilarity and value of 1 indicates maximal similarity (Methods). **B.** Differentiation protocols for Epi-NMP into Epi-meso and Epi-meso2 with modulation of BMP signalling

Development • Accepted manuscript

(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 (Fig. S5 and 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. S5C.

Development • Accepted manuscript

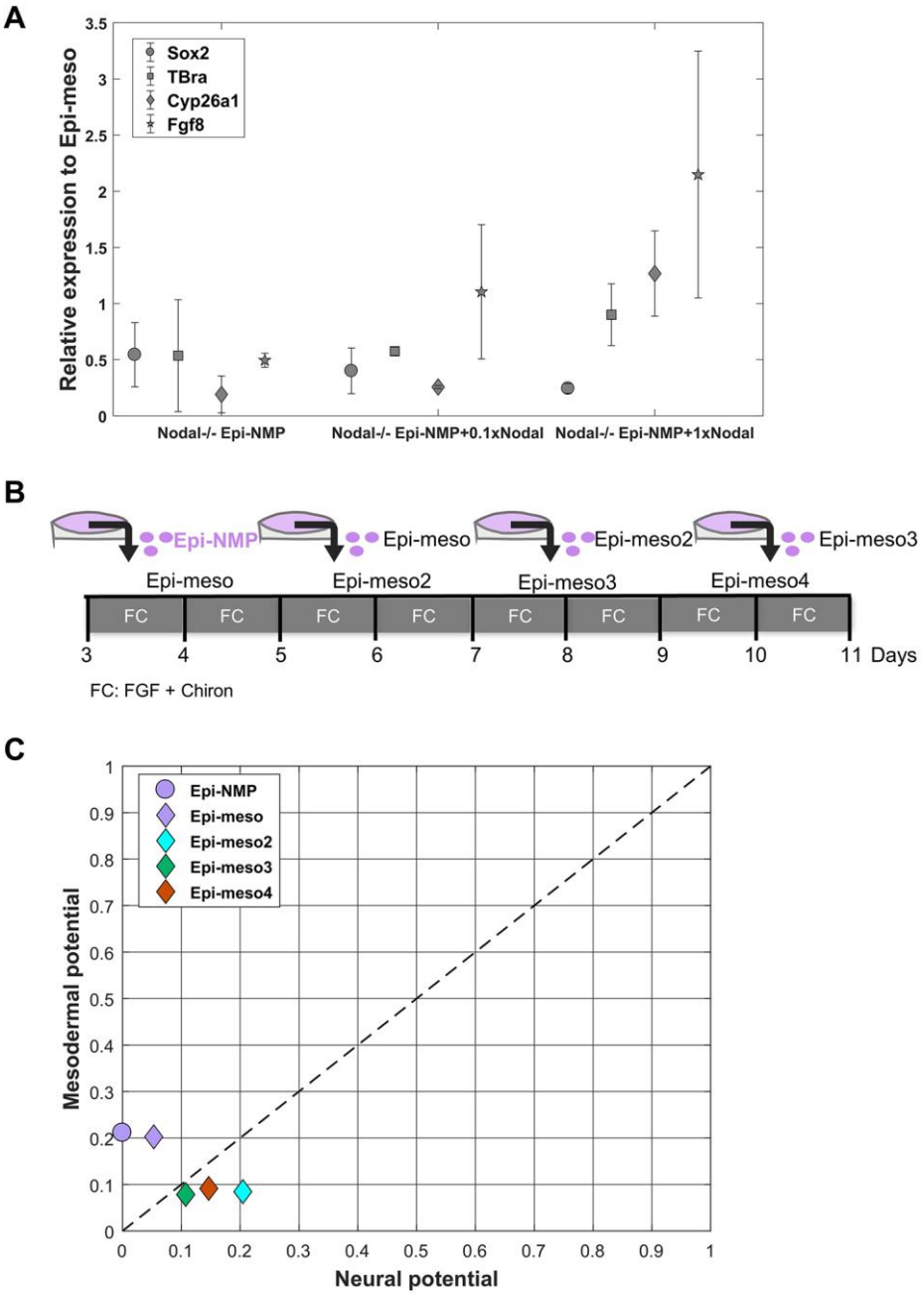

Fig 5. The effect of Nodal signalling and the maintenance of Epi-NMP in culture. **A.** Nodal mutant cells were cultured in the Epi-NMP protocol (Nodal -/- Epi-NMP). This population of cells was compared to the same population just with the addition of 2 doses of Nodal concentration to the growing medium 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 and the error bars represent the genes average expression and the standard error across biological replicates obtained by RT-qPCR respectively. The gene expression across the different conditions was normalized to Epi-meso condition (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 derivatives was calculated based on the Z-score of the normalized expression profile obtained by RT-qPCR shown in Fig. S6B.

