© 2022. Published by The Company of Biologists Ltd | Development (2022) 149, dev198812. doi:10.1242/dev.198812

RESEARCH ARTICLE

Epha1 is a cell-surface marker for the neuromesodermal competent population

Luisa de Lemos*, André Dias*, Ana Nóvoa and Moisés Mallo‡

ABSTRACT

The vertebrate body is built during embryonic development by the sequential addition of new tissue as the embryo grows at its caudal end. During this process, progenitor cells within the neuromesodermal competent (NMC) region generate the postcranial neural tube and paraxial mesoderm. Here, we have applied a genetic strategy to recover the NMC cell population from mouse embryonic tissues and have searched their transcriptome for cell-surface markers that would give access to these cells without previous genetic modifications. We found that Epha1 expression is restricted to the axial progenitor-containing areas of the mouse embryo. Epha1-positive cells isolated from the mouse tailbud generate neural and mesodermal derivatives when cultured *in vitro*. This observation, together with their enrichment in the $Sox2^{+}/Tbxt^{+}$ molecular phenotype, indicates a direct association between Epha1 and the NMC population. Additional analyses suggest that tailbud cells expressing low Epha1 levels might also contain notochord progenitors, and that high Epha1 expression might be associated with progenitors entering paraxial mesoderm differentiation. Epha1 could thus be a valuable cell-surface marker for labeling and recovering physiologically active axial progenitors from embryonic tissues.

KEY WORDS: NM research, NMC population, Axial progenitors, Epha1, EMT

INTRODUCTION

Vertebrate body axis extension occurs in a head-to-tail sequence and relies on populations of cells with self-renewing properties, collectively known as axial progenitors (Aires et al., 2018; Dias and Aires, 2020; Wymeersch et al., 2021). These progenitors, initially located in the caudal epiblast and later in the tailbud, include several cell pools classified according to their potential. The neuromesodermal competent population (NMC population), previously referred to generally as neuromesodermal progenitors (NMPs) (see Binagui-Casas et al., 2021) is known to harbor cells with developmental potential to generate both post-occipital neural and mesodermal derivatives (i.e. NMC cells) (Binagui-Casas et al., 2021; Cambray and Wilson, 2002, 2007; Guillot et al., 2021; Henrique et al., 2015; Tzouanacou et al., 2009; Wymeersch et al., 2021). Recently, transcriptome data analyses helped to identify another population, the notochord progenitors, that seems to maintain a stable molecular signature during axial elongation (including genes such as *Noto*, *Shh* and *Foxa2*), possibly acting as a stable progenitor niche, particularly for the NMC population (Wymeersch et al., 2019). Since the initial identification of axial progenitors, increasing efforts have been conducted to determine the precise molecular characteristics of each different cell population, with particular focus on the NMC population, and the capacity of individual cells to actually generate both post-occipital neural and mesodermal derivatives [i.e. the 'true' neuro-mesodermal progenitors (NMPs; see Binagui-Casas et al., 2021)]. Combined mapping and expression studies indicate that the early neural marker Sox2 and the mesodermal transcription factor brachyury (also known as Tbxt or T) are mainly co-expressed in regions known to contain NMC cells and, more recently, also in NMPs (Binagui-Casas et al., 2021; Cambray and Wilson, 2007; Guillot et al., 2021; Martin and Kimelman, 2012; Olivera-Martinez et al., 2012; Tsakiridis et al., 2014; Wymeersch et al., 2016, 2021). The introduction of improved high-throughput techniques allowing *in vivo* and *in vitro* transcriptomic analyses, and in particular single-cell RNA sequencing (scRNA-seq), has shed additional light on the role of these key transcription factors and has provided a deeper understanding of other molecular players (e.g. *Wnt3a*, *Gdf11*, *Lin28* and *Cdx2*) and gene regulatory networks involved in the maintenance and differentiation of the NMC population, both in the embryo and *in vitro* (Aires et al., 2019; Dias et al., 2020; Edri et al., 2019; Gouti et al., 2017; Guibentif et al., 2021; Guillot et al., 2021; Koch et al., 2017; Wymeersch et al., 2019, 2021). Interestingly, some of these studies have shown that the transcriptome of the NMC population changes extensively over time, including the activation of an incomplete epithelial-to-mesenchymal transition (EMT) when entering the tailbud (Dias et al., 2020; Guillot et al., 2021; Wymeersch et al., 2019). Additionally, expression of *Tbx6* in a subset of cells within the tailbud NMC population has also been reported, and this expression is involved in neuro-mesodermal fate decisions (Javali et al., 2017). Gene expression and lineage-tracing experiments have also indicated that *Nkx1-2* (previously known as *Sax1*) is present in the NMC region, and in early neural and mesodermal progenitors throughout axial extension (Albors et al., 2018).

Despite all these studies, it is still not possible to isolate the NMC population in a physiologically active form without relying on previous modifications to introduce reporter genes, mostly because the molecules typically used to identify those progenitors are transcription factors. In the present work, we searched for cell-surface markers that could facilitate isolation of the NMC population using conventional cell-sorting approaches. For this, we exploited the intrinsic self-renewal properties of these progenitors to obtain a cell population highly enriched in NMC cells (and possibly NMPs) from the tailbud. In our approach, we introduced a fluorescent marker into the axial progenitors early in

Instituto Gulbenkian de Ciência, Rua da Quinta Grande 6, 2780-156 Oeiras, Portugal.
*These authors contributed equally to this work

‡Author for correspondence (mallo@igc.gulbenkian.pt)

L.d.L., 0000-0002-4666-5388; A.D., 0000-0003-3337-6373; A.N., 0000-0002-5668-5630; M.M., 0000-0002-9744-0912

Handling Editor: Patrick Tam
Received 19 November 2020; Accepted 2 February 2022

Development (2022) 149, dev198812. doi:10.1242/dev.198812

development, which permitted us to identify and isolate the long-term progenitors (i.e. cells from the NMC population that are present for long periods of axial elongation, potentially NMC/NMP cells; Cambray and Wilson, 2007) from the tailbud at later developmental stages. From the genes enriched in these cells, we focused on *Epha1* because it encodes a cell-surface protein that is mostly localized in the progenitor-containing areas of the mouse embryo throughout axial extension and because of the existence of fluorescence-activated cell sorting (FACS)-tested antibodies that could facilitate the development of protocols for the isolation of NMC cells from embryonic tissues. Using both whole-mount *in situ* hybridization and immunofluorescence, we detected stronger Epha1 expression in regions of the caudal epiblast and the tailbud known to contain NMCs/NMPs and their early mesoderm derivatives. Importantly, we show that Epha1-positive cells isolated from the mouse tailbud were able to produce neural and mesodermal derivatives when cultured *in vitro* under NMP differentiating conditions. In addition, Epha1-positive cells from the tailbud were highly enriched in $Sox2^{+}$/$Tbxt^{+}$ cells, the most frequently used molecular signature of the NMC population (Binagui-Casas et al., 2021). FACS profiles obtained from tailbud cells revealed that Epha1-positive cells can be divided in two subpopulations, $Epha1^{High}$ and $Epha1^{Low}$, according to their Epha1 protein content. Whereas cells with low Epha1 expression levels were also obtained from tissues containing early differentiated NMC-population derivatives, $Epha1^{High}$ cells were exclusively found in axial progenitor regions. Analysis of the transcriptomes obtained from tailbud $Epha1^{High}$ and $Epha1^{Low}$ cells revealed that, in addition to an enrichment in NMC population-related transcripts, the profile obtained from $Epha1^{High}$ cells suggested that progenitors were entering mesodermal routes, whereas $Epha1^{Low}$ cells included a molecular signature congruent with the existence of notochord progenitors within this cell compartment. Together, our results indicate that Epha1 is a valuable cell-surface marker for the isolation of the NMC population from the mouse tailbud and that high Epha1 values might be a molecular signature of progenitors entering the mesodermal progenitor compartment.

RESULTS

Labeling and isolation of long-term axial progenitors and their immediate descendants from developing mouse embryos

Most transcriptome analyses of the NMC population from embryonic tissue have been performed on microdissected regions of the embryo containing NMCs/NMPs that also include other cell types (e.g. some that had already entered mesoderm differentiation routes) or on cells isolated on the basis of the expression of a gene enriched in these progenitors (Dias et al., 2020; Gouti et al., 2017; Guillot et al., 2021; Koch et al., 2017; Wymeersch et al., 2019). Here, we used an unbiased approach to isolate NMPs exploiting their self-renewing properties. In particular, as the NMC population in the tailbud is derived mainly from the cells located in the node streak border (NSB) at earlier developmental stages (Cambray and Wilson, 2007), we designed a genetic strategy to label axial progenitors when they are part of the caudal epiblast and isolated them from the tailbud at later developmental stages. This system combines the *Cdx2P-Cre^{ERT}* transgene (Jurberg et al., 2013) with the *ROSA26-YFP-Cre* reporter (Srinivas et al., 2001). By administering a single low tamoxifen dose at embryonic day (E) 7.5, we induced a short pulse of permanent YFP label into a subset of axial progenitors that allowed their fate to be followed at later developmental stages. We have previously used this system with the *ROSA26-βgal-Cre* reporter (Soriano, 1999), which proved the efficiency of the method, as estimated by the presence of labeled cells in the embryonic tissues posterior to the position of Cre-mediated recombination, all the way down to the tail tip (Aires et al., 2019) (Fig. 1A-C).

Descendants of progenitors labeled with YFP at E7.5 were then recovered from the tail region at E10.5 by FACS. Two sets of YFP-positive cells were recovered: the first set was isolated from the tailbud ($Tail^{Prog}$), which is expected to be highly enriched in progenitors; a second set was recovered from more-anterior tail regions, where labeled cells are already part of the tissues derived from the progenitors ($Tail^{Desc}$). Comparison of the transcriptome of these two cell pools, obtained by RNA-sequencing (RNA-seq), identified 2465 genes showing differential expression ($P<0.05$) between the two cell groups (Table S1). Of these, 847 genes were highly expressed in the $Tail^{Prog}$, whereas 611 genes were upregulated in the $Tail^{Desc}$ sample. A selection of 12 differentially expressed genes, including NMC population-related genes and others encoding membrane proteins (*Epha1*, *Cldn9* and *Nkd2*), was then used to validate the RNA-seq data by reverse transcription-quantitative polymerase chain reaction (RT-qPCR) (Fig. 1E).

Initial analysis of the differentially expressed genes revealed high enrichment of the $Tail^{Prog}$ cells in factors that have been linked to neuro-mesodermal identity (Fig. 1D). For example, *Cdx2*, *Cdx4* and *Tbxt*, which are known to be highly expressed in the NMC population and proven to be essential for their activity (Amin et al., 2016; Chawengsaksophak et al., 2004; Herrmann et al., 1990; Savory et al., 2011; van Rooijen et al., 2012), were among the most strongly upregulated genes in the $Tail^{Prog}$ compartment (Fig. 1D). Similarly, other genes whose expression has been shown to be enriched in tailbud NMC region, including *Mnx1*, *Nkx1-2*, *Fgf8*, *Fgf4*, *Fgf3*, *Evx1*, *Cyp26a1*, *Hoxb1*, *Sp5*, *Gdf11*, *Wnt5a* or *Wnt3a* (Abu-Abed et al., 2001; Aires et al., 2019; Albors et al., 2018; Boulet and Capecchi, 2012; Cambray and Wilson, 2007; Dush and Martin, 1992; Greco et al., 1996; Guillot et al., 2021; Harrison et al., 2000; Javali et al., 2017; McPherron et al., 1999; Murphy and Hill, 1991; Naiche et al., 2011; Robinton et al., 2019; Sakai et al., 2001; Takada et al., 1994; Wymeersch et al., 2019; Yamaguchi et al., 1999), also showed significant differential expression in the $Tail^{Prog}$ cell pool (Fig. 1D). In addition, we also found other genes previously not linked to the NMC population, such as *Epha1*, *Arl4d*, *Efna1*, *Cldn9*, *Gad1* and *Scara5*, to be highly expressed in the progenitor compartment (Fig. 1D).

Conversely, we found enrichment of the $Tail^{Desc}$ cell pool in markers for neural and mesodermal derivatives of the NMC population, including *Ngn2*, *Sox1*, *Olig2*, *Olig3*, *Sox10*, *Meox2*, *Raldh2* (*Aldh1a2*), *Pax6*, *Fst* and *Nr2f2* (Albano et al., 1994; Aubert et al., 2003; Candia et al., 1992; Gradwohl et al., 1996; Jonk et al., 1994; Kuhlbrodt et al., 1998; Niederreither et al., 1997; Takeichi et al., 2002; Walther and Gruss, 1991) (Fig. 1D), indicating that the $Tail^{Prog}$ cells have the differentiation potential expected from the NMC population. *Sox2*, one of the components of the typical NMC-population signature, was expressed at slightly higher levels in $Tail^{Desc}$ than in $Tail^{Prog}$ cells (Fig. 1D). This is not surprising, as *Sox2* is highly expressed in the neural tube, which is one of the tissues from which the $Tail^{Desc}$ cells were recovered. Interestingly, both $Tail^{Prog}$ and $Tail^{Desc}$ datasets do not contain notochord markers such as *Shh*, *Noto* or *Foxa2*, which were readily found in a dataset obtained from similarly staged unsorted tailbuds (Aires et al., 2019) ($Tail^{Tot}$) (Fig. 1F; Table S2), indicating that the $Tail^{Prog}$ cell pool is specifically enriched in cells from the NMC population. Together, these data indicate that the genetic lineage-tracing strategy described

