High Epha1 expression is a potential cell surface marker for embryonic neuro-mesodermal progenitors

Luisa de Lemos[1], Ana Nóvoa[1] and Moisés Mallo[1,*]

1 Instituto Gulbenkian de Ciência, Rua da Quinta Grande 6, 2780-156 Oeiras, Portugal

* correspondence to Moisés Mallo
e-mail: mallo@igc.gulbenkian.pt

Keywords: Neuro-mesodermal progenitors; axial progenitors; Epha1; embryonic development.

Running title: Epha1 expression and NMPs

ABSTRACT

The basic layout of the vertebrate body is built during the initial stages of embryonic development by the sequential addition of new tissue as the embryo grows at its caudal end. During this process the neuro-mesodermal progenitors (NMPs) are thought to generate the postcranial neural tube and paraxial mesoderm. In recent years, several approaches have been designed to determine the NMP molecular fingerprint but a simple method to isolate them from embryos without the need of transgenic markers is still missing. We wanted to identify a suitable cell surface marker allowing isolation of NMPs from the embryo without the need of previous genetic modifications. We used a genetic strategy to recover NMPs on the basis of their ability to populate the tail bud and searched their transcriptome for cell surface markers specifically enriched in these cells. We found a distinct Epha1 expression profile in progenitor-containing areas of the mouse embryo, consisting in at least two subpopulations of Epha1-positive cells according to their Epha1 expression levels. We show that double Sox2/T(Bra) positive cells are preferentially associated with the Epha1^{High} compartment, indicating that NMPs might be contained within this cell pool. Transcriptional profiling of Epha1-positive tail bud cells also showed enrichment of Epha1^{High} cells in known NMP markers. Interestingly, the Epha1^{Low} compartment contains a molecular signature compatible with notochord progenitor identity. Our results thus indicate that Epha1 could represent a valuable cell surface marker for different subsets of mouse embryonic axial progenitors.

INTRODUCTION

The basic layout of the vertebrate body is formed at early stages of embryonic development by progressive extension of the body axis resulting from the addition of new tissue at the posterior embryonic end (Aires et al., 2018; Stern et al., 2006; Steventon and Martinez Arias, 2017; Wilson et al., 2009). The process of embryonic axial extension depends on a population of cells collectively known as axial progenitors. These progenitors include several cell pools classified according to the tissues they generate. One such cell compartment is the so-called neuro-mesodermal progenitors (NMPs). These cells constitute a population of bipotent cells with self-renewing properties that generate both neural and mesodermal progenitors, which will later originate the spinal cord and the paraxial mesoderm, respectively (Attardi et al., 2018; Cambray and Wilson, 2007; Henrique et al., 2015; Tzouanacou et al., 2009; Wilson et al., 2009).

Since their identification, increasing efforts have been conducted to determine the precise molecular characteristics defining the NMPs. Combined mapping and expression studies first led to their characterization as a population co-expressing the early neural marker Sox2 and the mesodermal transcription factor T (Brachyury) (Cambray and Wilson, 2007; Wymeersch et al., 2016). More recently, the introduction of improved high throughput techniques for transcriptomic analyses have contributed to expand the NMP molecular fingerprint. Accordingly, single-cell transcriptome studies of mouse NMPs provided a deeper understanding of the molecular players and gene networks involved in the regulation of NMP maintenance and differentiation (Gouti et al., 2017; Koch et al., 2017). Those studies also expanded the core signature of NMPs, now including expression of the S*ox2*, *T*, *Nkx1.2*, *Cdx2* and *Cdx4* transcription factors. Another recent study also reported expression of *Tbx6* in a cell subpopulation within the NMP niche, while also providing evidence for its relevance in NMP cell fate decisions (Javali et al., 2017). While most molecular analyses of NMPs have been performed on mouse models, a recent study has reported a transcriptional profile of human NMP-like cells (Verrier et al., 2018). This study also compared this transcriptional signature with that of mouse NMPs, showing a remarkable conservation of many

components of their molecular fingerprints.