Development • Accepted manuscript

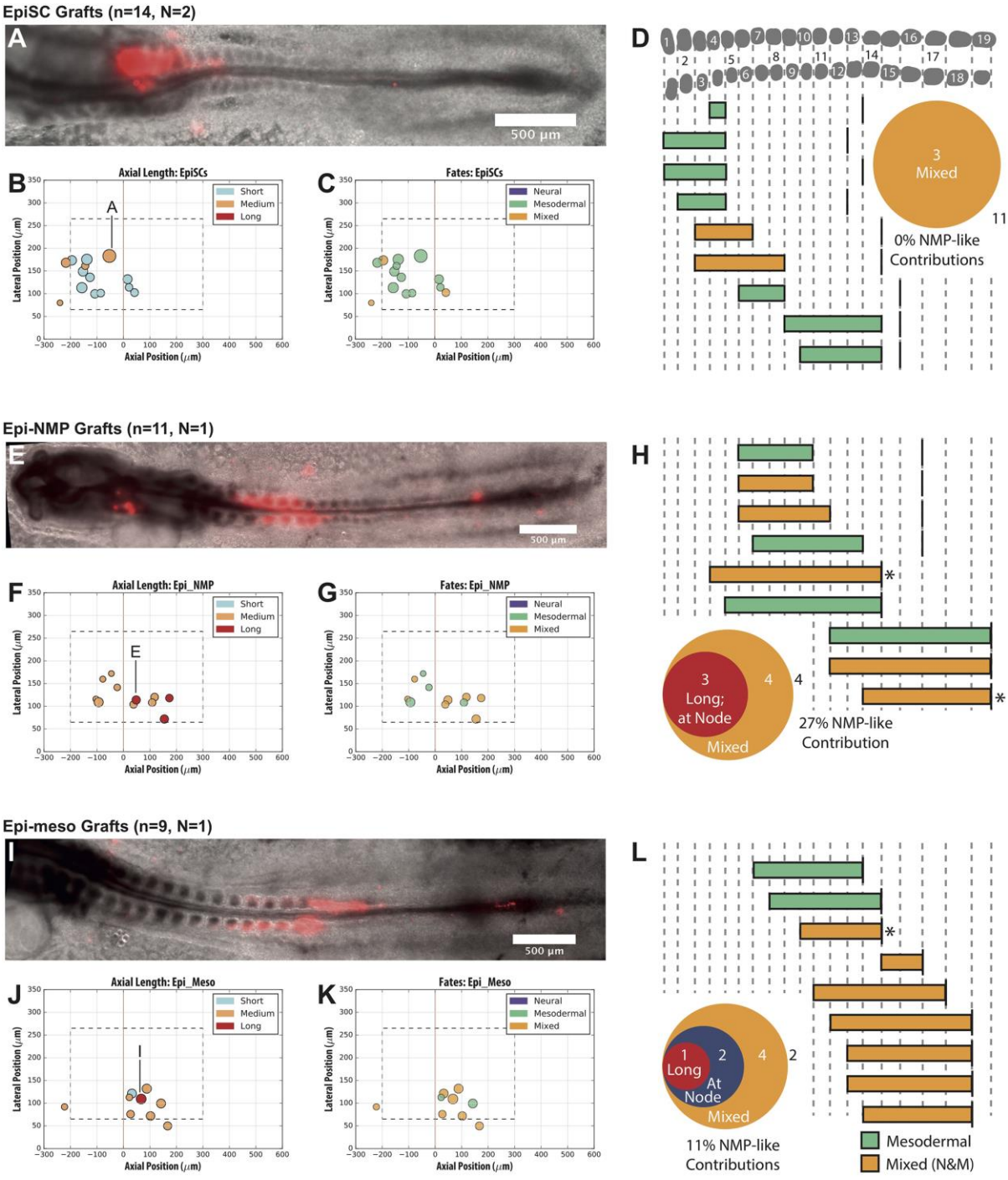

Fig 6. Epi-NMP and Epi-meso progressively contribute to more posterior portions of the embryonic body axis than EpiSCs. Colonies of EpiSC, Epi-NMP and Epi-meso cultures were labelled with a membrane dye (DiI; Methods) and were grafted into a region of the caudal lateral epiblast of the chicken embryo at HH stages 6-9. **A**, **E** and **I**. Representative images of recipient embryos are shown 15-18 hours after grafting. The initial position of each graft was measured axially and laterally from the caudal limit of the node and was used to plot the positions of the grafts in relation to the node (at 0,0). n denotes the number of embryos analysed and N indicates the number of biological replicate experiments. **B**, **F** and **J.** The starting position of the grafts, coloured according to the length of the final contribution of the labelled cells to the body axis (short, <500µm, blue; medium, 500-1750µm, orange; long, >1750µm, red). **C**, **G** and **K.** The starting position of the grafts,

Development • Accepted manuscript

coloured according to the contribution of the labelled cells to either neural tissue only (purple), mesodermal tissue only (green) or both tissues (orange). **D**, **H.** and **L.** Schematic of the grafted cell contributions to the embryonic axis, alongside a cartoon of the somatic compartment of an HH stage 13 embryo. Grafts made into hosts earlier than HH Stage 7 were excluded from this representation as they contributed more frequently to anterior somitic mesoderm (somites 1-6). The length of each grafted cell contribution is shown as a solid bar, coloured according to whether the cells contributed to mesoderm only (green) or both neural and mesodermal tissues (orange). In each case, a solid line shows the caudal boundary of the most recently formed somite of the host embryo at the endpoint. Where the bar abuts the line, labelled cells could be found into the unsegmented region of the body axis. Long contributions (>1750μm) are denoted with asterisks. The Venn diagrams in D, H and L show the numbers of grafts from the whole dataset that contributed to both neural and mesodermal tissues (orange), those that additionally retained cells in the region around the node (blue) and those that were also long (>1750μm, red). The numbers of grafts that did not follow these three criteria is outside the Venn diagram. The percentage of NMP-like contributions is calculated as the fraction that fulfilled all three criteria (*i.e.* the red circle).

Development • Accepted manuscript

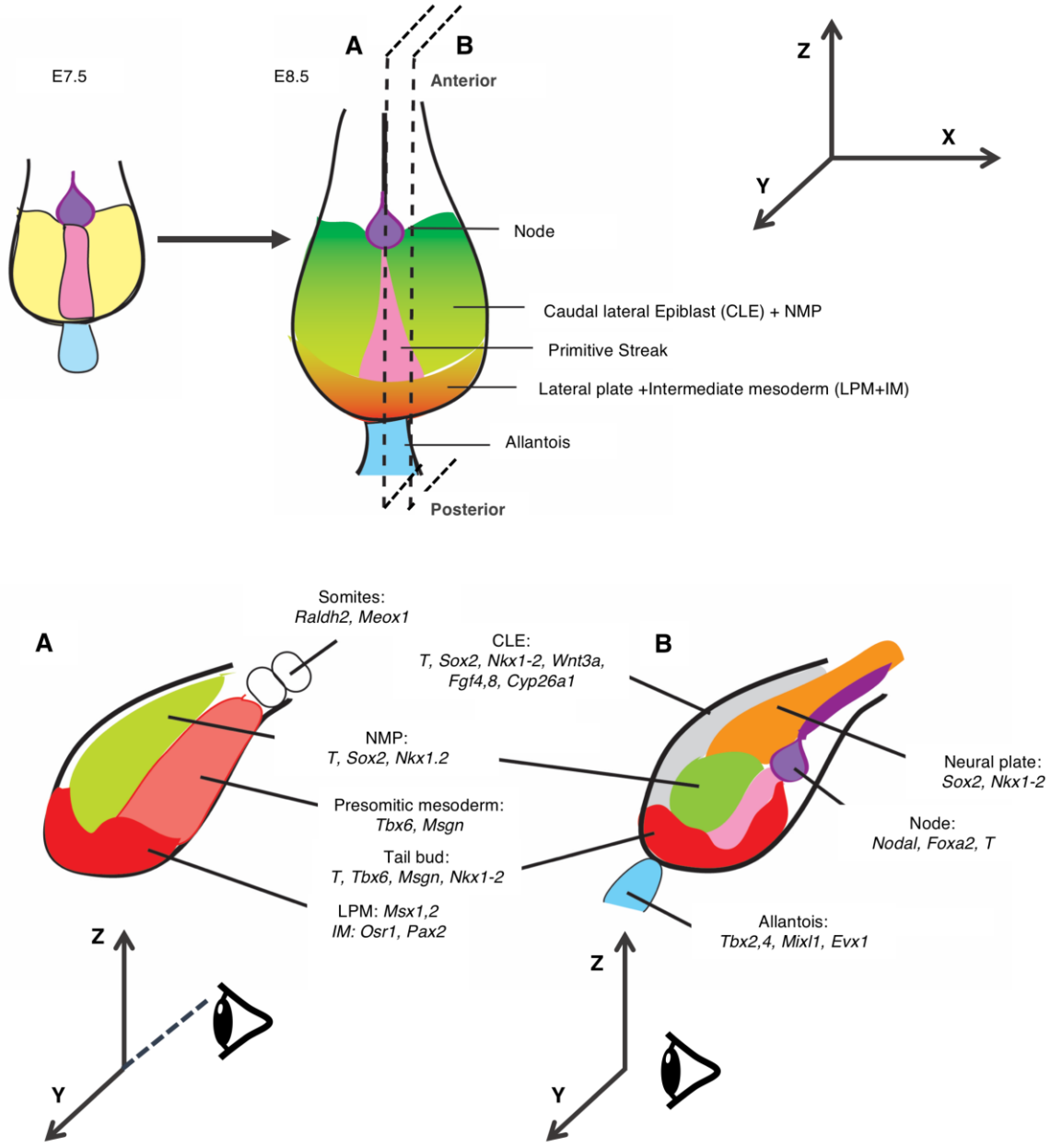

Fig S1 - Organization and gene expression patterns in the caudal region of the mouse embryo at E8.5; top, ventral views; bottom: lateral (**A**) and medial (**B**) views. The caudal region of the embryo is derived from the posterior epiblast of E7.5 (yellow) when the primitive streak (pink) reaches the 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 S1.