Development (2022) 149, dev198812. doi:10.1242/dev.198812

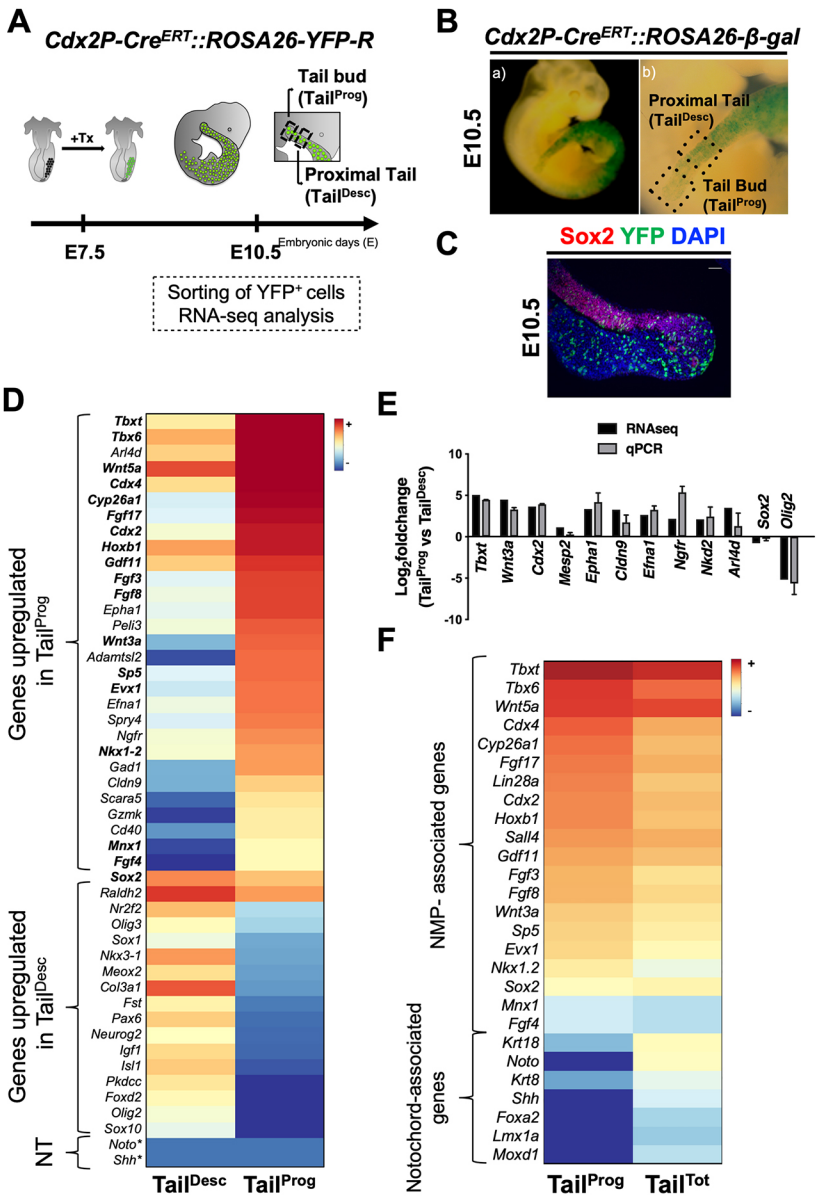

Fig. 1. Labeling, isolation and transcriptomic analysis of long-term axial progenitors and their immediate descendants. (A) Schematic representation of the genetic labeling strategy used to induce a permanent YFP label in axial progenitors in *Cdx2P-Cre^{ERT}::ROSA26-YFP* reporter embryos. Tx, tamoxifen. (Ba,b) β-Gal staining of E10.5 embryos labeled with the same genetic scheme but with a ROSA26 reporter to show contribution of labeled cells to the tailbud. (b) Higher magnification of the tail region with the regions from where Tail^{Prog} and Tail^{Desc} populations were obtained indicated by dotted lines. (C) Immunofluorescence for Sox2 (magenta) and YFP (green) in sagittal sections of *Cdx2P-Cre^{ERT}::ROSA26-YFP* tails. DAPI is shown in blue. Scale bar: 100 µm. (D) Heat map displaying several differentially expressed genes between Tail^{Prog} and Tail^{Desc} cell populations. Genes associated with NMPs are in bold. NT represents notochord genes. The key represents the average of normalized counts on a logarithmic scale. Values labeled with an asterisk are equal to zero; therefore, a logarithmic scale could not be applied. (E) Validation of RNA-seq data by RT-qPCR comparing the data obtained by RNA-seq (black bars) and by RT-qPCR (grey bars). The error bars represent the s.d. of three independent replicates. (F) Heat maps comparing transcriptome data from Tail^{Prog} YFP cells and from the entire tailbud (Tail^{Tot}) (same region) (obtained from Aires et al., 2019). Tail^{Prog} is highly enriched in NMPs but seem to be lacking in notochord progenitors. The key represents the average of normalized counts on a logarithmic scale.

here labels the tailbud NMC populations and their neural and mesodermal descendants, and that the isolated Tail^{Prog} cell pool is highly enriched in cells belonging to the NMC population.

Epha1 is expressed in axial progenitors and early mesodermal-fated cells

From the genes differentially expressed (P<0.05) between the Tail^{Prog} and the Tail^{Desc} cell populations, we concentrated on those encoding membrane proteins that could be used to isolate physiologically active cells of the NMC population without previous genomic modifications (e.g. transgenic reporters). Gene ontology (GO) categorization (Ashburner et al., 2000) of genes differentially upregulated in Tail^{Prog} cells with a $\log_2$ fold change>2 and q<0.05, identified 61 genes assigned to the category 'membrane' (GO:0005886). From these, we further selected the genes with higher expression, reducing the list to 16 genes (Fig. S1A). Expression analyses at E10.5 by *in situ* hybridization revealed that the staining patterns for some of those genes included a strong signal in the tail region, although these patterns differed among the various genes (Fig. S1B). In addition, for most of them, tailbud expression represented a subset of more-complex expression patterns that included other embryonic regions.

From these genes, we focused on *Epha1* based not only on its expression pattern, which is apparently restricted to the progenitor zone at different developmental stages (Fig. S1B) (Duffy et al., 2006), but also on the availability of FACS-validated antibodies, which are able to provide reliable data with cells obtained from solid embryonic tissues. By *in situ* hybridization, we detected *Epha1* expression as early as E8.5, when it was mostly observed in the caudal lateral epiblast (CLE), a region containing a NMC population (Binagui-Casas et al., 2021; Cambray and Wilson, 2007; Wymeersch et al., 2016), as well as in cells entering in the mesodermal compartment (Fig. 2Aa,a′). *Epha1* expression was also observed in the caudal epiblast of E9.5 embryos, fading anteriorly when entering the regions corresponding to the presomitic mesoderm and the caudal neural tube (Fig. 2Ab,b′); at E10.5 and E11.5, *Epha1* expression was essentially restricted to the tailbud (Fig. 2Ac-d′). Analysis of Epha1 by immunostaining confirmed the expression patterns obtained by *in situ* hybridization (Fig. 2B), although slightly broader, possibly derived from higher stability of

Development (2022) 149, dev198812. doi:10.1242/dev.198812

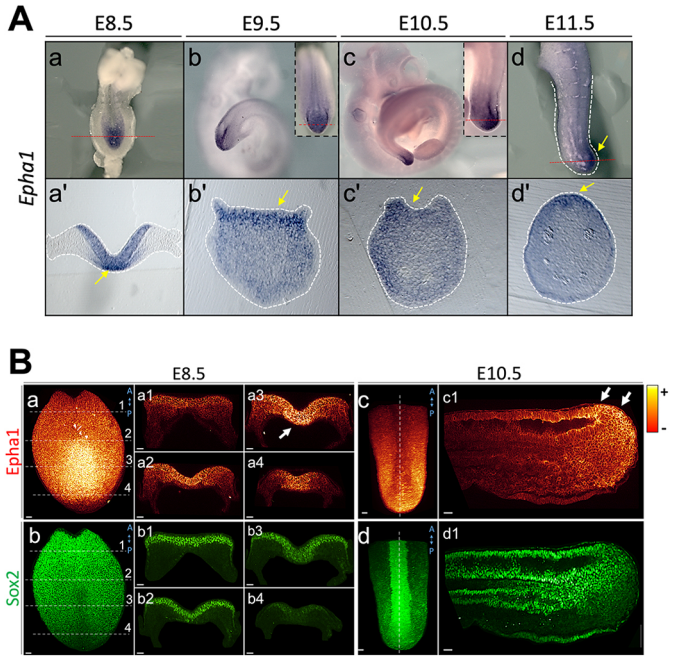

Fig. 2. Analysis of Epha1 expression during embryonic development. (A) Whole-mount *in situ* hybridization in E8.5 (a,a′), E9.5 (b,b′) E10.5 (c,c′) and E11.5 (d,d′) showing *Epha1* expression in the caudal epiblast and in the tailbud progenitor regions (arrow in d). Insets show higher magnifications of the posterior regions. (a′-d′) show transverse sections at the level indicated by the dotted red lines in the whole-mount embryos. Yellow arrows in a′,b′,c′,d′ indicate the regions where stronger *Epha1* expression is observed. (B) Immunofluorescence staining for Epha1 (a,c; red/yellow gradient) and Sox2 (b,d; green) of E8.5 (a,b) and E10.5 (c,d) embryos. (a1-a4, b1-b4,c1,d1) Transverse (a1-a4,b1-b4) and sagittal (c1,d1) sections through the areas indicated in a-d. High expression of Epha1 (white arrows) was found in NMC regions and in early mesoderm cells. Scale bars: 50µm.

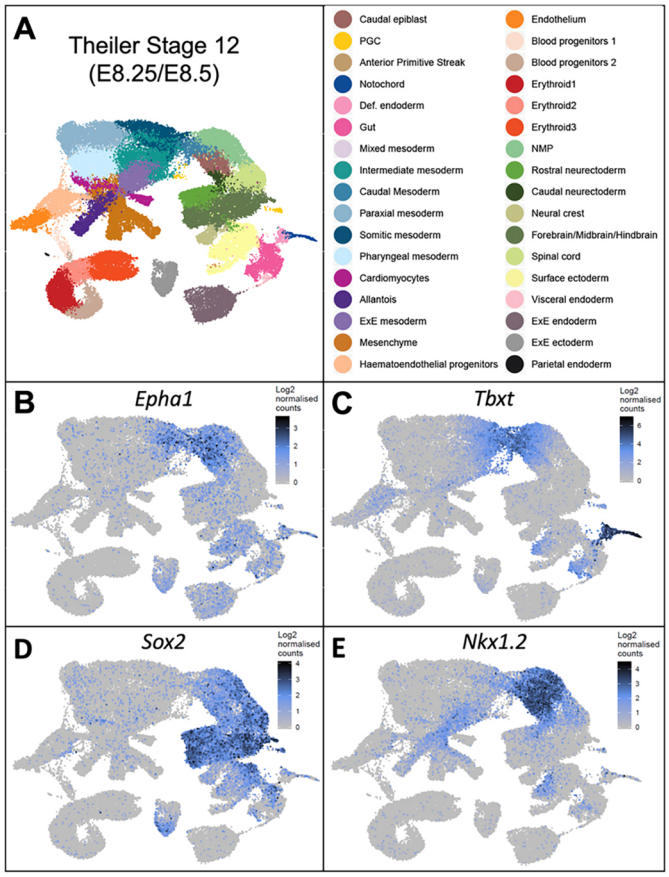

Fig. 3. *Epha1* expression in the scRNA-seq dataset of Theiler stage 12 (around E8.5) mouse embryos. (A) UMAP (uniform manifold approximation and projection) showing the general cluster cell distribution. (B-E) The distribution of *Epha1* expression (B) is shown together with that of *Tbxt* (C), *Sox2* (D) and *Nkx1-2* (E), which are known to include cells from NMC populations and their early mesoderm derivatives [clusters termed 'NMP population' (light green), 'caudal epiblast' (brown) and 'caudal mesoderm' (mid-blue)].

the protein than of the mRNA. In addition, although this technique is not purely quantitative, it showed that the Epha1 protein levels were not uniform in the positive domain. In particular, in E8.5 embryos, Epha1 expression was strong in the region of the primitive streak (PS) (Fig. 2Ba3, white arrow), which is known to contain progenitor cells undergoing an EMT to generate the mesodermal layer (Acloque et al., 2009; Hay, 1968; Wilson et al., 2009), and at E10.5 in the region abutting the posterior end of the neural tube, which is positive for both Tbxt and Sox2, as well as in mesenchyme caudal to this region thought to contain mesoderm progenitors (Dias et al., 2020; McGrew et al., 2008; Wymeersch et al., 2016, 2019) (Fig. 2Bc1, white arrows). Analysis of published mouse bulk RNA-seq datasets (Koch et al., 2017; Wymeersch et al., 2019) indicates that *Epha1* is not only strongly expressed in NMC regions [NSB, CLE and chordoneural hinge (CNH)] and their early mesoderm derivatives (e.g. posterior CNH), but is also highly enriched specifically in isolated cells with the $Tbxt^{+}/Sox2^{+}$ phenotype, the currently most-used criterion to identify the NMC population (Fig. S2A,B). At the single-cell level, using an early organogenesis stage (Theiler stage 12) mouse scRNA-seq dataset (Pijuan-Sala et al., 2019), we also found *Epha1* to be highly expressed in the clusters identified as the NMC population (previously referred to as NMPs) and the caudal mesoderm (Fig. 3), overlapping with the region co-expressing genes associated with the NMC population. Consistent with this observation, we also found that *Epha1* is present in several CLE single cells (Gouti et al., 2017) that co-express *Tbxt* and *Sox2* at E8.5 (Fig. S2C). Our expression data, together with the published transcriptomes, therefore suggest an association between Epha1 expression and the NMC population.

Epha1-expressing cells are enriched in Tbxt and Sox2 double-positive cells

To further characterize the Epha1-positive cell population, we analyzed cells from the tailbud and adjacent anterior tail region of E10.5 embryos by FACS, using an antibody against Epha1. Both areas contained a high proportion of Epha1-positive cells (Fig. 4A-C), which fits with the immunofluorescence data. Interestingly, we noticed that the staining patterns in the two embryonic regions were different, as cells from the tailbud included a population with higher staining intensity that was never observed in the FACS plots obtained from the anterior tail region (Fig. 4B). These results indicate the existence of cells with different amounts of Epha1 in the tailbud, consistent with the non-uniform staining intensities observed by immunofluorescence. Based on their Epha1 content, we grouped Epha1-positive tailbud cells into two subpopulations, which will be referred to as $Epha1^{High}$ and $Epha1^{Low}$ (Fig. S3). We also obtained similar $Epha1^{high}$ and $Epha1^{low}$ cell compartments in the progenitor-containing area of E8.5 embryos, although the proportion of $Epha1^{high}$ cells was lower in this tissue than in the tailbud from E10.5 embryos (Fig. S4).