Despite all these studies, we are still lacking cell surface markers that could be used to isolate these progenitors in a physiologically active form without relying on previous modifications to introduce reporter genes. In an effort to fill this gap, we adopted a genetic strategy to label the axial progenitors in the mouse embryo in a way that allows their identification and isolation from the tail bud, and searched for candidate genes coding for cell surface markers specifically enriched in NMPs. Detailed analysis of potential candidates revealed expression of Epha1 within the progenitor-containing areas of the mouse embryo. Additional analysis of Epha1-positive cells from the tail bud revealed the presence of at least two subpopulations that could be separated on the basis of their Epha1 expression levels by Fluorescent Activated Cell Sorting (FACS). Interestingly, transcriptional profiling of those cell compartments showed that the Epha1^{High} subpopulation is enriched in known NMP markers, whereas the Epha1^{Low} compartment displays a molecular signature congruent with the existence of notochord progenitors contained within this cell pool. Our results thus indicate that Epha1 could represent a valuable cell surface marker for the identification of different subsets of mouse embryonic axial progenitors.

RESULTS

Labeling and isolation of axial progenitors from developing mouse embryos

To isolate axial progenitors in a way compatible with the analysis of their transcriptional signature we took advantage of a genetic approach that we have recently developed to follow the fate of axial progenitors after labeling them at early developmental stages (Aires et al., 2019). With this system, which combines the *Cdx2P-Cre^{ERT}* transgene (Jurberg et al., 2013) with a ROSA26 reporter (Soriano, 1999; Srinivas et al., 2001), a single low tamoxifen dose administered at embryonic stage (E) 7.5 is able to induce a short pulse of permanent labeling into a subset of axial progenitors. Due to the continuous and cumulative nature of axial growth, labeled cells will contribute to the embryonic tissues posterior to the position of effective cre-mediated recombination, all the way down to the tail tip (Aires et al.,

2019), including the NMPs in the tail bud. We therefore labeled axial progenitors with a fluorescent marker using the ROSA26-YFP-reporter (Srinivas et al., 2001) in combination with the *Cdx2P-Cre*^{ERT} transgene. Descendants of progenitors labeled at E7.5 were then recovered from the tail region at E10.5 by fluorescence-activated cell sorting (FACS). YFP-positive cells were recovered from two areas of the tail (Fig 1A): the tail bud (Tail^{Prog}), which is expected to contain the axial progenitors, and a more anterior region, where labeled cells have already generated axial progenitor derivatives (Tail^{Contr}). The transcriptome of these two cell pools was obtained by RNA-seq, and then compared to identify genes differentially expressed in the Tail^{Prog} compartment that could thus provide a molecular signature of the axial progenitors (Fig. 1 and Table S1). This comparison identified 1458 genes showing differential expression ($p<0,05$) between the two cell groups (Fig. 1B and Table S1). Of these, 847 genes were highly expressed in the Tail^{Prog}, whereas 611 genes were up-regulated in the Tail^{Contr} sample. A selection of 12 differentially expressed genes was then used to validate the RNA-seq data by RT-qPCR (Fig. 1C).

Initial analysis of these data revealed a high enrichment of the Tail^{Prog} cells in factors that have been linked to axial progenitor identity (Fig. 1D). For instance, *Cdx2*, *Cdx4* and *T* (*Brachyury*), known to be highly expressed in axial progenitors 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 up-regulated genes in the Tail^{Prog} compartment. Similarly, other genes whose expression is known to be enriched in axial progenitors, including *Nkx1.2*, *Fgf8*, *Fgf4*, *Fgf3*, *Evx1*, *Cyp26a*, *Hoxb1*, *Sp5*, *Gdf11*, *Wnt5a* or *Wnt3a* (Abu-Abed et al., 2001; Boulet and Capecchi, 2012; Dush and Martin, 1992; Greco et al., 1996; Harrison et al., 2000; McKay et al., 1996; McPherron et al., 1999; Murphy and Hill, 1991; Naiche et al., 2011; Sakai et al., 2001; Takada et al., 1994; Verrier et al., 2018; Yamaguchi et al., 1999), also showed significant differential expression in the Tail^{Prog} cell pool. In addition, *Tbx6*, which was recently described as a marker for tail bud NMPs (Javali et al., 2017) was significantly up-regulated in Tail^{Prog} cells. These cells also contained significant