Development • Supplementary information

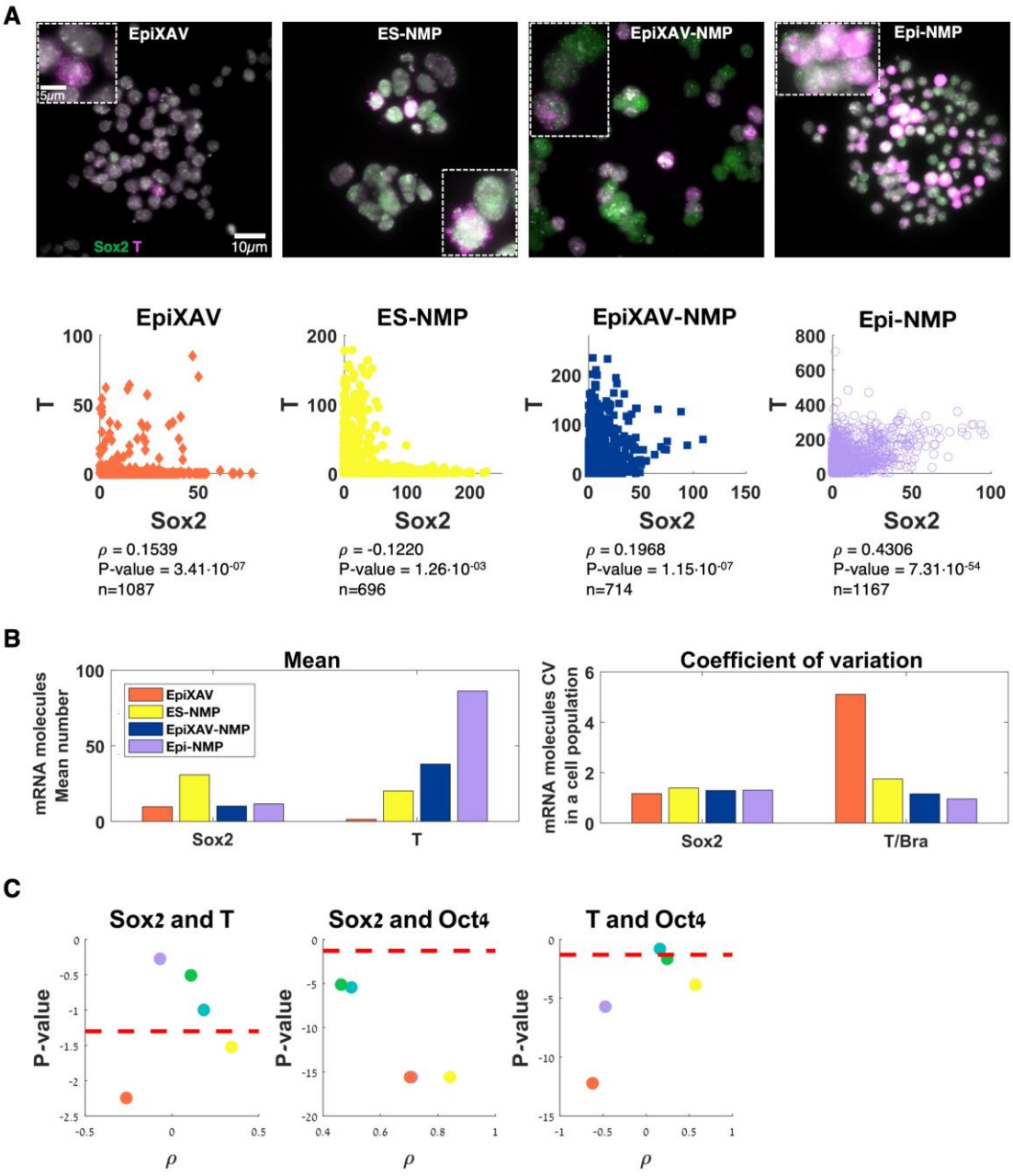

Fig S2. Coexpression of *Sox2* and *T* at mRNA levels in EpiSCs and the different NMP protocols. A. Images, obtained by using single molecule fluorescence in-situ hybridization (sm-FISH), of cells expressing *Sox2* (in green) and *T* (in magenta) mRNA in different conditions: ES-NMP, EpiXAV, EpiXAV-NMP 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 underneath the quantification plot. **B.** The mean and the coefficient of variation

Development • Supplementary information

(CV) of the distribution of *Sox2* and *T* in the different conditions: EpiXAV in orange, ES-NMP in yellow, EpiXAV-NMP in blue and Epi-NMP in purple. **C.** Spearman correlation between the fluorescence intensity representing the protein levels of *Sox2*, *T* and *Oct4* (Fig. 1C-D). The y-axis represents the P-values in logarithmic scale and the x-axis the Spearman coefficient correlation, noted as ρ. The red dashed line indicates P-value =0.05.

Development • Supplementary information

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 replicates obtained by RT-qPCR and the error bars indicate the standard error between those replicates. The gene expression across the different conditions was normalized to Epi-meso condition (Methods).