Development (2022) 149, dev198812. doi:10.1242/dev.198812

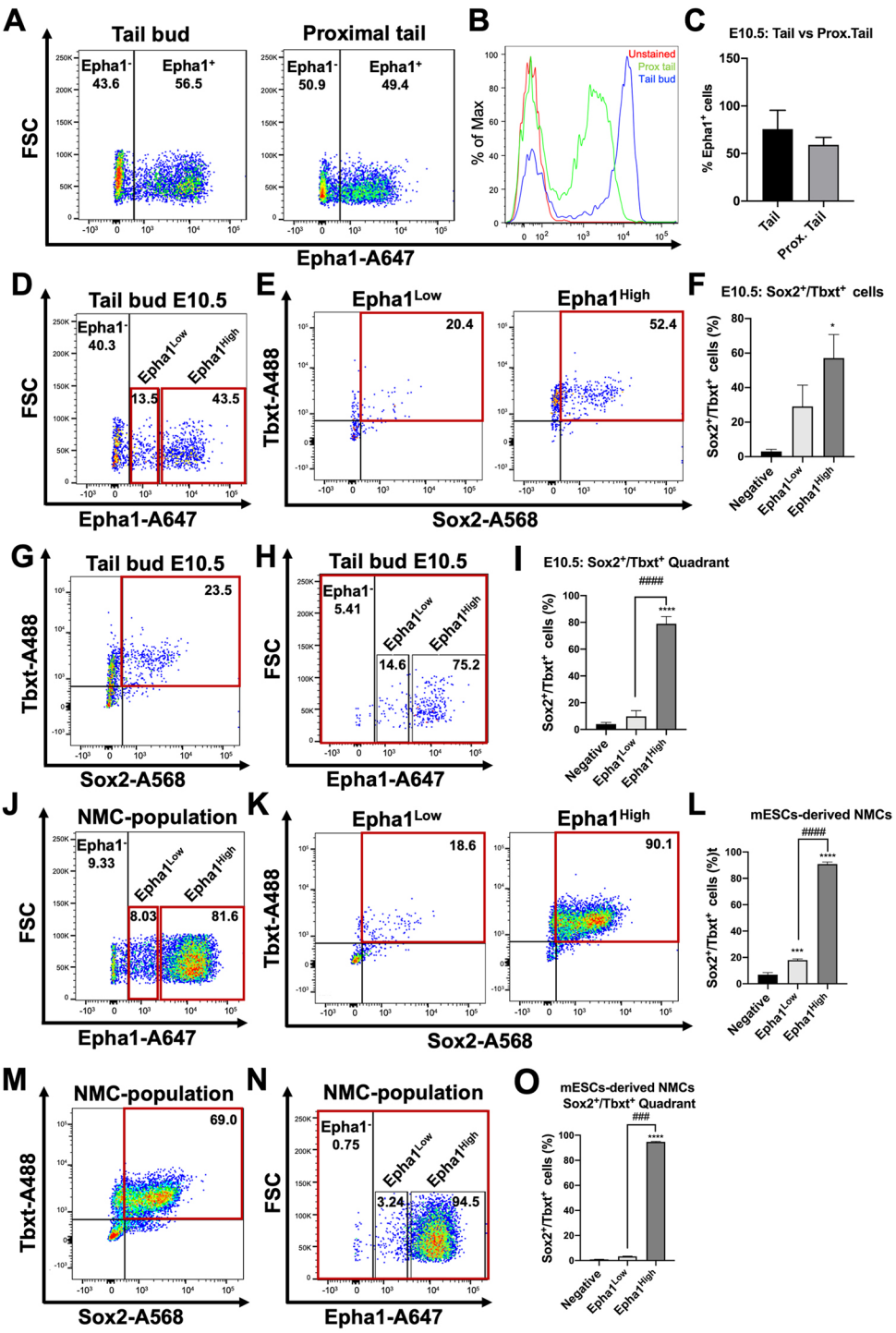

Fig. 4. Sox2 and Tbxt expression in Epha1 subpopulations from E10.5 tailbuds and *in vitro*-derived NMC populations. (A) FACS analysis showing fluorescence spread into the Epha1 channel of cells from the tailbud and proximal tail of E10.5 embryos. (B) Single parameter histogram showing the different fluorescent intensity of Epha1-positive cells from the tailbud (blue) and proximal tail (green) from E10.5 embryos, and the unstained control (red). (C) Percentage of Epha1-positive cells found in the tailbud and proximal tail of E10.5 embryos. (D) FACS dot-plot displaying Epha1 subpopulations in the tailbud of E10.5 embryos. (E) FACS profiles showing Sox2 and Tbxt expression in the Epha1^{Low} and Epha1^{High} compartments indicated in red in D. (F) Quantification of Sox2$^+$/Tbxt$^+$ cells within the different Epha1 subpopulations in the tailbud of E10.5 embryos (from Table 1). (G) FACS dot-plot displaying Sox2 and Tbxt expression in the tailbud of E10.5 embryos. (H) FACS profiles showing the distribution of Sox2$^+$/Tbxt$^+$ cells (indicated in red in G) among the various Epha1 populations. (I) Quantification of Sox2$^+$/Tbxt$^+$ cells in the different Epha1 subpopulations in the tailbud of E10.5 embryos (from Table 2). (J) FACS dot-plot displaying Epha1 subsets within *in vitro*-derived NMC cell populations. (K) FACS profiles showing Sox2 and Tbxt expression within the Epha1^{Low} and Epha1^{High} compartments, as indicated in red in J. (L) Quantification of double-positive Sox2/Tbxt cells within the different Epha1 compartments of *in vitro*-derived NMC cell populations (from Table 1). (M) FACS dot-plot displaying Sox2 and Tbxt expression in cells from *in vitro*-derived NMC populations. (N) FACS profiles showing the distribution of Sox2$^+$/Tbxt$^+$ cells (indicated in red in J) among the various Epha1 populations. (O) Quantification of Sox2$^+$/Tbxt$^+$ cells from *in vitro*-derived NMC populations in the different Epha1 compartments (from Table 2). Proportion of cells was calculated using at least three independent experiments and one-way analysis of variance ANOVA was used to determine statistical significance. (*P<0.05 and ****P<0.0001 versus negative; $^{\#\#\#}P$<0.001 and $^{\#\#\#\#}P$<0.0001 versus Epha1^{Low}). Error bars indicate the s.d. Gating strategy is shown in Fig. S4. mESCs, mouse embryonic stem cells.

We then assessed the distribution of Sox2$^+$/Tbxt$^+$ cells among the different Epha1 compartments obtained from the tailbud of E10.5 embryos. We observed that the proportion of cells co-expressing Sox2 and Tbxt differed significantly among the three cell compartments, being (on average) 57% for Epha1^{High} cells, 29% for Epha1^{Low} cells and almost absent in the Epha1-negative pool (3%), consistent with cells from the NMC population being present in the Epha1-positive compartment (Fig. 4D-F; Table 1). We then performed a complementary analysis by first isolating Sox2$^+$/Tbxt$^+$ cells from the E10.5 tailbud and analyzing their distribution among the different Epha1 compartments. We observed that, on average, 79% of the cells co-expressing Sox2 and Tbxt were Epha1^{High}, whereas only 9.8% and 4% of these cells were Epha1^{Low} or Epha1 negative, respectively (Fig. 4G-I; Table 2). These results are consistent with the Epha1^{High} subpopulation being enriched in the NMC cell population.

A similar distribution of Sox2$^+$/Tbxt$^+$ cells among the various Epha1 compartments and of Epha1^{High} and Epha1^{Low} from the Sox2$^+$/Tbxt$^+$ gated cells was also observed at E8.5 (Fig. S4). On average 53.4% of Epha1^{High} and 22.5% of Epha1^{Low} were found within the Sox2$^+$/Tbxt$^+$ cells (Fig. S4A-C; Table 1), and 70% and 18.4% of the Sox2$^+$/Tbxt$^+$ cells showed Epha1^{High} and Epha1^{Low}

Development (2022) 149, dev198812. doi:10.1242/dev.198812

Table 1. FACS data of Epha1 subpopulations

Sample	Epha1 channel	Distribution in Sox2 and Tbxt channels			
		Sox2$^-$/Tbxt$^-$	**Sox2$^-$/Tbxt$^+$**	**Sox2$^+$/Tbxt$^+$**	**Sox2$^+$/Tbxt$^-$**
E10.5 embryo	Epha1^{Neg}	58±2.55	33.60±1.41	3.07±1.26	5.34±0.15
	Epha1^{Low}	51.87±13.83	15.92±6.62	29.10±12.30	9.22±6.09
	Epha1^{High}	11.78±6.51	36.90±5.37	57.17±13.59	2.42±1.45
E8.5 embryo	Epha1^{Neg}	55.07±9.98	25.67±1.55	11.17±2.75	8.12±8.78
	Epha1^{Low}	46.40±13.48	21.87±12.35	22.50±4.87	9.22±6.09
	Epha1^{High}	15.98±11.81	25.47±13.07	53.37±10.20	5.17±7.49
mESCs-derived NMC population	Epha1^{Neg}	58.17±0.8	0.57±0.37	6.85±1.65	34.4±1.25
	Epha1^{Low}	59.83±3.52	4.05±2.09	17.97±0.93	18.13±2.28
	Epha1^{High}	1.61±0.6	5.72±0.57	90.83±1.63	1.86±0.67

Percentage of Epha1 cells (Epha1^{Neg}, Epha1^{Low} and Epha1^{High}) within Tbxt and Sox2 quadrants in wild-type embryos at E10.5 and E8.5, and in *in vitro*-derived NMC populations. Data are mean±s.d. from at least three independent experiments.

profiles, respectively (Fig. S4D-F; Table 2). In addition, NMC populations obtained from mouse ES cell cultures according to standard inducing conditions (Gouti et al., 2014; Turner et al., 2014) were preferentially found in the Epha1^{High} compartment, from which an average of 91% were positive for both Tbxt and Sox2 (Fig. 4J-L; Table 1), also further suggesting enrichment in cells from the NMC population in this Epha1-positive compartment of *in vitro* differentiated ES cells. Consistent with this, about 95% of Sox2$^+$/Tbxt$^+$ cells obtained from ES cell-derived NMC populations mapped to the Epha1^{High} compartment (Fig. 4M-O; Table 2). The proportion of Epha1-negative cells in the *in vitro*-derived NMC populations was almost residual, matching the high proportion of Sox2$^+$/Tbxt$^+$ cells in this compartment (Fig. 4N-O; Tables 1 and 2). Together, taking the simultaneous expression of Tbxt and Sox2 as the defining criterion for the NMC population, these results suggest that these cells might be mostly located within the Epha1^{High} compartment, although a significant number can also have an Epha1^{Low} phenotype, at least in embryonic tissues. This might reflect heterogeneity within the NMC population, when defined only as Sox2$^+$/Tbxt$^+$ cells, and is in line with observations made by other laboratories (e.g. Romanos et al., 2021; Wymeersch et al., 2016).

Epha1-positive cells differentiate into neural and mesodermal derivatives

To test whether the Epha1-positive compartment have the functional properties expected for the NMC population, we FACS to separate Epha1-positive and -negative cells from E10.5 embryonic tailbuds, and evaluated their capacity to differentiate *in vitro* along the neural and mesodermal routes (Fig. 5, Figs S5 and S6). When Epha1-positive cells were incubated under neural-promoting conditions, we observed sustained expression of Sox2 together with the absence of mesodermal markers (Fig. 5B and Fig. S6B,E). This result matches the molecular phenotype observed during neural differentiation of *in vitro*-derived NMC populations (Gouti et al., 2014, 2017). Interestingly, after 3.5 days under neural differentiation conditions, we observed the presence of Sox2-positive structures resembling neural rosettes, as well as the presence of neurite-like cellular extensions, suggesting the presence of neurons in these cultures. This was confirmed by the identification of Tuj1 expression in the periphery of the rosettes associated with the neurites (Fig. 5C). Conversely, when Epha1-positive cells were incubated under mesoderm-promoting conditions, they activated Tbx6 expression, indicating that they were able to enter paraxial mesodermal fates (Fig. 5A and Fig. S6A, E), which is also consistent with the patterns observed during mesoderm differentiation of *in vitro*-derived NMC populations (Gouti et al., 2014, 2017). It should be noted that clumps of Sox2-positive cells were often observed mixed with Tbx6-positive cells (Fig. 5A), although they both segregated from each other, even when present in the same cluster. A similar observation has been previously reported with *in vitro*-differentiated NMC populations (Gouti et al., 2017). So far, we have no explanation for the presence of those Sox2 clusters under mesodermal differentiation conditions. Importantly, Epha1-negative cells mostly failed to form colonies *in vitro* under the same culture conditions, remaining as single cells, and the few colonies formed by these cells were unable to follow neural or mesodermal fates when incubated under differentiation

Table 2. FACS data of Epha1 subpopulations within the different Tbxt/Sox2 quadrants from embryos at E10.5 and E8.5, and *in vitro*-derived NMC populations

Sample	Quadrant (Q)	Epha1 channel		
		Epha1^{Neg}	**Epha1^{Low}**	**Epha1^{High}**
E10.5 embryos	Q1: Tbxt$^+$/Sox2$^-$	40.30±1.75	8.75±2.79	46.70±2.51
	Q2: Tbxt$^+$/Sox2$^+$	4.04±1.25	9.79±4.30	78.97±5.44
	Q3: Tbxt$^-$/Sox2$^-$	67.03±5.33	21.47±3.74	6.33±2.45
	Q4: Tbxt$^-$/Sox2$^+$	48.73±8.90	32.27±2.29	10.10±2.59
E8.5 embryos	Q1: Tbxt$^+$/Sox2$^-$	20.47±2.36	29.67±3.67	46.00±6.24
	Q2: Tbxt$^+$/Sox2$^+$	7.76±0.50	18.40±6.06	70.30±7.47
	Q3: Tbxt$^-$/Sox2$^-$	42.20±18.35	39.17±7.31	12.85±10.97
	Q4: Tbxt$^-$/Sox2$^+$	36.93±20.69	42.17±2.36	20.27±17.56
mESCs-derived NMC population	Q1: Tbxt$^+$/Sox2$^-$	0.76±0.4	9.11±1.11	88.37±1.72
	Q2: Tbxt$^+$/Sox2$^+$	0.79±0.15	3.33±0.24	94.73±0.40
	Q3: Tbxt$^-$/Sox2$^-$	35.60±0.92	49.77±3.76	9.45±1.97
	Q4: Tbxt$^-$/Sox2$^+$	48.37±5.73	31.17±3.72	17.00±3.15

Data are mean±s.d.

Development (2022) 149, dev198812. doi:10.1242/dev.198812

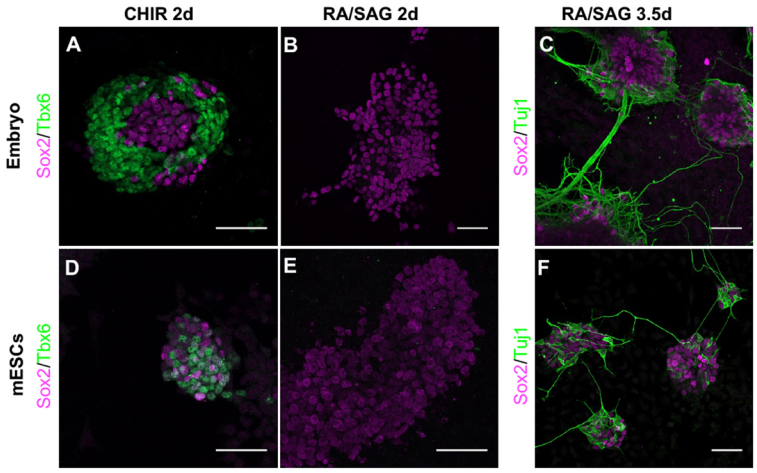

Fig. 5. Neural and mesodermal differentiation of Epha1^{+} sorted cells. (A,D) Immunofluorescence staining for Sox2 and Tbx6 in Epha1^{+} cells from E10.5 tailbuds (A) or from *in vitro*-derived NMC populations (D) differentiated with CHIR for 2 days after sorting. (B,C, E,F) Epha1^{+} cells from E10.5 tailbuds (B,C) or from *in vitro*-derived NMC populations (E,F) after differentiation with RA and SAG for 2 (B, E) and 3.5 (C,F) days after sorting, stained for Sox2 and Tbx6 (B,E), and Sox2 and Tuj1 (C,F). mESCs, mouse embryonic stem cells. Scale bars: 50 µm.

conditions (Fig. S6E). Epha1-positive sorted cells from *in vitro*-derived NMC populations, obtained from mouse embryonic stem (ES) cells, were also able to enter neural and paraxial mesodermal differentiation routes (Fig. 5D-F and Fig. S6C,D). In this case, however, it was not possible to compare with the Epha1-negative population from the same cultures, because they were present in only residual amounts (Fig. 4J and Fig. S5). Together, these results indicate a close functional association between Epha1 expression and the NMC population.

High Epha1 might label early mesodermal progenitors

To further characterize the Epha1-positive cell populations in the tailbud, we isolated the Epha1^{High} and Epha1^{Low} compartments from this region of E10.5 embryos, and analyzed their transcript content by RNA-seq. Both cell compartments showed enrichment in genes that have been associated with the NMC population (e.g. *Tbxt*, *Wnt3a*, *Tbx6* and *Nkx1-2*) to levels similar to those observed in Tail^{Prog} cells, an effect particularly clear in Epha1^{High} cells (Fig. 6A). However, Epha1^{High} and Epha1^{Low} cells seemed to represent two different cell compartments, as 875 genes showed differential expression (P<0.05) between these two Epha1 cell populations (Table S3). These included several early mesoderm-related genes, such as *Tbx6*, *Dll1*, *Msgn1*, *Cited1*, *Lfng*, *Snai1* or *Wnt3a* (Carver et al., 2001; Dias et al., 2020; Gouti et al., 2017; Koch et al., 2017; Serth et al., 2003; Wymeersch et al., 2019), that were upregulated in the Epha1^{High} compartment (Fig. 6B), suggesting that this cell population might contain cells from the NMC population that are already entering mesodermal fates. Interestingly, we also found that this cell population displays a more mesenchymal state than that observed in the axial progenitor pool (Tail^{Prog}) (Fig. 6C), thus suggesting that cells in the tailbud NMC population might complete the partial EMT that brought them into the tailbud (Dias et al., 2020) when entering mesodermal fates.