levels of *Sox2* transcripts, suggesting the existence of double Sox2+/T+ cells contained within this population, which have been extensively used as signature of NMP cells (Cambray and Wilson, 2007; Koch et al., 2017; Wymeersch et al., 2016). *Sox2* expression was, however, higher in Tail^{Contr} than in Tail^{Prog} cells (Fig 1D). This is likely to result from the extensive contribution of the Tail^{Contr} cell pool to the neural lineage, which is characterized by higher levels of *Sox2* expression than the axial progenitors. Expression of other known neural regulators, such as *Sox1*, *Ngn2, Pax6, Olig2* or *Olig3* (Aubert et al., 2003; Gradwohl et al., 1996; Takeichi et al., 2002; Walther and Gruss, 1991), was also significantly higher in Tail^{Contr} than in Tail^{Prog} cells (Fig 1D), which is consistent with this interpretation.

Together, these data indicate that the Tail^{Prog} compartment is highly enriched in tail bud axial progenitors. Moreover, we found high concordance between our dataset and previous published data from *in vivo* single-cell and *in vitro* differentiated hESCs or mESCs RNA-seq analysis (Gouti et al., 2017; Koch et al., 2017; Verrier et al., 2018), which further validates our lineage tracing strategy.

Identification of a cell surface marker of NMPs/Axial progenitors

We then concentrated our attention on genes coding for membrane proteins that could be used to isolate physiologically active NMPs without previous genomic modifications. Gene Ontology (GO) categorization (Ashburner et al., 2000) of genes differentially up-regulated in Tail^{Prog} cells with a $\log_2$ Fold change > 2 and q value < 0.05 identified 61 genes assigned to the category “membrane” (GO:0005886). From these, we further selected the 16 genes that had a read value >10 in the Tail^{Prog} RNA-seq dataset (Fig. 2A), as we have observed experimentally that, in these datasets, this is the lower threshold level allowing mRNA or protein detection in the embryo using conventional methods. 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. 2B). In addition, for most of them, tail bud expression was only a part of somewhat more complex expression patterns that included other embryonic regions.

From these genes we focused on *Epha1* based not only on its expression pattern at different developmental stages, but also due to the existence of FACS-validated antibodies able to provide reliable data with cells obtained from solid embryonic tissues. In particular, at E8.5 we observed *Epha1* expression in the caudal lateral epiblast (Fig. 2Ca-c'), which coincides with the region containing NMPs at early somite-stages (Baillie-Johnson et al., 2018; Cambray and Wilson, 2007, 2002; McGrew et al., 2008; Wymeersch et al., 2016). Strong *Epha1* expression was also observed in the tail tip of E9.5 embryos, fading anteriorly when entering the regions corresponding to the presomitic mesoderm and the caudal neural tube (Fig. 2Cd-f'). This expression pattern was maintained in E10.5 embryos (Fig. 2Cg-i'). Together, the *Epha1* expression pattern shows considerable overlap with embryonic regions known to contain axial progenitors at various developmental stages. This transcript distribution is consistent with previous reports, where *Epha1* expression was also observed in the primitive streak of early head fold embryos (Duffy et al., 2006), thus fitting this progenitor-rich domain.