Development • Supplementary information

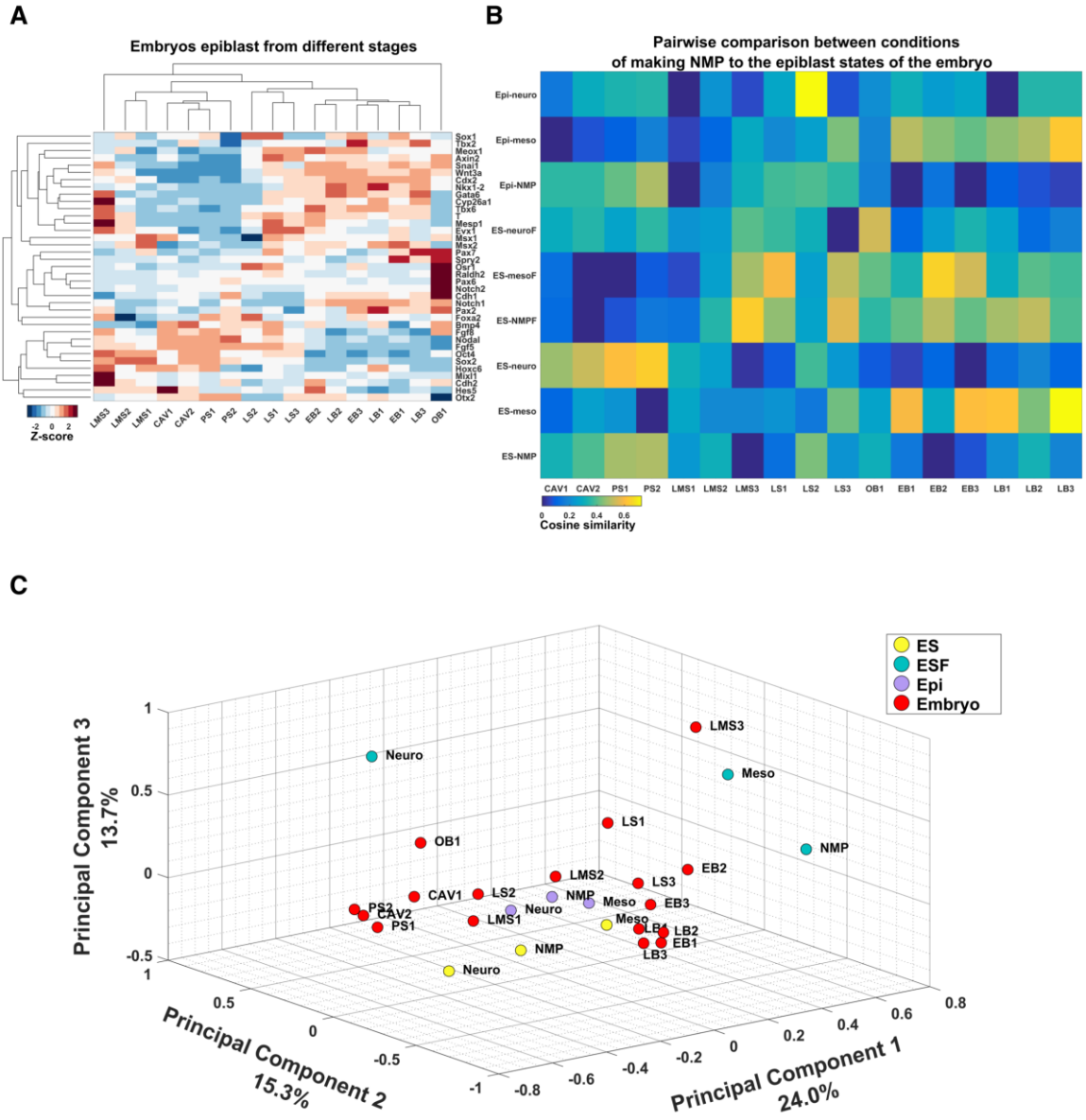

Fig S4. Comparison of expression of set of genes between the 3 *in vitro* protocols and the different embryo epiblast stages. **A.** Gene profile of the embryonic epiblast at different stages. Heatmap of the Z-score expression of the 27 genes, was calculated from the microarray gene expression data of 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. S4A 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 (Methods). **C**. PCA was performed on the expression of the 27 genes, indicated in Fig. S4A, expressed in the *in vitro*-derived NMP populations, their differentiated derivatives and the different stages of the mouse embryo epiblast.

Development • Supplementary information

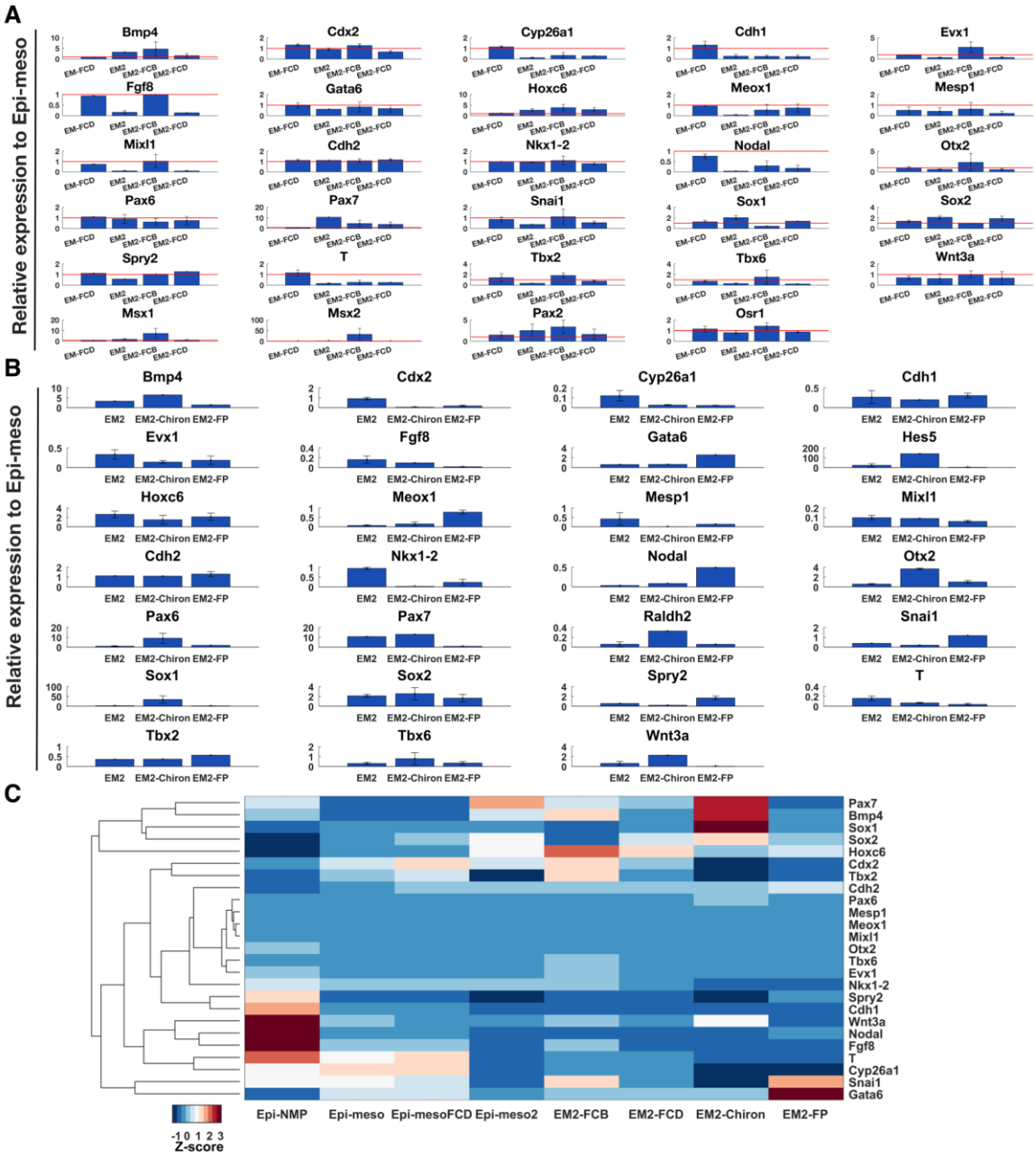

Fig S5. The influence of promoting or inhibiting BMP, FGF and Wnt. A-B. Expression of a set of chosen genes to monitor the differentiation of cells grown as indicated in Fig. 4B. The bars represent the genes' average expression across biological replicates obtained by RT-qPCR and the error bars indicate the standard error between these replicates. The gene expression across the different conditions was normalized to Epi-meso condition (Methods). **A.** The effect of BMP: measuring the expression of 29 genes in Epi-meso (indicated as the horizontal red line) in comparison to Epi-meso condition supplemented with the 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 the Epi-meso2 condition with Chiron alone (no exogenous FGF, EM2-Chiron) or FGF with Wnt pathway inhibitor IWP-2 (EM2-FP) (Methods). **C.** Expression heatmap of Fig. S5A-B: all measurements were obtained by RT-qPCR and the normalized

Development • Supplementary information

expression of each gene across the different conditions was scaled via calculating the Z-score (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 culture 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.