The transcriptome profile of Epha1^{Low} cells also showed substantial enrichment in some markers for the NMC population, which is consistent with the presence of Sox2^{+}/Tbxt^{+} cells in this cell compartment (Fig. 6A). Interestingly, these cells also contained a high abundance of transcripts for *Shh*, *Noto*, *Foxa2*, *Lmxa1*, *Krt8* and *Krt18*, which are not present in either the Epha1^{High} or the Tail^{Prog} cell populations (Figs 6B and 1D). These genes are commonly expressed in the notochord (Abdelkhalek et al., 2004; Ang et al., 1993; Echelard et al., 1993; Rodrigues-Pinto et al., 2016; Wymeersch et al., 2019), suggesting that the tailbud Epha1^{Low} cell population might also be enriched in notochord progenitors. In addition, this compartment also expressed markers for other midline structures, including those in common with the notochord (*Foxa2* and *Shh*), and others specific for the floor plate, such as *Slit2* and *Spon1* (Brose et al., 1999; Klar et al., 1992) (Fig. 6B), suggesting that the Epha1^{Low} cells might also contain progenitors for the ventral spinal cord.

Together, our results indicate that *Epha1* is a valuable cell-surface marker that can be used to label and isolate cells from the NMC population, and is able to generate neural and mesodermal derivatives. In addition, although higher Epha1 expression might be a hallmark of cells of the NMC population entering mesodermal routes, the midline progenitor population contributing to notochord and floor plate might be included within the low Epha1 expression compartment.

DISCUSSION

In this study, we have used a genetic strategy to label and isolate physiologically active cells from the NMC population of the tailbud of mouse embryos and analyzed their mRNA content using high-throughput methods. Although the Cre driver used to label the cells is not expected to be exclusive for the progenitors at early stages, it is expected that from the cells labeled by a pulse of Cre activity, only bona fide progenitors will carry the reporter to the tailbud. Indeed, labeled cells can be observed along the whole axis caudal to the region where the reporter was activated (see Fig. 1B or Aires et al., 2019). Both the enrichment of tailbud cells isolated with this strategy (Tail^{Prog}) in known markers for the NMC population, as well as the high concordance between our dataset and previously published data for that progenitor population obtained from embryos or *in vitro*-differentiated human or mouse ES cells (Dias et al., 2020; Gouti et al., 2017; Koch et al., 2017; Verrier et al., 2018; Wymeersch et al., 2019), support our conclusion that our lineage-tracing strategy is a non-biased reliable tailbud axial progenitor labeling method. However, it should be noted that the labeling conditions used to restrict the time frame of effective cell labeling might also have reduced the number of effectively labeled progenitors. Therefore, the tailbud cells isolated and analyzed using this approach might represent only a fraction of the actual progenitors and it is therefore possible that the tailbud NMC-population molecular fingerprint obtained in these experiments is not fully comprehensive.

Our initial goal was to identify cell-surface markers that could facilitate isolation of physiologically active cells from the NMC population. From the list of new genes that we found to be

Development (2022) 149, dev198812. doi:10.1242/dev.198812

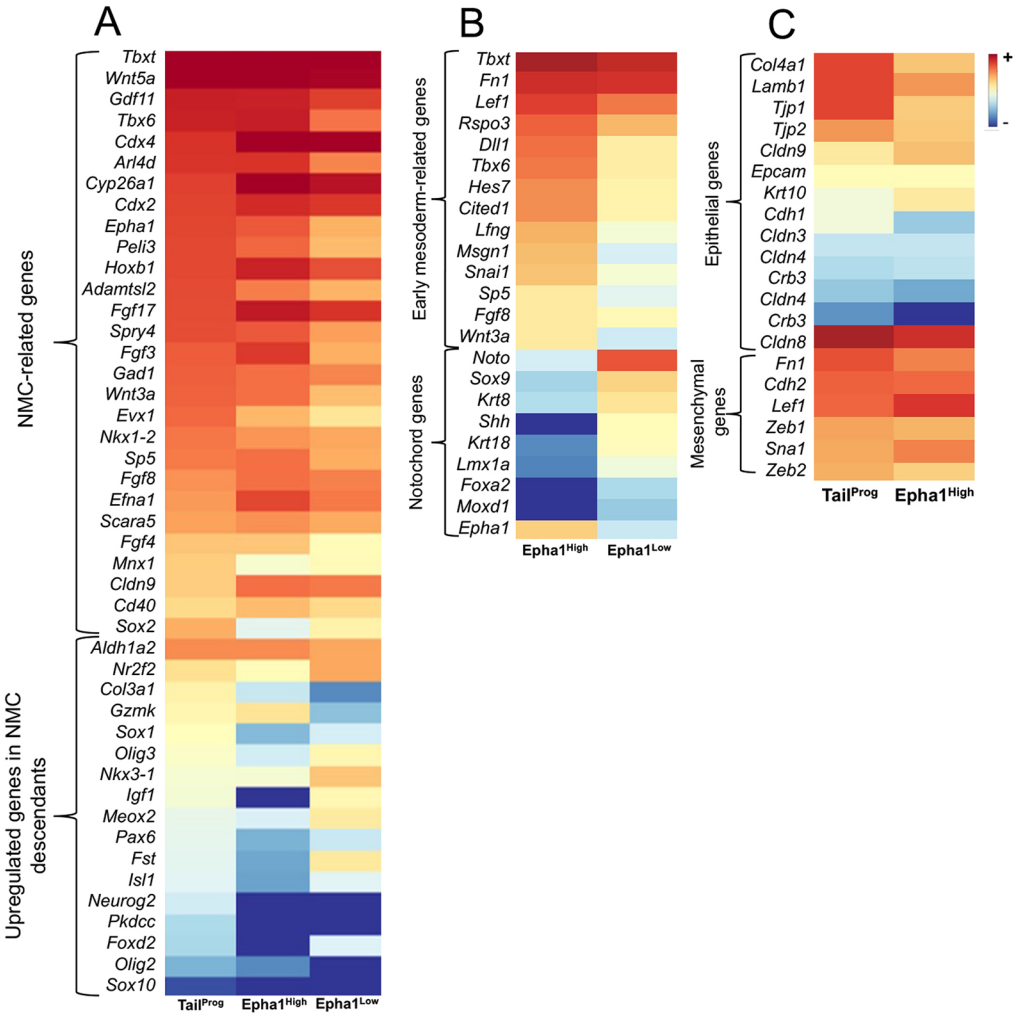

Fig. 6. Transcriptomic profile of two Epha1-positive subdomains within the axial progenitor cell population. (A) Heatmap showing expression of genes associated with NMC populations and their descendants in Epha1^{High}, Epha1^{Low} cells taking as a reference their levels in Tail^{Prog} cells. (B) Heatmap highlighting some differentially expressed genes between the two Epha1 cell populations. Epha1^{High} cells show high expression of several mesoderm-associated genes and the Epha1^{Low} compartment is enriched in notochord markers (e.g. *Noto*, *Shh* and *Foxa2*). (C) Heatmap comparing epithelial and mesenchymal markers between the Epha1^{High} and Tail^{Prog} cell populations. Epha1^{High} seem to have a more-complete mesenchymal phenotype, once several epithelial-associated genes are downregulated in the cell population. The key represents the average of normalized counts on a logarithmic scale.

differentially expressed between NMC populations and their derivatives, only Epha1 showed expression restricted to the progenitor-containing areas. Whether any of the other candidate surface molecules, alone or in combination, could help in developing protocols for NMP isolation will require additional work.

Functional and expression criteria indicate that the NMC population is contained within the Epha1-positive cell compartment of the tailbud. On the one hand, the Epha1-positive cells were enriched in Sox2^{+}/Tbxt^{+} cells, the standard signature for the NMC population. In addition, and most importantly, Epha1-expressing cells were able to generate both mesodermal and neural derivatives when cultured *in vitro* under the regular neuro-mesodermal differentiation conditions, whereas Epha1-negative cells obtained from the same tissue were mostly unable to grow and failed to enter any of those differentiation fates under the same conditions. From our data, it is not possible to determine whether all or only a subset of Epha1-expressing cells has NMC properties. This actually resembles the previous observation that only a subset of the Sox2^{+}/Tbxt^{+} cells, probably those expressing moderate levels of both proteins, might be bona fide NMC cells (Wymeersch et al., 2016).

Our experiments also indicate that the levels of Epha1 are not uniform in the caudal epiblast or the tailbud cells expressing this molecule. This was clearer from the FACS profiles showing the existence of at least two populations of Epha1-positive cells in these embryonic regions. The observation that the Epha1^{High} population was particularly enriched in Sox2^{+}/Tbxt^{+} cells suggested that it could represent the bona fide marker for the NMC population. However, the association between Epha1 expression levels and the NMC-population identity might be more complex, as a significant number of Sox2^{+}/Tbxt^{+} cells were also found in cells with lower Epha1 expression levels, indicating that the NMC population could be distributed between the Epha1^{High} and Epha1^{Low} compartments. Consistent with this, the transcriptome profile of tailbud Epha1^{Low} cells shows considerable enrichment in genes known to be expressed in the NMC population. The issue, then, is to understand the identity of cells containing high and low Epha1 expression levels. The transcriptome profile of the Epha1^{High} compartment, in addition to the presence of markers for the NMC population, also contains a significant enrichment in mesoderm-associated genes, thus indicating that at least part of these cells could represent progenitors entering mesodermal fates. Indeed, scRNA-seq analysis of Theiler stage 12 (around E8.5) (Pijuan-Sala et al., 2019) indicates that *Epha1* is strongly expressed in cells that were classified as belonging to NMC population and caudal mesodermal clusters, colocalizing with the *Sox2*$^{+}$*/Tbxt*$^{+}$*/Nkx1-2*$^{+}$ domain. The distribution of higher Epha1 immunoreactivity in the caudal tissue of E8.5 embryos and in the tailbud is also consistent with this hypothesis, as it is associated with areas representing early steps of mesodermal formation (e.g. primitive streak) (McGrew et al., 2008; Wymeersch et al., 2016, 2019). Similarly, the general patterns

Development (2022) 149, dev198812. doi:10.1242/dev.198812

observed within the $Sox2^{+}/Tbxt^{+}$ gate in the FACS plots from $Epha1^{High}$ cells might indicate the presence of cells at early stages in the progression towards the mesodermal lineage, as the cell distribution in the Sox2 expression axis includes a significant proportion of cells with relatively low Sox2 content but comparatively high Tbxt expression. This pattern is not observed in the $Sox2^{+}/Tbxt^{+}$ cells from the $Epha1^{Low}$ gate, further highlighting the differences between the $Epha1^{High}$ and $Epha1^{Low}$ compartments. Intriguingly, the transcriptomic profile of $Epha1^{Low}$ cells indicates that they might also contain precursors for the notochord and the floor plate, which are clearly not present in the $Epha1^{high}$ compartment. This finding is interesting in light of observations in chicken and zebrafish embryos that indicate the existence of a common progenitor for both structures (Row et al., 2016; Teillet et al., 1998). The presence of a similar progenitor in mice is not clear (Jeong and Epstein, 2003). Although our RNA-seq data cannot provide definitive proof for the existence of a bipotent progenitor contributing to both the notochord and floor plate, they might provide a new tool to investigate this interesting issue. Based on the above observations, we suggest that Epha1 is associated with axial progenitors, including the NMC population and those for the midline (notochord and floor plate), and that a transient increase in Epha1 expression is a hallmark of cells within the NMC population entering the mesodermal progenitor compartment that will generate paraxial mesoderm during axial extension. Interestingly, tailbud $Epha1^{High}$ cells seem to contain a more-complete mesenchymal profile than the tailbud NMC population, which has been shown to contain an intermediate epithelial/mesenchymal signature derived from an incomplete EMT (Dias et al., 2020; Guibentif et al., 2021; Guillot et al., 2021). This finding might indicate that, in the tailbud, mesodermal differentiation from the NMC population includes completion of the partial EMT that is characteristic of the axial progenitors. Such an intermediate EMT state has also been recently described for axial progenitors in zebrafish embryos, acting as a checkpoint to guarantee proper control of their differentiation (Goto et al., 2017; Kinney et al., 2020), indicating that epithelial-to-mesenchymal plasticity might be a general feature of these cells in different vertebrate clades (Binagui-Casas et al., 2021).

In conclusion, Epha1 and its expression levels might be a useful marker with which to access cells within the NMC-region/population in a physiologically active state without resorting to previous incorporation of genetic reporters. However, whether this molecule plays a specific role in NMCs and/or NMPs behavior remains to be determined. Epha1 is a member of the ephrin receptor family (Lisabeth et al., 2013). These receptors interact with ephrins on the surface of adjacent cells to control a variety of differentiation and morphogenetic processes involving tissue compartmentalization. It is therefore possible that increased Epha1 levels could be part of the mechanism promoting progenitor exit from their niche and their segregation into the mesenchymal compartment. Whether or not this is the case will need further investigation. Epha1 inactivation in mice resulted in a rather mild phenotype in axial structures: kinked tails (Duffy et al., 2008). However, a full appreciation of the role of Epha1 in NMP biology might be obscured by possible functional redundancy with other ephrin receptors, most particularly Epha2, as it binds the same ligands as Epha1 and its inactivation also affects development of axial structures (Naruse-Nakajima et al., 2001).

MATERIALS AND METHODS

Mice and embryos

Embryo staging was defined according to the standard timed mating approach, considering E0.5 the morning on which a mating plug was found. To isolate axial progenitors from developing embryos, matings were set up between *Cdx2P-Cre^{ERT}* transgenics (Jurberg et al., 2013) and *ROSA26-YFP-Cre* reporter (Srinivas et al., 2001) mice. Pregnant females were treated at E7.5 with a single intraperitoneal injection of 200 µl of 1 mg/ml tamoxifen (Sigma-Aldrich, T5648) in corn oil. Embryos were then collected at E10.5 by caesarean section, dissected in ice-cold phosphate-buffered saline (PBS) (Biowest, L0615) and processed for cell sorting (see below). Wild-type embryos for *in situ* hybridization or immunofluorescence were dissected in ice-cold PBS, fixed in 4% paraformaldehyde (PFA) (Sigma-Aldrich, P6148) in PBS and stored in methanol. Mice were genotyped as described by Aires et al. (2019), using primers listed in Table S4.

All animal procedures were performed in accordance with Portuguese (Portaria 1005/92) and European (directive 2010/63/EU) legislation and guidance on animal use in bioscience research. The project was reviewed and approved by the Ethics Committee of the Instituto Gulbenkian de Ciência and by the Portuguese National Entity Direcção Geral de Alimentação Veterinária (license 014308).

Expression analysis on embryos

Whole-mount *in situ* hybridization was performed according to the protocol described by Aires et al. (2019). Probes for *Efna1*, *Epha1*, *Ngfr*, *Cldn9*, *Nkd2* and *Arl4d* were prepared by amplifying cDNA fragments and cloning them into appropriate vectors for *in vitro* transcription. The sequences of all primers used to amplify these cDNAs are listed in Table S4. Whole-mount stained embryos were embedded in gelatin and sectioned using a vibratome following the protocol described by Dias et al. (2020). The β-galactosidase staining was performed as described by Aires et al. (2019).