We then performed FACS analysis of cells obtained from the tail bud and adjacent anterior tail region of E10.5 embryos using antibodies against Epha1. Both areas contained a high proportion of Epha1 positive cells (Fig. 3A-C), which was somehow surprising considering the significant differences observed in the transcriptomic data obtained from these two regions. This could indicate higher stability of the Epha1 protein than of its mRNA. Interestingly, however, the staining patterns in the two cell populations were different, as cells from the tail bud included an additional population with higher staining intensity that was never observed in the FACS plots from the anterior tail region (Fig. 3A, B). We therefore grouped Epha1-positive tail bud cells in two subpopulations on the basis of their Epha1 content, which will be referred to as Epha1^{High} and Epha1^{Low}. Interestingly, we also obtained Epha1^{High} and Epha1^{Low} cell compartments in the epiblast-containing area of E8.5 embryos, although the proportion of Epha1^{High} cells was lower in this tissue than in the tail bud from E10.5 embryos (Fig. 3D, G).

The above data suggest that axial progenitors might indeed be contained within the Epha1^{High} cell population. To assess this possibility we first evaluated the

distribution of Sox2^{+}/T^{+} cells, a common criterion used to identify NMPs (Cambray and Wilson, 2007; Koch et al., 2017; Wymeersch et al., 2016), among the different Epha1 compartments. We therefore isolated Epha1-negative, Epha1^{Low} and Epha1^{High} cell pools from the tail bud of E10.5 embryos and analyzed them for Sox2 and T expression (Fig. 3D-F; Table 1). We observed that the proportion of cells co-expressing Sox2 and T differed significantly among the three cell populations, being 56,93 ± 5,33% for Epha1^{High} cells, 28,03 ± 3,75% for Epha1^{Low} cells and considerably lower from the Epha1-negative pool (10,12 ± 5,00%), consistent with NMPs being indeed enriched in the Epha1^{High} compartment (Fig. 3F). Interestingly, Epha1-negative cells presented a higher proportion of Sox2^{-}/T^{+} cells (70,63 ± 16,48%) than any of the Epha1-positive cell populations (Fig. 3E and Table 1), suggesting that they might represent cells entering mesodermal differentiation.

Enrichment in Sox2^{+}/T^{+} cells within the Epha1^{High} subpopulation was also observed at E8.5. In particular, we obtained 66,67 ± 15,55% and 27,20 ± 9,46% of Sox2$^{+/}$T^{+} cells for Epha1^{High} and Epha1^{Low} cell compartments, respectively (Fig. 3H, I; Table 1), suggesting that NMPs might also be enriched within the Epha1^{High} pool in the E8.5 epiblast. Likewise, we also detected Epha1^{High} and Epha1^{Low} subsets in ES cell cultures incubated under conditions promoting NMP differentiation (Gouti et al., 2014; Turner et al., 2014) (Fig. 3J). When we assessed the Sox2 and T profiles in these cells, Sox2^{+}/T^{+} cells were preferentially found in the Epha1^{High} population. Indeed, most of the cells in this cell compartment (97,45 ± 1,04%) were positive for both markers (Fig. 3K-L; Table 1), further suggesting NMP enrichment also in this Epha1-positive compartment of *in vitro* differentiated ES cells.

Tail bud-NMPs have been recently associated with the Tbx6^{+}/Sox2^{+} expression phenotype (Javali et al., 2017). Analysis of these two markers in the different Epha1 compartments obtained from E10.5 tail buds showed a slight enrichment of double positive Tbx6 and Sox2 cells in the Epha1^{High} pool (62,30 ± 8,78%) (Fig. 3M-O; Table 1). This observation gives further support to the connection between the Epha1^{High} phenotype and NMP progenitors in the tail bud.