Development • Supplementary information

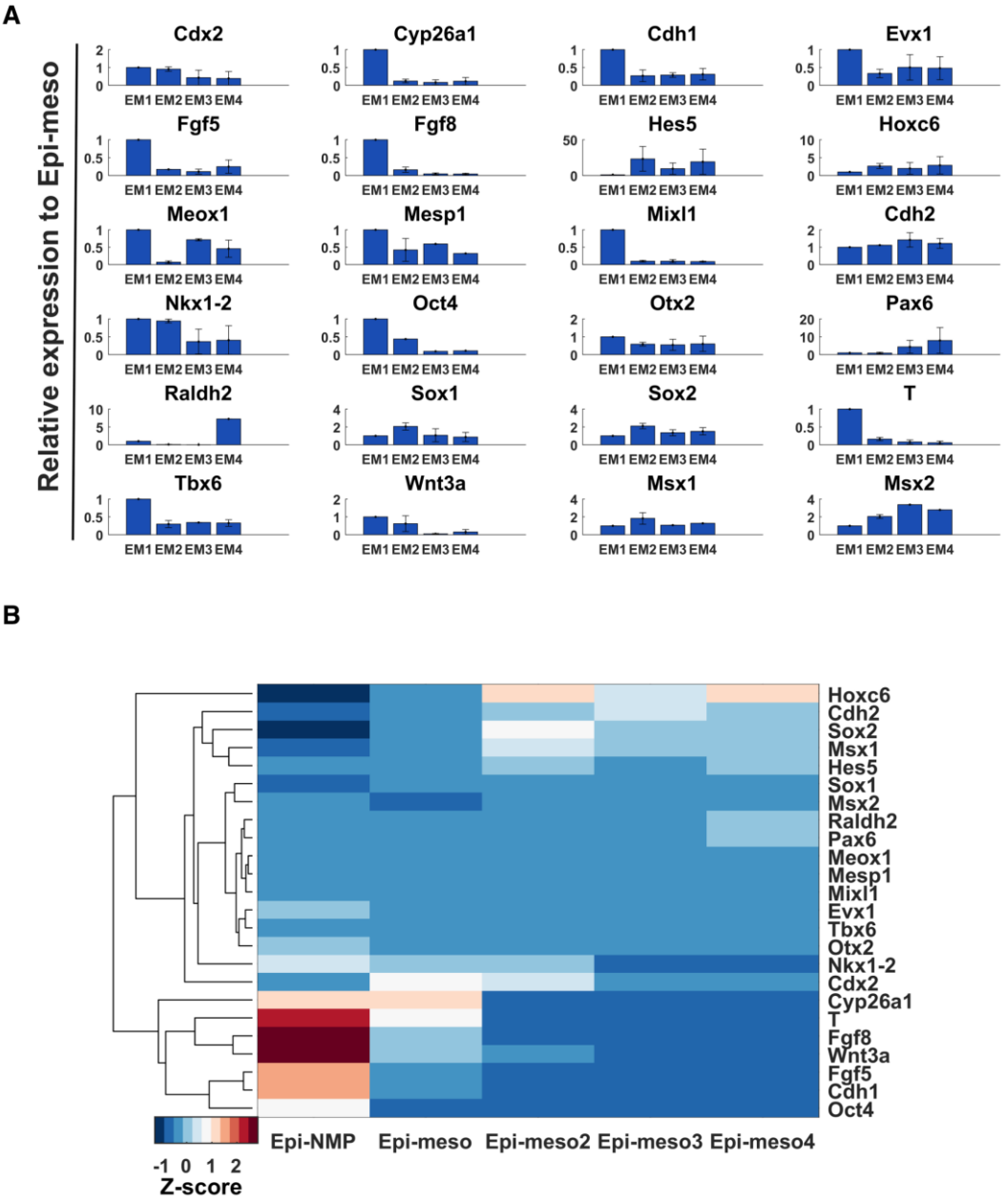

Fig S6. Culturing Epi-NMP over time. **A.** Measuring the expression of 24 genes in Epi-meso (EM1) and its derivatives: Epi-meso2, Epi-mes3 and Epi-mes4 (EM2, EM3 and EM4). The bars represent the genes average expression across biological replicates obtained by RT-qPCR and the error bars indicate the standard error between those replicates. The gene expression across the different conditions was normalized to Epi-meso condition (Methods). **B.** Expression heatmap of Fig. S6A: 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 (Methods).

Development • Supplementary information

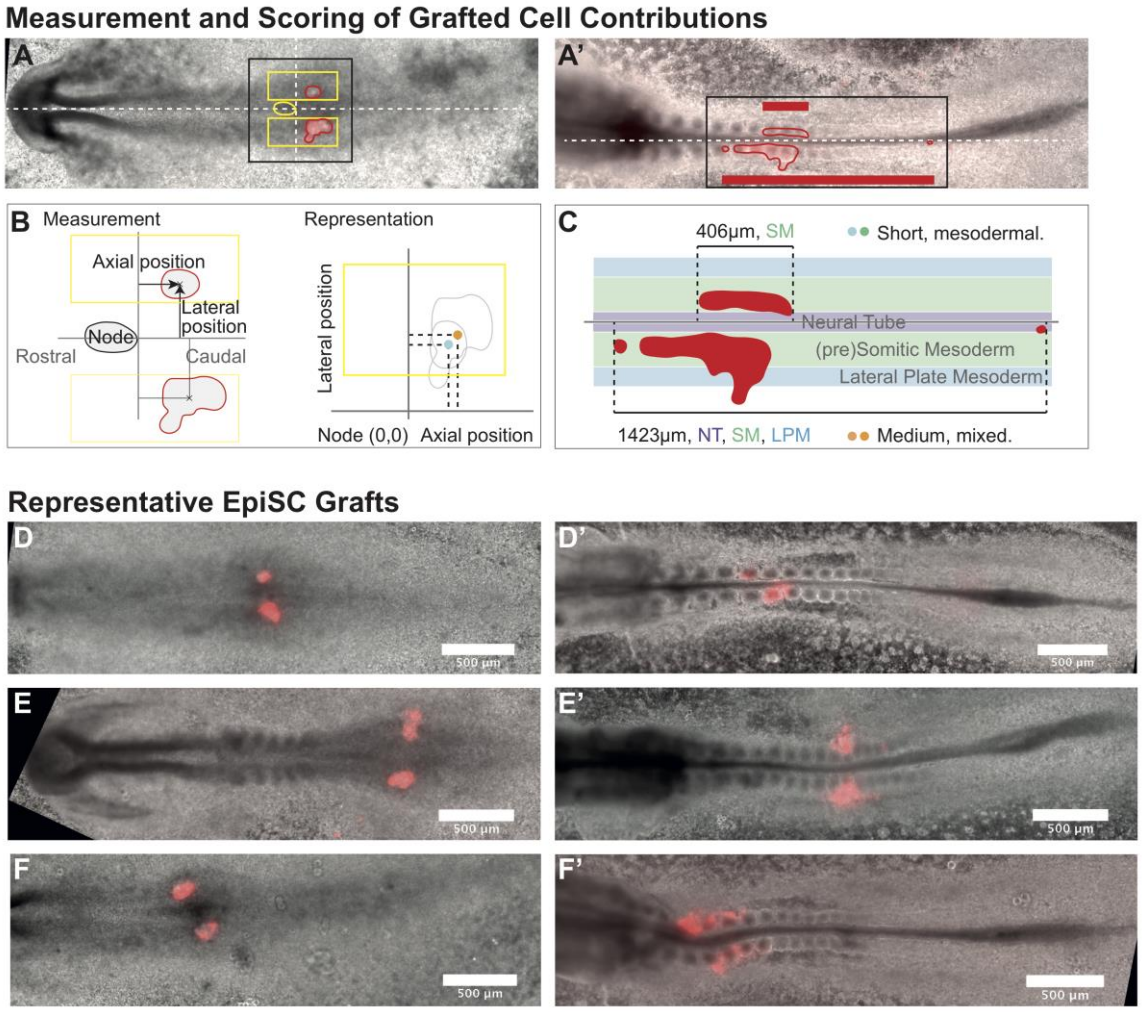

Fig S7. Measurement and scoring strategy for assessing grafted cell contributions to the embryonic axial tissues. A. Different populations of DiI-labelled cells were grafted into host chicken embryos between HH Stages 6-9. Grafted embryos were imaged immediately after grafting (as in A.) and after 15-18 hours' incubation (as in A.'). **B.** The starting position of the grafts was measured along the rostro-caudal axis and the medio-lateral axis, from the caudal limit of the node to the centre of the graft. These data were collated and plotted on Cartesian coordinates as x- and y-displacements from the origin (representing the node). **C.** The final position of the labelled cells within the embryo was determined by inspection, using the visible boundaries between the neural tube (NT), somitic mesodermal (SM) and lateral plate mesodermal (LPM) compartments to classify the grafts into neural only, mesodermal only, or mixed (neural and mesodermal) contributions. The length of the grafted cells' contribution was determined by measuring along the rostro-caudal axis from the most rostral to the most caudal labelled cell on one side of the midline; this is denoted by the red bars in A'. Live imaging data was used in some cases to more precisely follow grafted cell contributions. **D.-F'.** Representative images of single EpiSCs grafted embryos at either the start (D, E, F) or end (D', E' and F') of their incubation period.