Immunofluorescence analysis

Whole-mount immunostaining was carried out according to Dias et al. (2021). RapiClear 1.52 (SunJin Lab) was used for embryo clearing. Primary antibodies (used at 1:200 dilution) were as follows: rabbit anti-Sox2 (Abcam, AB92494) and goat anti-Epha1 antibody (R&D systems, AF3034). Secondary antibodies (all used at 1:1000 dilution) were as follows: donkey anti-goat 488 (Molecular Probes, A11055) and donkey anti-rabbit 568 (Molecular Probes, A10042). Images were acquired with a Prairie two-photon system, and the image dataset pre-processing was performed as described previously (Dias et al., 2021). No deconvolution or *z*-depth signal attenuation was performed in order to reduce possible pre-processing interferences that could misrepresent the real Epha1 expression levels in the tailbud.

Immunofluorescence staining of mouse sections and fixed sorted cells were performed as described previously (Aires et al., 2019). Primary antibodies (used at 1:200) were as follows: chicken anti-GFP (Abcam, AB13970), rabbit anti-Sox2 (Abcam, AB92494), mouse anti-Sox2 (Santa Cruz Biotechnology, sc-365823), rabbit anti-Tuj1 (Abcam, AB18207) and goat anti-Tbx6 (R&D, AF4744). Secondary antibodies (1:1000) were as follows: goat anti-chicken 488 (Thermo Fisher Scientific, A-11039), donkey anti-rabbit 568 (Molecular Probes, A10042) and donkey anti-goat 488 (Molecular Probes, A11055). Samples were analyzed using a Leica SP5 live or a Zeiss LSM 980 confocal microscope.

In vitro-derived NMC populations

CJ7 mouse ES cells (Swiatek and Gridley, 1993) were maintained in ES cell medium [DMEM High Glucose (Biowest, S17532L0102), 15% defined fetal bovine serum (Hyclone, GE Healthcare, SH30070.03), 1% MEM non-essential amino acid solution (Sigma-Aldrich, M-7145), 2 mM L-glutamine, 1% EmbryoMax Nucleosides (Millipore, ES-008-D), 100 U/ml penicillin and 100 µg/ml streptomycin (Sigma-Aldrich, P7539), 0.1 mM β-mercaptoethanol and 1000 U/ml LIF (Millipore, ESG1107)] on mitomycin C-inactivated primary mouse embryo fibroblasts. To start differentiation, ES cells were divided into NMC-like populations using a protocol adapted from Gouti et al. (2014). Briefly, ES cells were removed from feeders by dissociation using 0.05% trypsin-EDTA solution (Sigma-Aldrich, 59417C) and seeded at a density of 5000 cells/cm^2 on CellBIND Surface dishes (Corning, 3294) in N2B27 medium [Dulbecco's Modified Eagle Medium/F12 (Gibco, 21331-020) and Neurobasal medium (Gibco, 21103-049) (1:1), 40 µg/ml BSA, 0.1 mM β-mercaptoethanol and

Development (2022) 149, dev198812. doi:10.1242/dev.198812

supplemented with 1× N-2 (Gibco, LS17502048) and 1× B-27 minus vitamin A (Gibco, LS12587001)]. Cells were grown in N2B27 medium with 10 ng/ml bFgf (Peprotech, 100-18B) for 3 days (D1-D3). Neuromesodermal identity was induced by the addition of 5 µM CHIR99021 (Abcam, ab120890) from D2 to D3.

FACS of Epha1 cells for *in vitro* differentiation

Wild-type mouse embryos around E10/10.5 were dissected in ice-cold M2 medium (Sigma-Aldrich, M7167) and their tailbuds, up until the hindlimbs, dissociated into single cells through mechanical pipetting in cold Accutase solution (Sigma-Aldrich, A6964). NMC populations from differentiated mouse ES cells were also dissociated to single cells following a similar protocol. After adding two volumes of PBS/10% donkey serum (DS) (Biowest, S2170), the single-cell suspension was centrifuged at 120 ***g*** and washed twice with PBS/10% DS. After that, the single cells were incubated in blocking solution (10% DS with 1:100 dilution of 2.4G2 anti-mouse Fc block in PBS) on ice for 30 min and then stained (also on ice) for 30 min with a 1:100 dilution of goat anti-Epha1 antibody. After two washes with PBS/10% DS, cells were incubated with a 1:1500 dilution of donkey anti-goat A647 antibody (Thermo Fisher Scientific, A-21447) for an additional 30 min on ice and subsequently washed twice with PBS/10% DS. Finally, the single cells were resuspended in PBS, filtered through a 100 µm cell strainer and sorted on a FACSAria IIu (BD Biosciences). Epha1-positive and -negative cells were identified using a 633 nm excitation laser with filter detection of 660/20 following the gating conditions illustrated in Fig. S5 and then further processed similarly. Cell sorting was performed using a purity mask of 16 to avoid sort non-target particles in our samples and ensure high purity (>80-90%) of the sorted samples. Sorting experiments were performed independently three times.

In vitro differentiation of Epha1 sorted single cells

Epha1-positive and -negative FACS cells from embryonic tissue or from *in vitro*-generated NMC cells were collected in N2B27 medium supplemented with 10 ng/ml bFgf, 5 µM CHIR99021 and 10 µM Y-27632 inhibitor. To remove PBS added by the sorting machine, the single-cell suspension was centrifuged at 270 ***g*** and the culture medium replaced. Cells were plated in precoated 0.1% gelatin coverslips, at a density of around $3\times10^4/cm^2$ and cultured at 37°C, with 5% CO_2, for 3 to 24 h (the best results for mesoderm differentiation were achieved with only 3 h). After this period, for differentiation in mesoderm, cells were cultured with N2B27+5-8 µM of CHIR99021. For neural differentiation, we used N2B27 medium supplemented with 100 nM of retinoic acid (Sigma, R2625) and 500 nM of Shh agonist SAG (Calbiochem, 364590-63-6). After 2 or 3.5 days of differentiation, cells were washed twice with PBS, fixed in 4% PFA for 20 min and stored in PBS at 4°C for immunohistochemistry.

FACS of YFP and Epha1 cell populations for RNA-sequencing

Two regions were collected from E10.5 *Cdx2P-Cre^{ERT}::ROSA26-YFP* embryos: the tailbuds and a more-proximal region of the tail tip (up to the third somite). To obtain a single-cell suspension, tissue was incubated on ice for 15 min in Accutase. Digestion was terminated by adding two volumes of PBS/10% DS and washed twice with PBS/10% DS. Cells were then resuspended in PBS/2% DS and filtered through a 100 µm cell strainer. Cells were sorted according to their YFP-positive fluorescence in a MoFlo sorter (Beckman Coulter) using a 488 nm excitation laser with detector filter of 520/40. The YFP parameters were set using cells from *Cdx2P-Cre^{ERT}* tails dissected in parallel to serve as a YFP-negative control. YFP-positive cells were collected directly in TRI Reagent (Sigma-Aldrich, T9424) and stored at −80°C. The collected YFP-positive cells from the tail tip and the proximal tail were designated as Tail^{Prog} and Tail^{Desc}, respectively.

To isolate Epha1 cell populations for RNA-seq, single-cell suspensions were obtained from tailbuds and proximal tail regions of E10.5 wild-type embryos as mentioned above. Gating conditions of Epha1^{High} and Epha1^{Low} cell populations (illustrated in Fig. S3) were based on an apparent separation in total Epha1-positive cells histogram. For RNA-seq analyses, sorted cells were collected directly in TRI Reagent and kept at −80°C until further use. For all RNA-seq experiments, 15,000-20,000 purified-sorted cells per sample were collected to obtain a sufficient concentration of high-quality RNA.

RNA-sequencing analysis

Total RNA was isolated from the TRI Reagent suspension following the manufacturer's protocol, with the addition of 10 mg RNase-free glycogen (Roche, 10901393001) in the isopropanol step. RNA samples were then resuspended in RNase-free water. RNA concentration and purity were determined on an AATI Fragment Analyzer (Agilent). For the Tail^{Desc} and Tail^{Prog} samples, RNA-seq libraries were prepared from two biological replicates using TruSeq Stranded mRNA sample Prep Kit (Illumina, 20020594) and sequenced using Illumina HiSeq 2500 system at the CRG Genomics Unit (Barcelona, Spain). At least 25 million single end 50 bases reads were generated for each library. Read alignments were performed by TopHat2 v2.0.9 (Kim et al., 2013) with Bowtie2 v2.1.0.0. (Langmead and Salzberg, 2012). Differential expression analysis between Tail^{Desc}, Tail^{Prog} and Tail^{Tot} (Aires et al., 2019) was performed using CuffDiff v2.1.1 (Trapnell et al., 2013).

RNA-seq from Epha1^{High} and Epha1^{Low} cells was performed using two separate biological replicates. Libraries were prepared from total RNA using the SMART-Seq2 protocol (Picelli et al., 2014). Sequencing was performed on Illumina NextSeq500 at the IGC Genomics Facility, generating 20-25 million single-end 75 base reads per sample. Read alignments were performed as above. Read count normalization and differential expression between Epha1^{High} and Epha1^{Low} samples were analyzed using the DESeq2 R package (Love et al., 2014). To faithfully compare all samples (Tail^{Contr}, Tail^{Prog}, Epha1^{High} and Epha1^{Low}) from the above-mentioned RNA-seq independent experiments, normalization and differential expression were performed using the DESeq2 R package (Table S4). Representative heatmaps were created using BioVinci 2.0 data visualization software. The sequencing data of the RNA-seq experiments have been deposited in the NCBI trace and Short-read Archive (SRA) under accession numbers PRJNA527654 and PRJNA527619.

Quantitative RT-qPCR

Quantitative RT-qPCR was carried out as described previously (Aires et al., 2019). The sequences of the primers used are given in Table S5.

Protein expression profile analysis by FACS

Cells obtained from embryos and stained with the Epha1 antibody as described above were then washed twice with PBS/10% DS and processed for staining using True-Nuclear transcription factor Buffer Set (BioLegend, 424401) according to the manufacturer's instructions and following the protocol described previously (Aires et al., 2019). The following primary antibodies were used: rabbit anti-Sox2 (1:200; Abcam, AB92494) and mouse anti-Tbxt (1:100; Santa Cruz Biotechnology, sc-166962). Donkey anti-rabbit A568) (Molecular Probes, A10042; 1:1500) and donkey anti-mouse Alexa Fluor 488 (Abcam, ab150105; 1:1500) were used as secondary antibodies. For multicolor FACS analysis, single stain compensation controls were used to correct the spectral overlap between different fluorophores. Gating conditions of Epha1^{High} and Epha1^{Low} subsets were based on an apparent separation in a total Epha1-positive cells histogram (A647 channel). Quadrant gates were established according to fluorescence levels detected by the unstained controls and processed without primary antibodies; the negative control was determined by forelimbs cells exposed to the same conditions as the tail samples (Fig. S3). Flow cytometry data was analyzed using FlowJo 10 (BD Biosciences) software. Quadrant averages were calculated using at least three independent experiments and one-way analysis of variance ANOVA was used to determine statistical significance.

Single-cell analysis and visualization

The public single-cell molecular map of mouse gastrulation and early organogenesis (Pijuan-Sala et al., 2019) was used to assess the gene expression of *Epha1*, *Tbxt*, *Sox2* and *Nkx1-2* in Theiler stage 12 wild-type mouse embryos. E8.5 CLE single cells were obtained from Gouti et al. (2017), following the procedure described previously (Dias et al., 2020) and

using the single-cell consensus clustering (SC3) software (Kiselev et al., 2017). Gene expression visualization was obtained using SPRING (Weinreb et al., 2018), with the following parameters: minimum UMI total (for filtering cells)=1000; minimum number of cells with >=3 counts (for filtering genes) was set to 3; gene variability percentile (for filtering genes)=50; number of PCA dimensions (for building graph) was set to 20; number of nearest neighbors (for graph)=5.

Acknowledgements
We thank the IGC animal house, Flow Cytometry, Genomics and Imaging Facilities for their expert services, advice and assistance, and Daniel Sobral from the IGC Bioinformatics Unit for assistance with RNA-seq analysis; the SunJin laboratory for the RapiClear test sample; Val Wilson for sharing microarray processed data; and Anestis Tsakiridis for feedback on the culture and differentiation of FAC-sorted single cells. We also acknowledge all members of the Mallo laboratory for helpful discussions and comments throughout the course of the project, and Rita Aires and Ana Casaca for reading the manuscript.

Competing interests
The authors declare no competing or financial interests.

Author contributions
Conceptualization: L.d.L., M.M.; Methodology: L.d.L., A.D., A.N., M.M.; Validation: M.M.; Formal analysis: A.D.; Investigation: L.d.L., A.D., A.N.; Data curation: L.d.L., A.D.; Writing - original draft: L.d.L., A.D., M.M.; Supervision: M.M.; Project administration: M.M.; Funding acquisition: M.M.

Funding
This work was supported by the Fundacao para a Ciencia e a Tecnologia (PTDC/BEX-BID/0899/2014 and LISBOA-01-0145-FEDER-030254 to M.M.), by Santa Casa da Misericórdia de Lisboa (SCML-MC-60-2014 to M.M.), by a PhD fellowship from the Fundacao para a Ciencia e a Tecnologia (PD/BD/128426/2017 to A.D.) and by the research infrastructure Congento (LISBOA-01-0145-FEDER-022170).

Data availability
The sequencing data of the RNA-seq experiments have been deposited in the NCBI trace and Short-read Archive (SRA) under accession numbers PRJNA527654 and PRJNA527619.