Transcriptional analysis of distinct Epha1-positive cell subpopulations

To further characterize the Epha1-positive cell populations in the tail bud we isolated the Epha1^{High} and Epha1^{Low} cell compartments from this region of E10.5 embryos and analyzed their transcript content by RNA-seq (Fig. 4A; Table S2). We then performed a three-way Venn analysis of the genes differentially expressed with a q-value < 0.05 obtained from three pairwise comparisons between the RNA-seq data of Tail^{Contr} cells and each of the transcriptomes of the Epha1^{High}, Epha1^{Low} and Tail^{Prog} cell populations. This comparison revealed 258 differentially expressed genes shared by these three cell pools, from which 161 were up regulated and 97 down regulated genes (Fig. 4B and Table 2). Tail^{Prog} and Epha1^{High} shared additional 44 up regulated and 38 down regulated genes (Table 2), whereas only 7 differentially expressed genes were common exclusively in Tail^{Prog} and Epha1^{Low} cells (Table 2). A closer look at the 161 differentially up regulated genes shared by Epha1^{High}, Epha1^{Low} and Tail^{Prog} revealed that, despite their common differential expression relative to Tail^{Contr}, the expression levels of many of them were clearly different in the three cell compartments (Fig. 4A,C). Significantly, the expression profiles obtained from the Epha1^{High} and Tail^{Prog} compartments were often more similar than those obtained with Epha1^{Low} cells. This was particularly clear for many of the genes representing the known molecular signature of axial progenitors, some of which were also present in the gene set shared only by Epha1^{High} and Tail^{Prog} cells (Fig. 4C). Also, genes that have been shown to be enriched in progenitors that are leaving the bipotent compartment to enter paraxial mesodermal fates (Koch et al., 2017), including *Msgn1*, *Rspo3*, *Lef1*, *Snai1*, *Tbx6*, *Sp5* or *Hes7*, presented higher expression levels in the Epha1^{High} compartment (Fig. 4C). These results are consistent with the Epha1^{High} population being enriched in axial progenitors, in agreement with the profiles obtained for Sox2 and T expression.

Interestingly, these gene expression analyses also revealed that the Epha1^{Low} compartment contained high abundance of transcripts for *Shh*, *Noto*, *Foxa2*, *Krt8* and *Krt18*, none of which was detected at significant levels in any of the other datasets (Fig. 4C). All 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), thus suggesting that the tail bud Epha1^{Low} cell population is enriched in notochord progenitors. Consistent with this, *Epha2*, which in the tail region is restricted to the most caudal part of the notochord and required for its elongation (Naruse-Nakajima et al., 2001), was expressed at substantially higher levels in the Epha1^{Low} than in the Epha1^{High} compartment (Fig. 4C).

DISCUSSION

In this study, we have used a genetic strategy to isolate physiologically active NMPs from the tail bud of mouse embryos and analyzed their mRNA content using high throughput methods. This approach stems from the observation that epiblast and tail bud axial progenitors belong to the same cell lineage (Brown and Storey, 2000; Cambray and Wilson, 2007, 2002; Tzouanacou et al., 2009). Hence, a permanent label introduced into axial progenitors at early developmental stages allows isolation of their descendants in the tail bud. Differential expression of known markers for NMPs in cells isolated with this strategy indicates that they do represent bona fide axial progenitors. However, it should be noted that, to restrict the time frame of effective cell labeling, we used conditions that also limit the number of effectively labeled progenitors. As a consequence, the tail bud cells isolated and analyzed in this work represent a fraction of the actual progenitors and, therefore, it is most likely that in these experiments we did not obtain a full catalog of the tail bud NMP transcriptome.

Our analyses led to the identification of several cell surface proteins that could be tested for their use as NMP markers allowing their isolation without the need of previous genetic manipulations. From these we chose Epha1 for further investigation given the existence of reliable and FACS-validated antibodies working in solid embryonic tissues. It will be important to identify and verify antibodies that would allow similar studies for other candidates and determine whether they label similar or different subsets of progenitors.