Development • Supplementary information

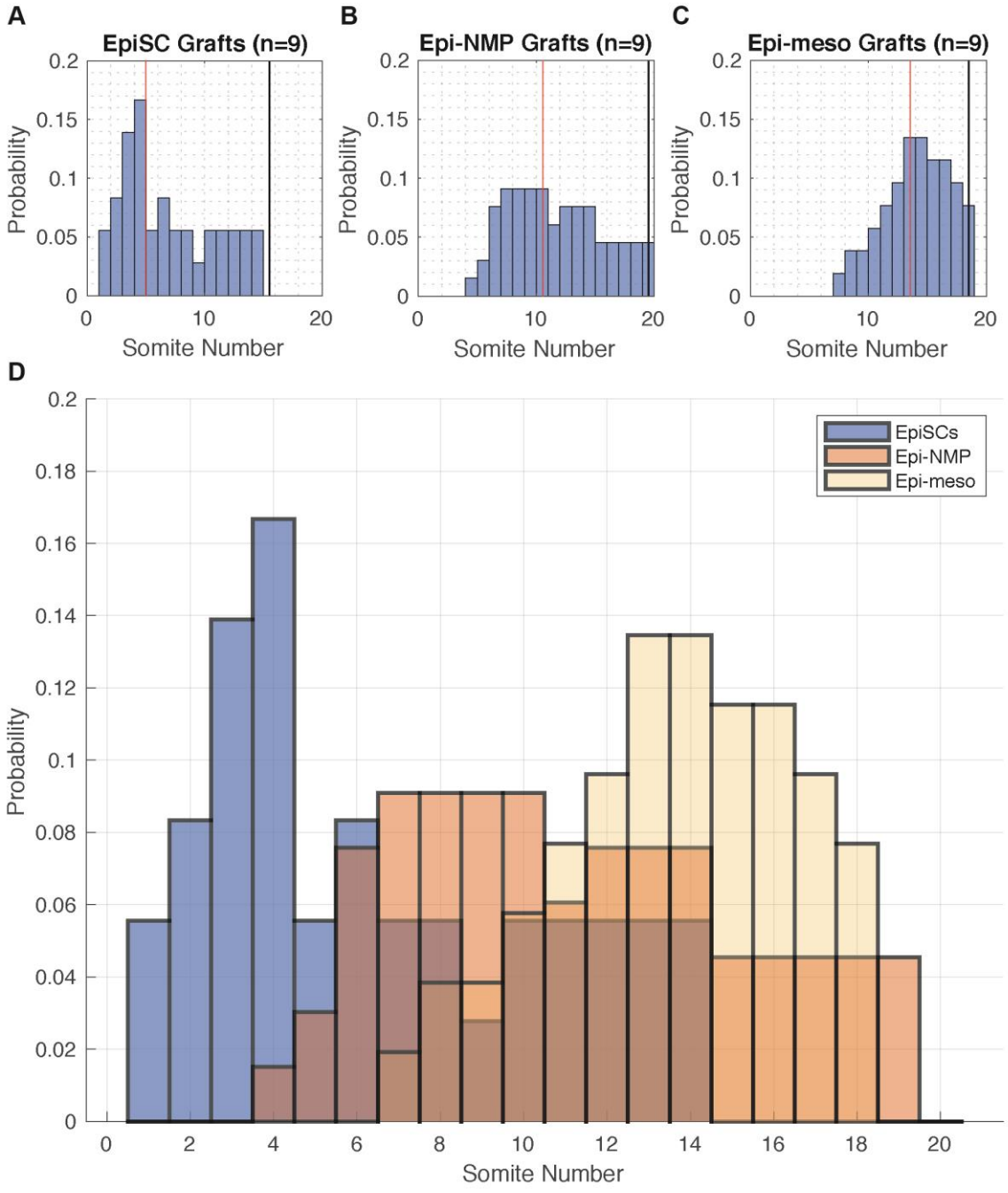

Fig S8. Histograms of Axial Contributions, Scored by Somite Level. The contributions of labelled cells to the host embryos was scored 15-18 hours after grafting according to the axial levels of the host's corresponding somites (summarised in Fig. 6D, H, L). The resulting axial levels were collated for each population, which allowed the probability to be calculated of a given somitic level containing the grafted cells. These results are presented as histograms for each population: EpiSC grafts in **A** (blue in **D**), Epi-NMP grafts in **B** (orange in **D**) and Epi-meso grafts in **C** (yellow in **D**). All three histograms are overlain in **D**, for ease of comparison. The red lines in **A**, **B** and **C** denote the median of each distribution, while the black line denotes the last formed somite in each dataset (*i.e.* the oldest embryo). On comparing these distributions with the

Development • Supplementary information

Kruskal-Wallis one-way analysis of variance and a follow-up test for unequal medians, all three populations were found to be significantly different from one another (with regards to their mean ranks). These samples could further be ordered in making more anterior to more posterior axial contributions after grafting as: EpiSCs, Epi-NMPs, Epi-meso (see Supplementary section 4).

Development • Supplementary information

Supplementary Materials

1. The caudal region of the mouse embryo

In order to evaluate whether the NMP-like 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 S1) to create a coarse grained reference map of the Caudal Epiblast (CE) at E8.5 (Fig. S1), as this is the first time that self renewing NMPs can be identified and tested functionally. These progenitors do not change much until the emergence of the Chordo Neural Hinge (CNH) at E9.5, and thus are a good reference for our study.

The E8.5 caudal region of the mouse embryo is derived from the posterior epiblast of the E7.5 embryo after proliferation and rearrangements and it is defined as a structure posterior and lateral to the node (Fig. S1). 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 a map of its gene expression landscape to divide the CE into several domains, each with a specific gene expression signature, as shown in Fig. S1 (see Supplementary information S1 for details). Two very clear domains can be observed within this region: an anterior to posterior gradient of *Sox2* and *Nkx1-2* expression and a posterior to anterior gradient of *T* expression. The Caudal Lateral Epiblast (CLE) region, harbours the NMPs (Cambray and Wilson, 2007; Wymeersch et al., 2016) and, in its most posterior region, 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 the region around and posterior-lateral to the node, by its gene expression signature as well as the genes expressed earlier at E7.0, that are not expressed at E8.5 *e.g. Cdh1*, *Oct4*, *Otx2* and *Fgf5* (Supplementary information S1).

2. Supplementary information S1

The gene expression patterns in the CLE at E8.5 shown in Fig. S1 were collated from several published sources; some examples of their expression can be found in the references below and in the Mouse Genome Informatics gene expression database (http://www.informatics.jax.org/). These references were also used to compile the table below that summarizes changes in expression between E7.5 and E8.5/9.0 in the caudal region of the embryo. Fig. S1 and the table below were used as a reference to compare the gene expression profiles in the embryo with those of the cells derived from the different *in vitro* differentiation protocols.