References
Abdelkhalek, H. B., Beckers, A., Schuster-gossler, K., Pavlova, M. N., Burkhardt, H., Lickert, H., Rossant, J., Reinhardt, R., Schalkwyk, L. C., Herrmann, B. G. et al. (2004). The mouse homeobox gene Not is required for caudal notochord development and affected by the truncate mutation. *Genes Dev.* **18**, 1725-1736. doi:10.1101/gad.303504

Abu-Abed, S., Dollé, P., Metzger, D., Beckett, B., Chambon, P. and Petkovich, M. (2001). The retinoic acid-metabolizing enzyme, CYP26A1, is essential for normal hindbrain patterning, vertebral identity, and development of posterior structures. *Genes Dev.* **15**, 226-240. doi:10.1101/gad.855001

Acloque, H., Adams, M. S., Fishwick, K., Bronner-Fraser, M. and Nieto, M. A. (2009). Epithelial-mesenchymal transitions : the importance of changing cell state in development and disease. *J. Clin. Invest.* **119**, 1438-1449. doi:10.1172/JCI38019

Aires, R., Dias, A. and Mallo, M. (2018). Deconstructing the molecular mechanisms shaping the vertebrate body plan. *Curr. Opin. Cell Biol.* **55**, 81-86. doi:10.1016/j.ceb.2018.05.009

Aires, R., de Lemos, L., Nóvoa, A., Jurberg, A. D., Mascrez, B., Duboule, D. and Mallo, M. (2019). Tail Bud Progenitor Activity Relies on a Network Comprising Gdf11, Lin28, and Hox13 Genes. *Dev. Cell* **48**, 383-395. doi:10.1016/j.devcel.2018.12.004

Albano, R. M., Arkell, R., Beddington, R. S. P. 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. doi:10.1242/dev.120.4.803

Albors, A. R., Halley, P. A. and Storey, K. G. (2018). Lineage tracing of axial progenitors using Nkx1-2CreER T2 mice defines their trunk and tail contributions. *Development* **145**, dev164319. doi:10.1242/dev.164319

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. doi:10.1016/j.celrep.2016.11.069

Ang, S.-L., Wierda, A., Wong, D., Stevens, K. A., Cascio, S., Rossant, J. and Zaret, K. S. (1993). The formation and maintenance of the definitive endoderm lineage in the mouse: involvement of HNF3/forkhead proteins. *Development* **119**, 1301-1315. doi:10.1242/dev.119.4.1301

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T. et al. (2000). Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25-29. doi:10.1038/75556

Aubert, J., Stavridis, M. P., Tweedie, S., Reilly, M. O., Vierlinger, K., Li, M., Ghazal, P., Pratt, T., Mason, J. O., Roy, D. et al. (2003). Screening for mammalian neural genes via fluorescence-activated cell sorter purification of neural precursors from Sox1–gfp knock-in mice. *Proc. Natl. Acad. Sci. USA* **100**, 11836-11841. doi:10.1073/pnas.1734197100

Binagui-Casas, A., Dias, A., Guillot, C., Metzis, V. and Saunders, D. (2021). Building consensus in neuromesodermal research: current advances and future biomedical perspectives. *Curr. Opin. Cell Biol.* **73**, 133-140. doi:10.1016/j.ceb.2021.08.003

Boulet, A. M. and Capecchi, M. R. (2012). Signaling by FGF4 and FGF8 is required for axial elongation of the mouse embryo. *Dev. Biol.* **371**, 235-245. doi:10.1016/j.ydbio.2012.08.017

Brose, K., Bland, K. S., Kuan, H. W., Arnott, D., Henzel, W., Goodman, C. S., Tessier-Lavigne, M. and Kidd, T. (1999). Slit proteins bind robo receptors and have an evolutionarily conserved role in repulsive axon guidance. *Cell* **96**, 795-806. doi:10.1016/S0092-8674(00)80590-5

Cambray, N. and Wilson, V. (2002). Axial progenitors with extensive potency are localised to the mouse chordoneural hinge. *Development* **129**, 4855-4866. doi:10.1242/dev.129.20.4855

Cambray, N. and Wilson, V. (2007). Two distinct sources for a population of maturing axial progenitors. *Development* **134**, 2829-2840. doi:10.1242/dev.02877

Candia, A. F., Hu, J., Crosby, J., Lalley, P. A., Noden, D., Nadeau, J. H. and Wright, C. V. E. (1992). Mox-1 and Mox-2 define a novel homeobox gene subfamily and are differentially expressed during early mesodermal patterning in mouse embryos. *Development* **116**, 1123-1136. doi:10.1242/dev.116.4.1123

Carver, E. A., Jiang, R., Lan, Y., Oram, K. F. and Gridley, T. (2001). The mouse Snail gene encodes a key regulator of the epithelial-mesenchymal transition. *Mol. Cell. Biol.* **21**, 8184-8188. doi:10.1128/MCB.21.23.8184-8188.2001

Chawengsaksophak, K., de Graaff, W., Rossant, J., Deschamps, J. and Beck, F. (2004). Cdx2 is essential for axial elongation in mouse development. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 7641-7645. doi:10.1073/pnas.0401654101

Dias, A. and Aires, R. (2020). Axial Stem Cells and the Formation of the Vertebrate Body. In *Concepts and Applications of Stem Cell Biology* (ed. G. Rodrigues and B. A. J. Roelen), pp. 131-158. Cham: Springer.

Dias, A., Lozovska, A., Wymeersch, F. J., Nóvoa, A., Binagui-Casas, A., Sobral, D., Martins, G. G., Wilson, V. and Mallo, M. (2020). A TgfbRI/Snai1-dependent developmental module at the core of vertebrate axial elongation. *eLife* **9**, e56615. doi:10.7554/eLife.56615

Dias, A., Martins, G. G., Lopes, A. and Mallo, M. (2021). Three and four-dimensional visualization and analysis approaches to study vertebrate axial elongation and segmentation. *J. Vis. Exp.* e62086. doi:10.3791/62086

Duffy, S. L., Steiner, K. A., Tam, P. P. L. and Boyd, A. W. (2006). Expression analysis of the Epha1 receptor tyrosine kinase and its high-affinity ligands Efna1 and Efna3 during early mouse development. *Gene Expr. Patterns* **6**, 719-723. doi:10.1016/j.modgep.2005.12.007

Duffy, S. L., Coulthard, M. G., Spanevello, M. D., Herath, N. I., Yeadon, T. M., McCarron, J. K., Carter, J. C., Tonka, I. D., Kay, G. F., Phillips, G. E. et al. (2008). Generation and characterization of EphA1 receptor tyrosine kinase reporter knockout mice. *Genesis* **46**, 553-561. doi:10.1002/dvg.20434

Dush, M. K. and Martin, G. R. (1992). Analysis of mouse Evx genes: Evx-1 displays graded expression in the primitive streak. *Dev. Biol.* **151**, 273-287. doi:10.1016/0012-1606(92)90232-6

Echelard, Y., Epstein, D. J., St-Jacques, B., Shen, L., Mohler, J., McMahon, J. A. and McMahon, A. P. (1993). Sonic hedgehog, a member of a family of putative signaling molecules, is implicated in the regulation of CNS polarity. *Cell* **75**, 1417-1430. doi:10.1016/0092-8674(93)90627-3

Edri, S., Hayward, P., Jawaid, W. and Martinez Arias, A. (2019). Neuro-mesodermal progenitors (NMPs): a comparative study between pluripotent stem cells and embryo-derived populations. *Development* **146**, dev180190. doi:10.1242/dev.180190

Goto, H., Kimmey, S. C., Row, R. H., Matus, D. Q. and Martin, B. L. (2017). FGF and canonical Wnt signaling cooperate to induce paraxial mesoderm from tailbud neuromesodermal progenitors through regulation of a two-step EMT. *Development* **144**, 1412-1424.

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 Biol.* **12**, e1001937. doi:10.1371/journal.pbio.1001937

Gouti, M., Delile, J., Stamataki, D., Wymeersch, F. J., Huang, Y., Kleinjung, J., Wilson, V. and Briscoe, J. (2017). A gene regulatory network balances neural and mesoderm specification during vertebrate trunk development. *Dev. Cell* **41**, 1-19. doi:10.1016/j.devcel.2017.04.002

Gradwohl, G., Fode, C. and Guillemot, F. (1996). Restricted expression of a novel murine atonal-related bHLH protein in undifferentiated neural precursors. *Dev. Biol.* **180**, 227-241. doi:10.1006/dbio.1996.0297
Greco, T. L., Takada, S., Newhouse, M. M., McMahon, J. A., McMahon, A. P. and Camper, S. A. (1996). Analysis of the vestigial tail mutation demonstrates that Wnt-3a gene dosage regulates mouse axial development. *Genes Dev.* **10**, 313-324. doi:10.1101/gad.10.3.313
Guibentif, C., Griffiths, J. A., Imaz-Rosshandler, I., Ghazanfar, S., Nichols, J., Wilson, V., Göttgens, B. and Marioni, J. C. (2021). Diverse Routes toward Early Somites in the Mouse Embryo. *Dev. Cell* **56**, 141-153.e6. doi:10.1016/j.devcel.2020.11.013
Guillot, C., Djeffal, Y., Michaut, A., Rabe, B. and Pourquié, O. (2021). Dynamics of primitive streak regression controls the fate of neuromesodermal progenitors in the chicken embryo. *eLife* **10**, e64819. doi:10.7554/eLife.64819
Harrison, S. M., Houzelstein, D., Dunwoodie, S. L. and Beddington, R. S. P. (2000). Sp5, a new member of the Sp1 family, is dynamically expressed during development and genetically interacts with Brachyury. *Dev. Biol.* **227**, 358-372. doi:10.1006/dbio.2000.9878
Hay, B. (1968). Organization and fine structure of epithelium and mesenchyme in the developing chick embryo. In *Epithelial-Mesenchymal Interactions* (ed. R. Fleischmajer and R. E. Billingham), pp. 31-55. Philadelphia: Williams & Wilkins Co.
Henrique, D., Abranches, E., Verrier, L. and Storey, K. G. (2015). Neuromesodermal progenitors and the making of the spinal cord. *Development* **142**, 2864-2875. doi:10.1242/dev.119768
Herrmann, B. G., Labeit, S., Poustka, A., King, T. R. and Lehrach, H. (1990). Cloning of the T gene required in mesoderm formation in the mouse. *Nature* **343**, 617-622. doi:10.1038/343617a0
Javali, A., Misra, A., Leonavicius, K., Acharya, D., Vyas, B. and Sambasivan, R. (2017). Co-expression of Tbx6 and Sox2 identifies a novel transient neuromesoderm progenitor cell state. *Development* **144**, 4522-4529.
Jeong, Y. and Epstein, D. J. (2003). Distinct regulators of Shh transcription in the floor plate and notochord indicate separate origins for these tissues in the mouse node. *Development* **130**, 3891-3902. doi:10.1242/dev.00590
Jonk, L. J. C., de Jonge, M. E. J., Pals, C. E. G. M., Wissink, S., Vervaart, J. M. A., Schoorlemmer, J. and Kruijer, W. (1994). Cloning and expression during development of three murine members of the COUP family of nuclear orphan receptors. *Mech. Dev.* **47**, 81-97. doi:10.1016/0925-4773(94)90098-1
Jurberg, A. D., Aires, R., Varela-Lasheras, I., Nóvoa, A. and Mallo, M. (2013). Switching axial progenitors from producing trunk to tail tissues in vertebrate embryos. *Dev. Cell* **25**, 451-462. doi:10.1016/j.devcel.2013.05.009
Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R. and Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36. doi:10.1186/gb-2013-14-4-r36
Kinney, B. A., Al Anber, A., Row, R. H., Tseng, Y. J., Weidmann, M. D., Knaut, H. and Martin, B. L. (2020). Sox2 and canonical Wnt signaling interact to activate a developmental checkpoint coordinating morphogenesis with mesoderm fate acquisition. *Cell Rep.* **33**, 108311. doi:10.1016/j.celrep.2020.108311
Kiselev, V. Y., Kirschner, K., Schaub, M. T., Andrews, T., Yiu, A., Chandra, T., Natarajan, K. N., Reik, W., Barahona, M., Green, A. R. et al. (2017). SC3: consensus clustering of single-cell RNA-seq data. *Nat. Methods* **14**, 483-486. doi:10.1038/nmeth.4236
Klar, A., Baldassare, M. and Jessell, T. M. (1992). F-spondin: a gene expressed at high levels in the floor plate encodes a secreted protein that promotes neural cell adhesion and neurite extension. *Cell* **69**, 95-110. doi:10.1016/0092-8674(92)90121-R
Koch, F., Scholze, M., Wittler, L., Schifferl, D., Sudheer, S., Grote, P., Timmermann, B., Macura, K. and Herrmann, B. G. (2017). Antagonistic activities of Sox2 and brachyury control the fate choice of neuro-mesodermal progenitors. *Dev. Cell* **42**, 514-526. doi:10.1016/j.devcel.2017.07.021
Kuhlbrodt, K., Herbarth, B., Sock, E., Hermans-Borgmeyer, I. and Wegner, M. (1998). Sox10, a novel transcriptional modulator in glial cells. *J. Neurosci.* **18**, 237-250. doi:10.1523/JNEUROSCI.18-01-00237.1998
Langmead, B. and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357-359. doi:10.1038/nmeth.1923
Lisabeth, E. M., Falivelli, G. and Pasquale, E. B. (2013). Eph receptor signaling and ephrins. *Cold Spring Harb. Perspect. Biol.* **5**, a009159. doi:10.1101/cshperspect.a009159
Love, M. I., Huber, W. and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550. doi:10.1186/s13059-014-0550-8
Martin, B. L. and Kimelman, D. (2012). Canonical Wnt signaling dynamically controls multiple stem cell fate decisions during vertebrate body formation. *Dev. Cell* **22**, 223-232. doi:10.1016/j.devcel.2011.11.001
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. doi:10.1242/dev.022020
McPherron, A. C., Lawle, A. M. and Lee, S.-J. (1999). Regulation of anterior/posterior patterning of the axial skeleton by growth/differentiation factor 11. *Nat. Genet.* **22**, 260-264. doi:10.1038/10320
Murphy, P. and Hill, R. E. (1991). Expression of the mouse labial-like homeobox-containing genes, Hox 2.9 and Hox 1.6, during segmentation of the hindbrain. *Development* **111**, 61-74. doi:10.1242/dev.111.1.61
Naiche, L. A., Holder, N. and Lewandoski, M. (2011). FGF4 and FGF8 comprise the wavefront activity that controls somitogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 4018-4023. doi:10.1073/pnas.1007417108
Naruse-Nakajima, C., Asano, M. and Iwakura, Y. (2001). Involvement of EphA2 in the formation of the tail notochord via interaction with ephrinA1. *Mech. Dev.* **102**, 95-105. doi:10.1016/S0925-4773(01)00290-8
Niederreither, K., McCaffery, P., Dräger, U. C., Chambon, P. and Dollé, P. (1997). Restricted expression and retinoic acid-induced downregulation of the retinaldehyde dehydrogenase type 2 (RALDH-2) gene during mouse development. *Mech. Dev.* **62**, 67-78. doi:10.1016/S0925-4773(96)00653-3
Olivera-Martinez, I., Harada, H., Halley, P. A. and Storey, K. G. (2012). Loss of FGF-dependent mesoderm identity and rise of endogenous retinoid signalling determine cessation of body axis elongation. *PLoS Biol.* **10**, e1001415. doi:10.1371/journal.pbio.1001415
Picelli, S., Faridani, O. R., Björklund, Å. K., Winberg, G., Sagasser, S. and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171-181. doi:10.1038/nprot.2014.006
Pijuan-Sala, B., Griffiths, J. A., Guibentif, C., Hiscock, T. W., Jawaid, W., Calero-Nieto, F. J., Mulas, C., Ibarra-Soria, X., Tyser, R. C. V., Ho, D. L. L. et al. (2019). A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* **566**, 490-495. doi:10.1038/s41586-019-0933-9
Robinton, D. A., Chal, J., Lummertz da Rocha, E., Han, A., Yermalovich, A. V., Oginuma, M., Schlaeger, T. M., Sousa, P., Rodriguez, A., Urbach, A. et al. (2019). The Lin28/let-7 Pathway Regulates the Mammalian Caudal Body Axis Elongation Program. *Dev. Cell* **48**, 396-405. doi:10.1016/j.devcel.2018.12.016
Rodrigues-Pinto, R., Berry, A., Piper-Hanley, K., Hanley, N., Richardson, S. M. and Hoyland, J. A. (2016). Spatiotemporal analysis of putative notochordal cell markers reveals CD24 and keratins 8, 18, and 19 as notochord-specific markers during early human intervertebral disc development. *J. Orthop. Res.* **34**, 1327-1340. doi:10.1002/jor.23205
Romanos, M., Allio, G., Roussigné, M., Combres, L., Escalas, N., Soula, C., Médevielle, F., Steventon, B., Trescases, A. and Bénazéraf, B. (2021). Cell-to-cell heterogeneity in Sox2 and Bra expression guides progenitor motility and destiny. *eLife* **10**, e66588. doi:10.7554/eLife.66588
Row, R. H., Tsotras, S. R., Goto, H. and Martin, B. L. (2016). The zebrafish tailbud contains two independent populations of midline progenitor cells that maintain long-term germ layer plasticity and differentiate in response to local signaling cues. *Development* **143**, 244-254.
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. doi:10.1101/gad.851501
Savory, J. G. A., Mansfield, M., Rijli, F. M. and Lohnes, D. (2011). Cdx mediates neural tube closure through transcriptional regulation of the planar cell polarity gene *Ptk7*. *Development* **138**, 1361-1370. doi:10.1242/dev.056622
Serth, K., Schuster-Gossler, K., Cordes, R. and Gossler, A. (2003). Transcriptional oscillation of Lunatic fringe is essential for somitogenesis. *Genes Dev.* **17**, 912-925. doi:10.1101/gad.250603
Soriano, P. (1999). Generalized lacZ expression with the ROSA26 Cre reporter strain. *Nat. Genet.* **21**, 70-71. doi:10.1038/5007
Srinivas, S., Watanabe, T., Lin, C.-S., William, C. M., Tanabe, Y., Jessell, T. M. and Costantini, F. (2001). Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus. *BMC Dev. Biol.* **1**, 4. doi:10.1186/1471-213X-1-4
Swiatek, P. J. and Gridley, T. (1993). Perinatal lethality and defects in hindbrain development in mice homozygous for a targeted mutation of the zinc finger gene Krox20. *Genes Dev.* **7**, 2071-2084. doi:10.1101/gad.7.11.2071
Takada, S., Stark, K. L., Shea, M. J., Vassileva, G., McMahon, J. A. and McMahon, A. P. (1994). Wnt-3a regulates somite and tailbud formation in the mouse embryo. *Genes Dev.* **8**, 174-189. doi:10.1101/gad.8.2.174
Takeichi, M., Okubo, K., Uchida, T., Ohtsuki, T., Chisaka, O., Takebayashi, H., Ikenaka, K., Kawamoto, S. and Nabeshima, Y. (2002). Non-overlapping expression of Olig3 and Olig2 in the embryonic neural tube. *Mech. Dev.* **113**, 169-174. doi:10.1016/S0925-4773(02)00021-7
Teillet, M.-A., Lapointe, F. and Le Douarin, N. M. (1998). The relationships between notochord and floor plate in vertebrate development revisited. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 11733-11738. doi:10.1073/pnas.95.20.11733
Trapnell, C., Hendrickson, D. G., Sauvageau, M., Goff, L., Rinn, J. L. and Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* **31**, 46-53. doi:10.1038/nbt.2450
Tsakiridis, A., Huang, Y., Blin, G., Skylaki, S., Wymeersch, F., Osorno, R. R., Economou, C., Karagianni, E., Zhao, S., Lowell, S. et al. (2014). Distinct Wnt-driven primitive streak-like populations reflect in vivo lineage precursors. *Development* **141**, 1209-1221. doi:10.1242/dev.101014