While mRNA *in situ* hybridization experiments suggested that *Epha1* expression is restricted to progenitor containing regions, FACS analyses using

Epha1 antibodies revealed a wider Epha1 distribution, indicating that the presence of Epha1 by itself cannot be regarded as a hallmark of NMPs. Indeed, Epha1-positive cells were not restricted to the tip of the tail bud but could also be identified in more anterior areas containing tissues that had already entered differentiation routes. Immunofluorescence analyses also revealed broader Epha1 protein than mRNA distribution, suggesting higher stability of the protein than of the transcript. However, Epha1 protein levels were not uniform in the different regions that we tested, a feature that seems to be relevant to define specific cell populations within the tail bud. In particular, we identified a subset cells containing high Epha1 levels that could be indeed enriched in axial progenitors. Accordingly, the Epha1^{High} cell compartment seemed to be exclusive of embryonic areas containing NMPs, such as the epiblast at E8.5 or the tail bud at E10.5, and could be also readily identified in ES cells incubated under NMP-promoting conditions. The significant enrichment in Sox2^{+}/T^{+} cells in the Epha1^{High} cell populations isolated from the various embryonic tissues and ES cell-differentiated NMPs further supports a correlation between high Epha1 levels and NMPs. The transcriptomic profile of Epha1^{High} cells was also consistent with this correlation as it is enriched in transcripts of genes typically associated with the NMP subset of axial progenitors. At this stage, however, we cannot be sure about whether the Epha1^{High} phenotype can be considered a general marker for NMPs or if it identifies a subpopulation within those cells. Yet, the RNA-seq data suggests that at least a fraction of the Epha1^{High} cell compartment might indeed represent progenitors entering mesodermal fates.

Quite surprisingly, the RNA-seq data from tail bud Epha1^{Low} cells revealed a distinct enrichment in notochord markers, suggesting that they might contain progenitors for this structure. Whether the tail bud Epha1^{Low} cell pool derives from Epha1^{High} cells that reduced their Epha1 protein load or they represent an independent cell population cannot be definitely established from our data and will require a direct cell tracing approach. The heat maps of their transcriptomes indicate that they embody different cell populations. However, the finding that Epha1^{High} and Epha1^{Low} cells share a large number of genes differentially expressed relative to Tail^{Contr} cells suggests that the main differences observed in

their heat maps might result fundamentally from dissimilarities in gene expression levels, which most likely reflects a gradual reduction of these transcripts over time specifically in the cells undergoing reduction of their Epha1 protein load. On the other hand, the extremely low or absent expression of notochord markers in Epha1^{High} cells argues against these cells being precursors the Epha1^{Low} subset, at least in what notochord progenitors are concerned. The finding that Tail^{Prog} cells also lack signs of notochord-specific gene expression likewise suggests that notochord progenitors do not derive from this cell population. Interestingly, some of these notochord markers are also expressed in the node (Abdelkhalek et al., 2004; Ang et al., 1993; Jeong and Epstein, 2003), indicating that this Epha1^{Low} cell population might have a different origin.

It is important to consider that both Epha1^{High} and Epha1^{Low} cell compartments might not be homogenous, but rather they consist of a mix of cell groups with different characteristics. In this context, Epha1^{Low} cells might include at least two subpopulations, one with low notochord marker expression that could indeed derive from Epha1^{High} cells, and another with features associated with notochord progenitors but with an independent origin of that of Epha1^{High} and Tail^{Prog} cells. Solving these issues will require single cell transcriptomic analyses. Also, it will be interesting to understand whether cell sorting strategies combining Epha1 with other surface markers could help sub fractioning Epha1^{High} and Epha1^{Low} cells into subgroups with defined functional profiles.

EXPERIMENTAL PROCEDURES

Mice and embryos

In this work 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, mattings were set up between *Cdx2P-cre^{ERT}* (Jurberg et al., 2013) and *ROSA26-YFP-R* (Srinivas et al., 2001) mice. Pregnant females were treated at E7.5 with a single intraperitoneal injection of 200 µl of a 1 mg/ml solution of tamoxifen (Sigma #T5648) in corn oil. Embryos were then collected at E10.5 by caesarean section, dissected in ice-cold PBS and

processed for cell sorting (see below). Wild type embryos were collected at E8.5, E9.5 and E10.5, dissected in ice-cold PBS and processed for *in situ* hybridization, gene expression or flow cytometry studies.