Gene	***E7.5***	***E8.5/9.0***	***Comment***
Sox2	ON	on	
T	ON	ON	
Cdh1	ON	OFF	
Eomes	ON	OFF	
Fgf5	ON	OFF	
Fst1	ON	OFF	
Mixl1	ON	OFF*	*At E8.5 there is some expression at the tip of the tail bud and in the allantois
Oct4	ON	OFF	
Otx2	ON	OFF	
Raldh2	ON	OFF	

Development • Supplementary information

Cdh2	OFF	ON	
Cyp26a1	OFF	ON	
Bmp4	ON	ON	
Hox5-9	OFF	ON	
Cdx2	ON	ON	
Evx1	on	on	Most of the expression is at the posterior end of the tail bud and in the allantois
Fgf4	ON	on	Expression moves from the primitive streak to a region of the dorsal CE
Fgf8	ON	ON	
Hox1-4	ON	ON	
Mesp1	ON	on*	*Mainly in the allantois
Tbx2	OFF	ON	
Tbx4	OFF	ON	
Tbx6	ON	ON	
Nodal	ON	on*	*In the node
Wnt3a	ON	ON	
Wnt5a	OFF	ON	
Nkx1-2	ON	ON	

Gene expression magnitude: ON > on > OFF (no expression).

References with examples of the spatial distribution of specific genes that have been used to underpin the expression patterns used in this study:

Bmp4 (Lawson et al., 1999; Zakin and De Robertis, 2004)
Cdx1,2, 4 (Deschamps and van Nes, 2005)
Cyp26a1 (Sakai et al., 2001; Sirbu et al., 2005)
Evx1 (Cambray and Wilson, 2007; Schubert et al., 1995)
Fgf5 (Hebert et al., 1991; Khoa le et al., 2016)
Fgf8 (Cunningham et al., 2015; Dunty et al., 2008; Sirbu and Duester, 2006)
Fgf4 (Niswander and Martin, 1992; Wright et al., 2003)
Fst1(Albano et al., 1994; Cunningham et al., 2016)
Mesp1(Bondue and Blanpain, 2010)
Mixl1 (Dunty et al., 2014; Robb et al., 2000; Wolfe and Downs, 2014)
Nkx1-2 (Henrique et al., 2015; Schubert et al., 1995)
Oct4 (Downs, 2008)
Otx2 (Acampora et al., 2009; Cajal et al., 2012)
Raldh2 (Hochgreb et al., 2003)
Sox2 (Henrique et al., 2015)
Tbx2 (Chang et al., 1999)
Tbx4 (Naiche et al., 2011; Papaioannou, 2014)
Tbx6 (Chalamalasetty et al., 2011; Chalamalasetty et al., 2014; Dunty et al., 2008)
Wnt3a (Cambray and Wilson, 2007; Giros et al., 2011; Parr et al., 1993)
Wnt5a (Yamaguchi et al., 1999)

Development • Supplementary information

3. Developmental stage reference

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 (E7.5) (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 (Fig. S4C) enabled us to recapitulate the trajectory obtained using the complete microarray data mainly along PC1, thus validating the use of our reduced set of genes for mapping states (Methods).

4. Grafts transplant statistics

There are two points to address statistically to understand the results of the *in vitro* grafts into chicken embryos.

Are there differences in the extent of the axial contribution of the three cell types?
This was addressed by comparing the absolute lengths of the labelled cells' contributions (measured as in Fig. S7) between the three populations. The length distributions were treated as independent populations, with the null hypothesis that the distribution of lengths within each population was the same (As for the construction of Figure 6 D,H,L; host embryos that were grafted at stage 6 or before were excluded from the analysis).
Pairwise comparisons were made between the populations using a two-tailed Wilcoxon Rank Sum test, with the following results:

Comparison (n)	Mean values (μm)	Standard deviation (μm)	P value
EpiSCs (9) vs. Epi-NMPs (11)	397.15 vs. 1365.8	145.10 vs. 734.7	2.66 x 10-4
EpiSCs (9) vs. Epi-meso (9)	397.15 vs. 986.24	145.10 vs. 370.86	2.88 x 10-4
Epi-NMPs (11) vs. Epi-meso(9)	1365.8 vs. 986.24	734.7 vs. 370.86	0.4033

Both of the Epi-NMP and the Epi-meso populations differed significantly in the lengths of their axial contributions from the EpiSC population.
There was not a significant difference in the lengths of the axial contributions between the Epi-NMP and Epi-meso populations (*i.e.* the data could have been sampled from populations with the same distribution).

Are there differences in the segmental level of the axial contribution of the three cell types?
This was addressed with a more detailed analysis of the data presented in Figure 6 D,H,L. For each cell type, the segmental level of its axial contributions was described as a list of somites to which it contributed. (For example, the first two EpiSC grafts in Fig 6D gave a list of somites 1, 2, 3, 4, 4). This allowed the probability of labelling for each somite to be calculated. For each cell type, the contributions to the first twenty somites were shown as histograms of probability mass in Figure S8.
The Kruskal-Wallis one-way analysis of variance was used to compare these discrete distributions with a non-parametric test. The null hypothesis was that the segmental level of the axial contributions has the same distribution across all three groups.
The returned P-value (2.2862×10^{-11}) rejects the null hypothesis that all three data samples come from a population with the same distribution at the 1% significance level.

Development • Supplementary information

A follow-up test for unequal medians (MATLAB's *multcompare* function) was used to further investigate differences between these distributions by comparing their mean ranks to their confidence intervals, testing the null hypothesis that all three populations were sampled from one with the same distribution. The returned P-values refute the null hypothesis and therefore show that all three groups have mean ranks that are significantly different from one another.

Population	Median Somite	Interquartile range	P-value
EpiSCs	5.5	7	<0.00001
Epi-NMP	11	6	<0.00001
Epi-meso	14	4	0.0056

The distributions of both the Epi-NMP and Epi-meso grafts have mean ranks that are significantly higher than for EpiSCs – *i.e.* they contribute to more posterior axial levels. Additionally, the distribution of the Epi-meso grafts has a mean rank that is significantly higher than for both the EpiSCs and Epi-NMP population. The samples can therefore be ordered from making more anterior to more posterior axial contributions after grafting as EpiSCs, Epi-NMPs, Epi-meso.