Turner, D. A., Hayward, P. C., Baillie-Johnson, P., Rué, P., Broome, R., Faunes, F. and Martinez Arias, A. (2014). Wnt/β-catenin and FGF signalling direct the specification and maintenance of a neuromesodermal axial progenitor in ensembles of mouse embryonic stem cells. *Development* **141**, 4243-4353. doi:10.1242/dev.112979

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. doi:10.1016/j.devcel.2009.08.002

van Rooijen, C., Simmini, S., Bialecka, M., Neijts, R., van de Ven, C., Beck, F. and Deschamps, J. (2012). Evolutionarily conserved requirement of Cdx for post-occipital tissue emergence. *Development* **139**, 2576-2583. doi:10.1242/dev.079848

Verrier, L., Davidson, L., Gierliński, M., Dady, A. and Storey, K. G. (2018). Neural differentiation, selection and transcriptomic profiling of human neuromesodermal progenitor-like cells in vitro. *Development* **145**, dev166215. doi:10.1242/dev.166215

Walther, C. and Gruss, P. (1991). Pax-6, a murine paired box gene, is expressed in the developing CNS. *Development* **113**, 1435-1449. doi:10.1242/dev.113.4.1435

Weinreb, C., Wolock, S. and Klein, A. M. (2018). SPRING: A kinetic interface for visualizing high dimensional single-cell expression data. *Bioinformatics* **34**, 1246-1248. doi:10.1093/bioinformatics/btx792

Wilson, V., Olivera-Martinez, I. and Storey, K. G. (2009). Stem cells, signals and vertebrate body axis extension. *Development* **136**, 2133. doi:10.1242/dev.039172

Wymeersch, F. J., Huang, Y., Blin, G., Cambray, N., Wilkie, R., Wong, F. C. K. and Wilson, V. (2016). Position-dependent plasticity of distinct progenitor types in the primitive streak. *eLife* **5**, e10042. doi:10.7554/eLife.10042

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**, dev168161.

Wymeersch, F. J., Wilson, V. and Tsakiridis, A. (2021). Understanding axial progenitor biology in vivo and in vitro. *Development* **148**, dev180612. doi:10.1242/dev.180612

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. doi:10.1242/dev.126.6.1211

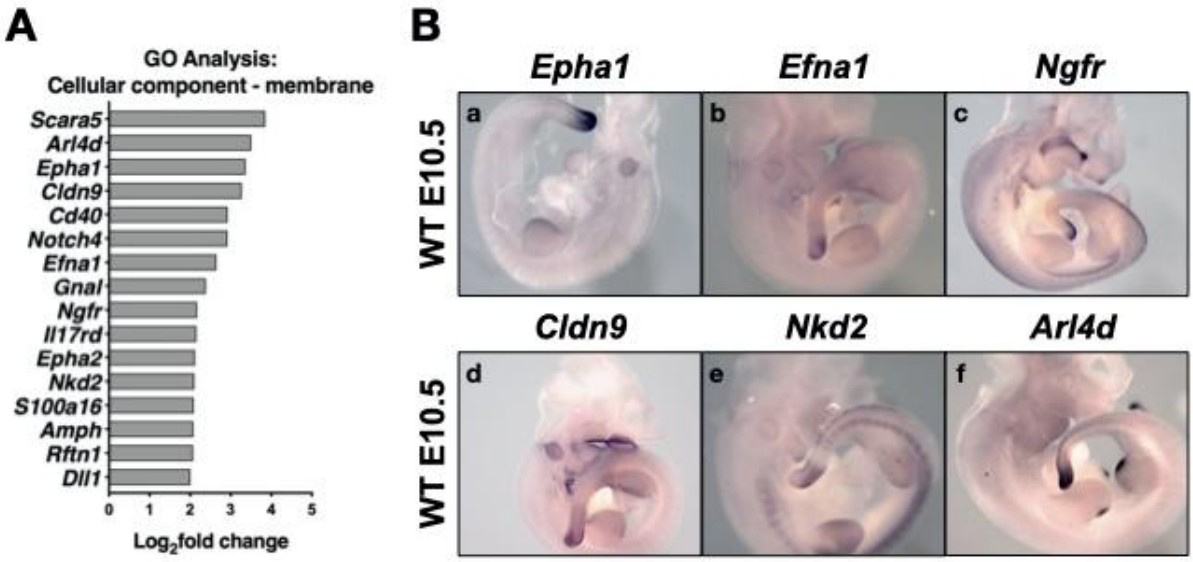

Fig. S1. Identification of potential cell-surface markers for axial progenitors. A. Differentially expressed genes encoding for plasma membrane proteins according to gene ontology classification. **B.** Whole mount ISH of E10.5 embryos using probes for *Epha1* (a), *Efna1* (b), *Ngfr* (c), *Cldn9* (d), *Nkd2* (e) and *Arl4d* (f).

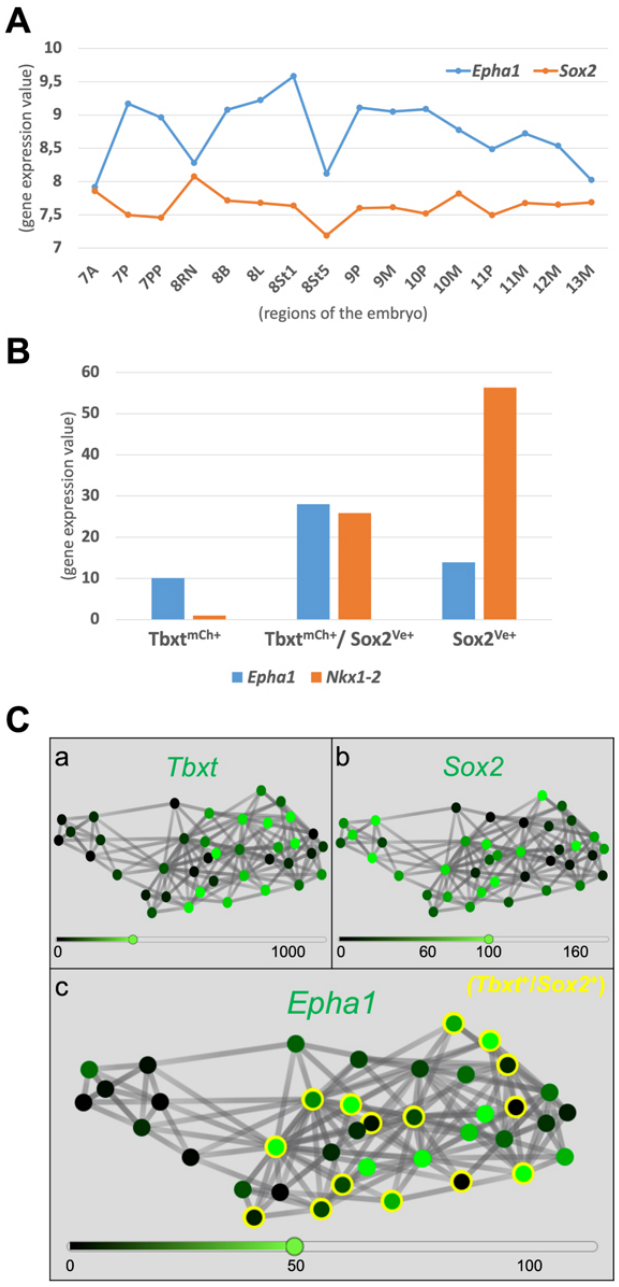

Fig. S2. *Epha1* expression in cells of NMC-populations. **A**. Microarray data from (Wymeersch et al., 2019) indicates that *Epha1* is expressed at the posterior region of the mouse embryo during the emergence of NMC cells ("7P" - E7.5 posterior epiblast) and by E8.5 is highly expressed in NMC-regions ("8B" = NSB and "8L" = CLE). During tailbud stages, *Epha1* is highly expressed in the CNH ("9M = E9.5 CNH; "10M" = E10.5 CNH; "11M" = E11.5 CNH and "12M" = E12.5 CNH) and its expression decays with the end of axis elongation ("13M" = E13.5 CNH). Contrasting with regions that contain lateral and paraxial mesoderm progenitors (LPMPs) (Wymeersch et al., 2016) ("7PP" = posterior-most part of the epiblast at E7.5 and "8St5" = posterior-most part of the primitive streak at E8.5), the regions with early mesoderm progenitors are also high *Epha1* positive ("8St1" = anterior part of the primitive streak at E8.5; "9P", "10P" and "11P" = region posterior to the CNH at E9.5, E10.5 and E11.5, respectively).

Epha1 is not expressed in the anterior part of the embryo at E7.5 ("7A") and also in the rostral node region at E8.5 ("8RN"), where notochord progenitors are present. **B**. Comparison between *Epha1* and *Nkx1-2* expression in sorted Tbxt$^+$, Sox2$^+$ and Tbxt$^+$/Sox2$^+$ cells at E8.5 (from Koch et al., 2017). *Epha1* is among the 154 genes that are highly expressed in Tbxt$^+$/Sox2$^+$ cells (group1 in Koch et al., 2017) and its expression in Tbxt-mCh$^+$/ Sox2-Ve$^+$ cells is slightly higher than *Nkx1-2*. **C**. *Epha1* is expressed in *Tbxt$^+$/Sox2$^+$* single-cells from E8.5 CLE (from Gouti et al., 2017). *Tbxt* (C**a**) and *Sox2* (**Cb**) expression in microdissected single cells from the E8.5 CLE. *Epha1* is expressed in the majority (2/3) of *Tbxt* and *Sox2* double positive cells (circled in yellow) (**Cc**).

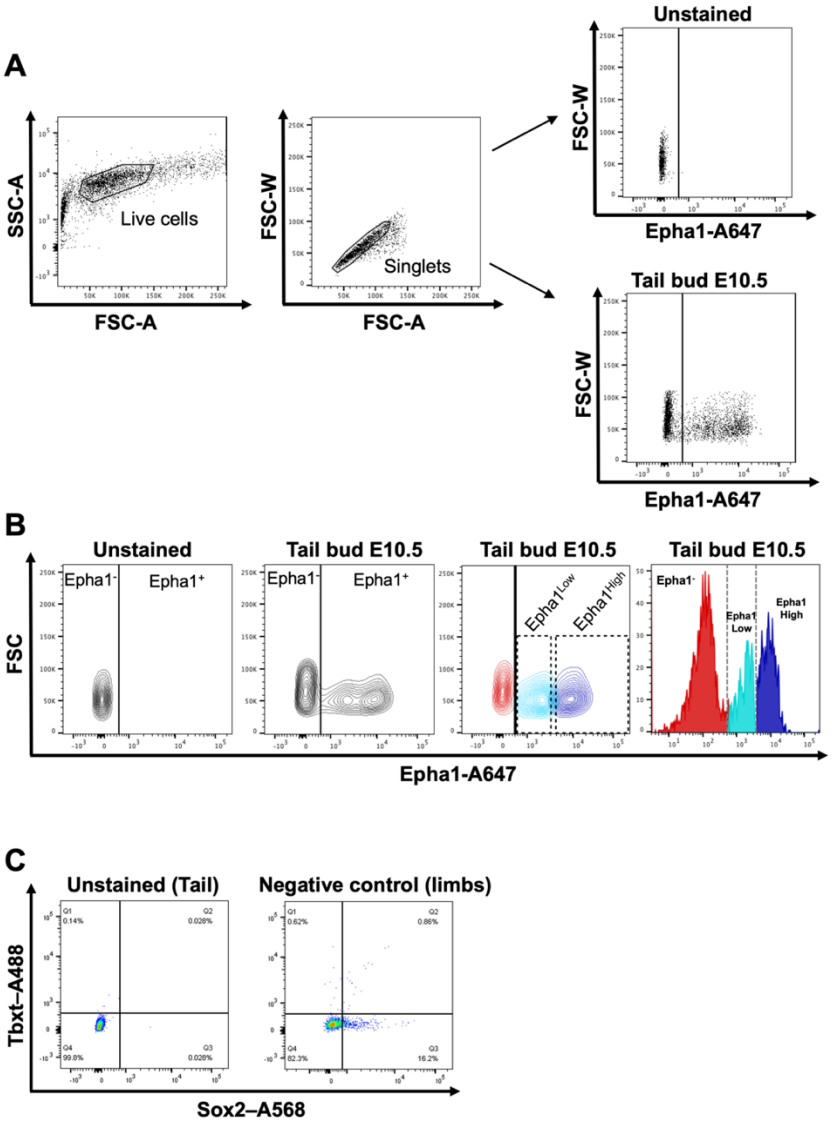

Fig. S3. Gating strategy for FACS analysis used to define Epha1 positive cells and Sox2/Tbxt quadrants. **A.** Representative example of exclusion of debris and doublets cells (using forward scatter and side scatter gates) and definition of Epha1 gating. **B.** Representative gating scheme illustrating the two subpopulations for Epha1-positive cells: Epha1^{Low} (light blue) and Epha1^{High} (dark blue), and a representative histogram showing two peaks with different fluorescent intensity in the Epha1-A647 channel, corresponding to the defined Epha1^{Low} and Epha1^{High} populations. All gates were firstly set using unstained control. **C.** Gating strategy for defining intracellular staining of Sox2-A568 and Tbxt-A488 on unstained cells from tail and stained cells from forelimbs (negative control) of E10.5 mouse embryos.

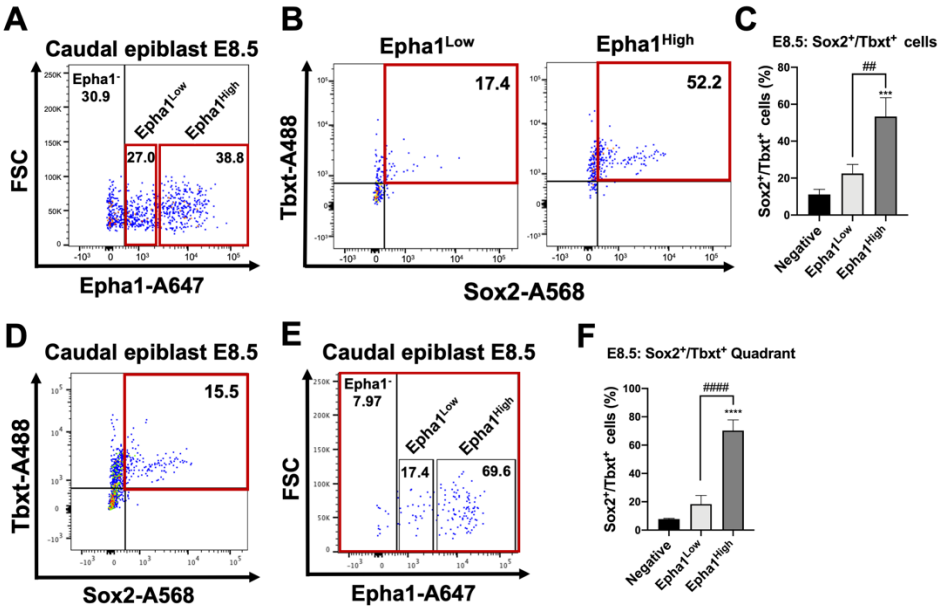

Fig. S4. Flow cytometry analysis of Epha1 subpopulations that co-express Sox2 and Tbxt in E8.5 embryos. A. FACS dot-plot displaying Epha1 subpopulations of E8.5 posterior ends. **B**. FACS profiles of Sox2 and Tbxt expression in cells from the Epha1^{Low} and Epha1^{High} compartments indicated in A. **C.** Quantification of Sox2^{+}/Tbxt^{+} cells within the different Epha1 subpopulations from posterior ends of E8.5 (from Table 1). **D.** FACS dot-plot displaying Sox2 and Tbxt expression in cells from the posterior regions of E8.5 embryos. **E.** FACS profile showing distribution of Sox2^{+}/Tbxt^{+} cells from the red window in D among the Epha1 compartments. **F.** Quantification of the distribution of Sox2^{+}/Tbxt^{+} cells from the posterior region of E8.5 embryos between the different Epha1 subpopulations (from Table 2).

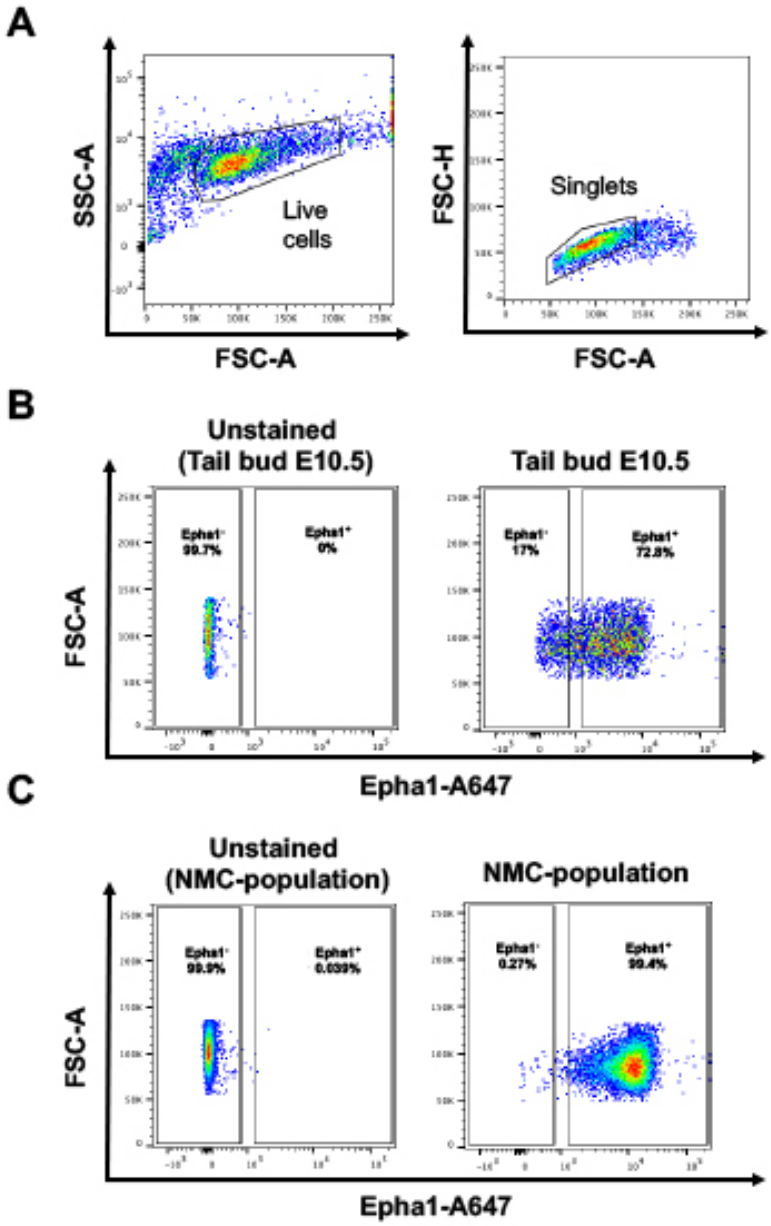

Fig. S5. Gating strategy for FACS-sorting of Epha1^{+} and Epha1^{-} cells from E10.5 tail buds and *in vitro* derived NMC-populations for in vitro culture experiments. A. Cells were first gated for singlets discrimination using side and forward scatter detectors. **B**. Representative sorting gates for Epha1 positive and negative populations from E10.5 Tail buds. **C**. Representative sorting gates for Epha1 positive and negative populations from in vitro derived NMC-populations.

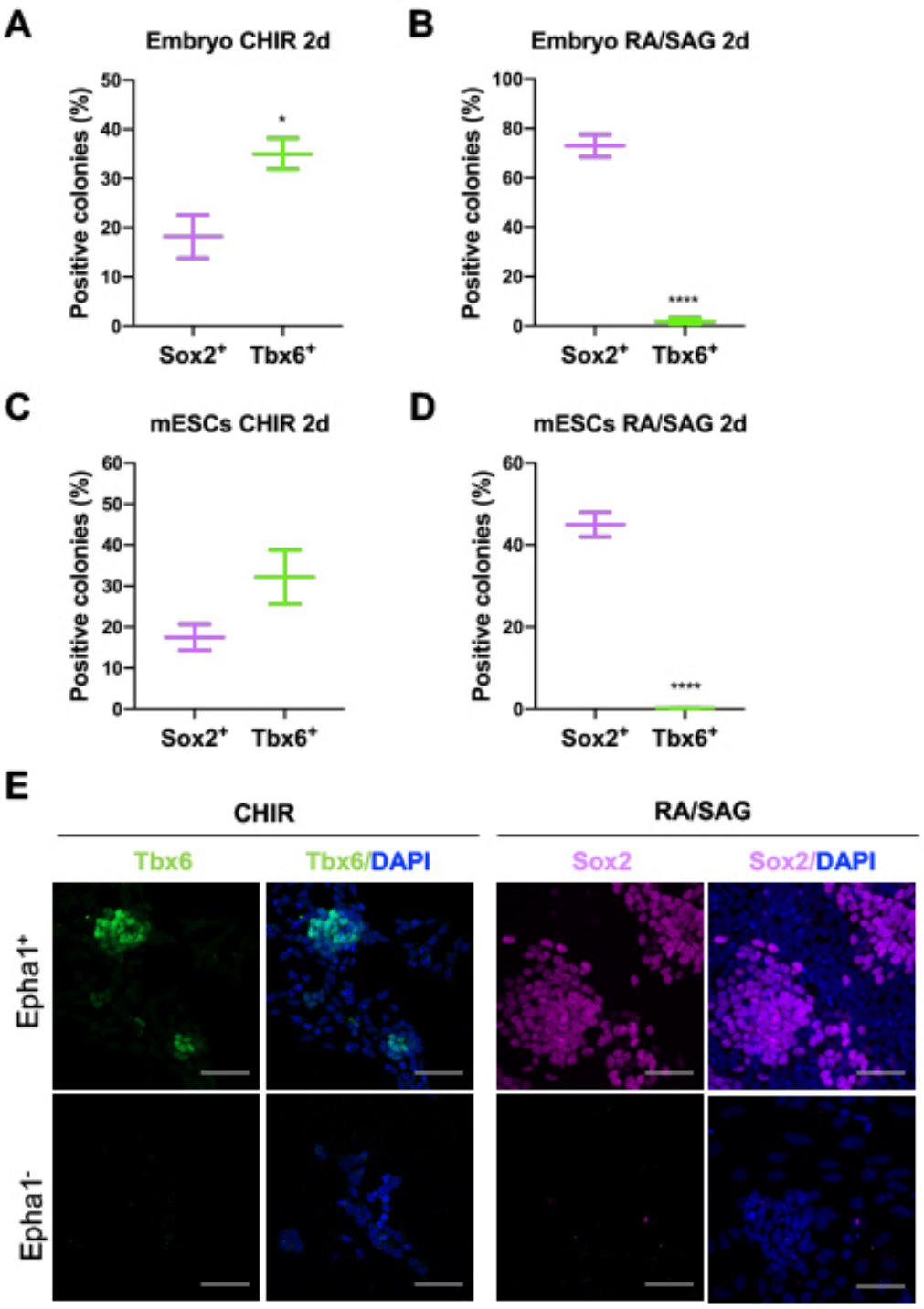

Fig. S6. Differentiation potential of sorted populations. A-D. Quantification of neural (Sox2$^+$) and mesodermal (Tbx6$^+$) derivatives over DAPI of Epha1$^+$ sorted cells from Embryo (E10.5) and mESCs differentiated with CHIR or RA/SAG. Five fields for each condition were counted and the results are presented as mean±SEM. Statistical analysis was assessed using Student's t-test. *p<0.05 and ****p<0.0001. **E.** Immunofluorescence staining of Epha1$^+$ and Epha1$^-$ sorted cells from E10.5 embryo tail buds, differentiated towards mesodermal (CHIR) or neural (RA/SAG) derivatives. Scale bar: 50 µm. "mESCs" means mouse embryonic stem cells.

Table S1. CuffDiff2 differential gene expression analysis between Tail^{Desc} and Tail^{Prog} cell populations. (attached exel file).

Click here to download Table S1

Table S2. CuffDiff2 differential gene expression analysis between Tail^{Desc} and Tail^{Tot} cell populations. (attached exel file).

Click here to download Table S2

Table S3. DEseq2 differential expression analysis between Epha1^{High} and Epha1^{Low} cell populations. (attached exel file).

Click here to download Table S3

Table S4. DEseq2 normalized counts from Tail^{Desc}, Tail^{Prog}, Epha1^{High} and Epha1^{Low} cell populations (attached exel file)

Click here to download Table S4

Table S5. List of primers used in this work.

Primers for genotyping (sequence 5' to 3')		
cre	Forward	CGAGTGATGAGGTTCGCAAG
	Reverse	CACCAGCTTGCATGATCT
***YFP* wild type Allele**	Forward	CTGGCTTCTGAGGACCG
	Reverse	CAGGACAACGCCCACACA
***YFP* mutant Allele**	Forward	AGGGCGAGGAGCTGTTCA
	Reverse	TGAAGTCGATGCCCTTCAG

Primers for RT-qPCR (sequence 5' to 3')		
β-Actin	Forward	ATGAAGATCCTGACCGAGCG
	Reverse	TACTTGCGCTCAGGAGGAGC
Arl4d	Forward	GCCTCGAGGGCTGAAGACACCCCAGCTT
	Reverse	CTGAATTCGCCTTGCTGATCCGGTGTAA
Cdx2	Forward	GCGAAACCTGTGCGAGTGGATG
	Reverse	TTTCCTCTCCTTGGCTCTGCG
Cldn9	Forward	GCCTCGAGGGCTGGCTAGGAACTTTGGT
	Reverse	CTGAATTCGGACACGTACAGCAGAGGAG
Efna1	Forward	GCCTCGAGCTCTCTTGGGTCTGTGCTGC
	Reverse	CTGAATTCGTACTTCCGGGTCATCTGCTT
Epha1	Forward	GCCTCGAGCAAGATTGCAAGACTGTGGC
	Reverse	CTGAATTCCCTCCCACATTACAATCCCA
Mesp2	Forward	GCCATGAGTAGTGGGGTGTC
	Reverse	GTCAGCGGCTCTTTCTAGGG
Ngfr	Forward	GCCTCGAGTGCCTGGACAGTGTTACGTT
	Reverse	CTGAATTCAGGAATGAGGTTGTCAGCGG
Nkd2	Forward	GCCTCGAGGGAGAGAGAGTCCCGAAGGG
	Reverse	CTGAATTCACATGTCCTCTCTGGTGACTT
Olig2	Forward	TTACAGACCGAGCCAACACC
	Reverse	TCAACCTTCCGAATGTGAATTAGA
Sox2	Forward	TTTGTCCGAGACCGAGAAGC
	Reverse	CTCCGGGAAGCGTGTACTTA
Tbxt	Forward	ACCCAGCTCTAAGGAACCAC
	Reverse	GCTGGCGTTATGACTCACAG
Wnt3a	Forward	ATTGAATTTGGAGGAATGGT
	Reverse	CTTGAAGTACGTGTAACGTG
Primers for in situ hybridization probes (sequence 5' to 3')		
Arl4d	Forward	GCCTCGAGGGCTGAAGACACCCCAGCTT
	Reverse	CTGAATTCGCCTTGCTGATCCGGTGTAA
Cldn9	Forward	GCCTCGAGGGCTGGCTAGGAACTTTGGT
	Reverse	CTGAATTCGGACACGTACAGCAGAGGAG
Efna1	Forward	GCCTCGAGCTCTCTTGGGTCTGTGCTGC
	Reverse	CTGAATTCGTACTTCCGGGTCATCTGCTT
Epha1	Forward	GCCTCGAGCAAGATTGCAAGACTGTGGC
	Reverse	CTGAATTCCCTCCCACATTACAATCCCA
Ngfr	Forward	GCCTCGAGTGCCTGGACAGTGTTACGTT
	Reverse	CTGAATTCAGGAATGAGGTTGTCAGCGG
Nkd2	Forward	GCCTCGAGGGAGAGAGAGTCCCGAAGGG
	Reverse	CTGAATTCACATGTCCTCTCTGGTGACTT