Development • Supplementary information

4. Table S1 - Primer sequences used for qRT-PCR

	Gene	Forward Primer Sequence	Reverse Primer Sequence
1	*Axin2*	CTAGACTACGGCCATCAGGAA	GCTGGCAGACAGGACATACA
2	*Bmp4*	CTCAAGGGAGTGGAGATTGG	ATGCTTGGGACTACGTTTGG
3	*Cdx2*	TCCTGCTGACTGCTTTCTGA	CCCTTCCTGATTTGTGGAGA
4	*Cyp26a1*	TCTGGGACCTGTACTGTGTGA	AAGCCGTATTTCCTGCGCTT
5	*Cdh1*	CAATGCCTGCTCTTGATGGT	GGGAGATCTGACTGCCTCTG
6	*Evx1*	TACAAACCTTCAGCGCCTCT	AAGGACCACTTCCTCCCAGT
7	*Fgf5*	TGGCATTATGTGGAATCTGG	CTGTGGACGCTGCACACTT
8	*Fgf8*	AGGACTGCGTATTCACAGAGAT	CATGTACCAGCCCTCGTACT
9	*Foxa2*	CATTACGCCTTCAACCACCC	GGTAGTGCATGACCTGTTCG
10	*Gata6*	GTTGCAGCAATCAGTGTTAAATC	GAAGTGGGCTGTGAGTGTAAG
11	*Hes5*	GGAACTGCAGGTGTTTCCTC	ATGCACCCACCCATACAAA
12	*Hoxc6*	CCCTCTCTTCTCCCTTGCTC	CCACGTCTGACTCCCTGTTT
13	*Meox1*	GCACAAGAGCTGATGGATGA	ACGCAGGATAGGTCCAAATG
14	*Mesp1*	CGCCTGCCTACCCTAGACC	CTGCTGAAGAGCGGAGATG
15	*Mixl1*	GGCAGCTTCCAGTTAACCAA	CTGAGTCCCAACCAGAAAGG
16	*Msx1*	TCTGCTGCCCTATACCACCT	GGCCTCTGCACCCTTAGTTT
17	*Msx2*	CGCCGCCCAGACATATGAG	CAGGTACTGTTTCTGGCGGA
18	*Cdh2*	GGGATGAGACCACAAGATAGGA	AAACTCCCTTTATCTGCAACCA
19	*Nkx1-2*	ACAACCACACAAGCCACTGA	CCATCCTGGGAACCCTTATT
20	*Nodal*	AGCCACTGTCCAGTTCTCCAG	GTGTCTGCCAAGCATACATCTC
21	*Notch1*	TCCAATGTGCATTGTGGACT	TGCAAGAATCTGCTGTGAGC
22	*Notch2*	CTGACTTATGCGATGGTGGG	ATGCAAGACTTCAAGTGGCC
23	*Oct4*	CCAATCAGCTTGGGCTAGAG	CTGGGAAAGGTGTCCCTGTA
24	*Osr1*	GAAGTCTAGTTCGCCAGGGG	TCAAATATCTTGTTCCGC
25	*Otx2*	CTGGGCTGAACATTCCAGTT	GTCCATTTCAGGTTGCTGGT
26	*Pax2*	CGTTGTGACCGGTCGTGATA	TGCTGAATCTCCAAGCCTCA
27	*Pax6*	AAGCACTTCACTTTGTAACTGTCC	CCAACTGATACCGTGCCTTC
28	*Pax7*	CAAGGTCTGGACAAGAGGAAAG	GAGCAAGGAATGTGGAGGAG
29	*Raldh2*	AAGACACGAGCCCATTGGAG	GGAAAGCCAGCCTCCTTGAT
30	*Snail1*	AGCCAGACTCTTGGTGCTTG	ACCCACTCGGATGTGAAGAG
31	*Sox1*	AGACAGCGTGCCTTTGATTT	TGGGATAAGACCTGGGTGAG
32	*Sox2*	CATGAGAGCAAGTACTGGCAAG	CCAACGATATCAACCTGCATGG
33	*Sprouty2*	ACAATTCAGCTAATGGAACCCG	TCTTCGCCTAGGAGTGTTGG
34	*Tbra*	CTGGGAGCTCAGTTCTTTCG	GTCCACGAGGCTATGAGGAG
35	*Tbx2*	GACCGACTTCATCGCTGTCA	TAGCTGCTTCCTTTTCTCCCG
36	*Tbx6*	CCAGAACCCTAGGATCACACA	CCCGAAGTTTCCTCTTCACA
37	*Wnt3a*	CATACAGGAGTGTGCCTGGA	AATCCAGTGGTGGGTGGATA

Table S1 - Primer sequences used for qRT-PCR.

Development • Supplementary information

5. Table S2 - Probes sequences with the conjugated dye used for sm-FISH

Brachyury (Quasar® 670/Quasar® 570)	Sox2(CAL Fluor Red 610)	Otx2 (Quasar® 570)	Cyp26a1 (Quasar® 670)
aaaaagggtcactctgcgga	Stellaris control probe sets: VSMF-3073-5	ggatggagtgagaccagata	ctgctcacacagtacaggtc
aacagccaccttcacttctc		ggcacaggcttttaaaggag	aaagaatgggaagcccatgg
tgacacatttaccttcagca		cacgttccagcactaactaa	aagcaccatctgcaatgttt
aagagtacatggcattgggg		gtgggtagatttggagtgac	tcatctgcagaaacttcctc
ttgtcagccgtcacgaagtc		ttgtttggaggcgcaaagtc	atgaagccgtatttcctgcg
tggatgtagacgcagctggg		ggcggttgctttagataaga	cccaaacagatgcgtcttgt
cccaaaattgggcgagtctg		catacccgaagtggtcagac	aagatgcgccgcacattatc
ctttgctgaaagacacaggc		ctagtaaatgtcgtcctctc	gaatcgtgcaggttggagag
attgagcttgttggtgagtt		aacagagcttccagaacgtc	ataatcaccttctttcgctg
ggttcatacttatgcaagga		catgaagatgtctgggtacc	tccagacaactgctgacttc
ccaactctcacgatgtgaat		cacttagctcttcgattctt	atggcgatgcggaacatgag
gaaagcagtggctggtgatc		ctggagagctcttcttcttg	tcgatgcgcgcgtgtataag
catcaaggaaggctttagca		acttgttccactctctgaac	gatcttggcgcgaatgttct
tccattacatctttgtggtc		ctagaggggggactgaactg	gagtgctcaatcaggagctg
caaactgagggtgggagctg		tggcaatggttgggactgag	ttgttttagtgcctgcatat
ttagagctgggtacctctca		cagatagacactggagcact	ccaaagaggagctctgttga
agctgttccgatgagcatag		ctgagtataggtcatgggat	actggctgtagtttcatgac
atgaattgtccgcataggtt		catagccttgactataacct	taagtgatcagtgacgtcgc
ctgcagcatggacagacaag		ccaaagtaggaagttgagcc	tggaggacatgtgggtagag
aggctagaccagttatcatg		ataagatccacagtccatgc	gctctttatctcttctcgaa
gagcctcgaaagaactgagc		taacagcattggtacccatg	ttgctcttgcaaagtaagcc
cgtcagtggtgtgtaatgtg		ctggggactgattgagatgg	ccatgtctaacttgttgtct
gaagacgaggacgtggcagc		atatccctgggtggaaagag	ccagtgtatttaagctgttc
cccttcatacatcggagaac		agtggttgagttaaaaccca	cagggtctccttaatgacac
actggctgtcagaaatgtct		ttggtccttataatccaagc	ctggaaccggaggattcaat
gaggctatgaggaggctttg		aagttaagcttccaagaggc	tcttcagagcaacccgaaac
attcacatagatgggggtga		ataatccaagcagtcagcat	tggtatccattcagctcaaa
ctcagcacatggaggaaagt		tggaatttccatgaggacgt	gaagatatctgccacatcgt
gttgacaccggttgttacaa		ttcagcccagcatatttaaa	ggattaaattcctccttgtt
ccactttgagcctagaagat		tttaaccaatgcctggctaa	atgaggcactataaagcggt
ggacatgtattattgctctt		agagcatcgttccatctaac	aagctgaaccgggaagcatc
ggtatcaaacaagaggctgt		ctcggtaactttgatcagga	aaggcctcctccaaatggaa
gtaggacacatatcactact		gttgatggacccttctaagg	ttgcaaactctttgcctaca
aaaaaactggaccacagccg		taaaacaccggatcacctct	tagaagctgccaatcacagt
actctgattaactgccagtc		ccattcctaagattcaacca	aatcttgcagggagattgtc
acaagatataggaccctacc		gaaacgtgaatgagcctggg	cagatatctccctggaagtg
ccaggatttcaaagtcacat		ttttcagtgccaactacctg	
taggacaggggtggacgaa		aatccacacagccctgaaaa	
tgtaccttgtgtctcactaa		catctaggacaatcagtcgc	
ggcaacaagggaggacatta		catattgactccgtatgagc	
tcctcaaagttaactactct		cagttttgaagtctagcaca	
cttctggctatgctttttc		ggagttcaaggttgcataca	
cagttgacggttcagttaca		ttaatcacagaagaacccct	
agctacgtgtgcttaaagta		cagttgctctgaattttgct	
atacagttgacttcccaaca		cccaaggtaatctttctaga	
		ggtgaattagggtccttttg	
		aaagtcatcagggtcagagc	
		tccagttaacatctgcaagc	

Table S2 - Probes sequences with the conjugated dye used for sm-FISH

Development • Supplementary information