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

Genotyping

Adult mice were genotyped from tail biopsies. Samples were incubated overnight at 55°C in PBND buffer (50 mM KCl, 10 mM Tris-Hcl pH 8.3, 2.5 mM $MgCl_2$, 0.1 mg/ml Nonidet P40, 0.25% Tween-20) containing 200 µg/ml proteinase K (Roche #P4032). Lysates were then heat-inactivated for 15 minutes at 95°C before being used for the genotyping PCR reactions. To genotype embryos yolk sacs were incubated overnight at 55°C in yolk sac lysis buffer (50 mM KCl, 10 mM Tris-HCl pH 8.3, 2 mM $MgCl_2$, 0.45% Tween-20, 0.45% Nonidet P40) containing 200 µg/ml proteinase K. Proteinase K was inactivated as above and 2 µl of the lysate was then used to set up the genotyping PCR reactions. Primers used for genotyping transgenic mice are listed in Table 3.

Cell Sorting for RNA-sequencing by Fluorescence-Activated Cell sorting (FACS)

Two regions were collected from E10.5 *Cdx2P-cre*^{ERT}::*ROSA26-YFP-R* embryos: the tail buds and a more proximal region of the tail tip (the last 3 somites). To obtain a single cell suspension, tissue was incubated on ice for 5 minutes in Accutase solution (Sigma #A6964). Digestion was terminated by adding two volumes of PBS/10% donkey serum (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 YFP negative control. YFP-positive cells were collected directly in TRI Reagent® (Sigma #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^{Contr}, respectively.

To isolate Epha1 cell populations, single cell suspensions were obtained from tail buds and proximal tail regions of E10.5 or from the epiblast of E8.5 wild type embryos as mentioned above. Cells were incubated with 100 µl of blocking solution (10% DS with 1:100 dilution of 2.4G2 anti-mouse Fc block in PBS) on ice for 30 minutes and then stained on ice for 1 hour with a 1:100 dilution of goat anti-Epha1 primary antibody (R&D systems #AF3034). After two washes with PBS/10% DS, cells were incubated with a 1:1500 dilution of donkey anti-goat A647 (Thermo Fisher #A-21447) secondary antibody for another hour on ice and subsequently washed twice again with PBS/10% DS. Stained cells were resuspended in PBS/2% DS, filtered through a cell strainer and sorted on a FACSAria IIu (BD Biosciences). Epha1-positive cells were identified using a 633 nm excitation laser with filter detection of 660/20. Gating conditions of Epha1^{High} and Epha1^{Low} cell populations 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, in order to obtain a sufficient concentration of high quality RNA.

RNA-sequencing analysis

Total RNA was isolated from the TRI Reagent® suspension following the manufacture's protocol, with the addition of 10 µg RNase-free glycogen (Roche #10901393001) in the isopropanol step. RNA samples were then resuspended in RNase-free water. RNA concentration and purity was determined on an AATI Fragment Analyzer (Agilent). For the Tail^{Contr} 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 Ilumina 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 $Tail^{Contr}$ and $Tail^{Prog}$ 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 bases 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). In order 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 DESeq2 R package. These data was then used to build a Venn diagram using the VennDiagram R package.

The sequencing data of the RNA-seq experiments was deposited in the NCBI trace and Short-read Archive (SRA), accession numbers PRJNA527654 and PRJNA527619.

Gene ontology (GO) analysis (Ashburner et al., 2000) was used on differentially expressed genes according to the criteria $\log_2$ Fold change > 2 and a *q* value < 0.05.

Quantitative RT-PCR

1 µg of total RNA was reverse-transcribed into complementary DNA (cDNA) and following the protocol of the NZY Reverse transcriptase enzyme (NZYTech #MB08301) using random hexamer priming. Real-time PCR was performed with iQ™ SYBR® Green Supermix (Bio-Rad #1708880) according to manufacturer's instructions and using the CFX384 Real Time PCR detection system (Bio-Rad). The expression levels were normalized to *β-actin* and changes in fold expression

were calculated using $2^{-\Delta\Delta Ct}$ method. Gene expression data was presented as the mean±SD. The sequences of the primers used are given in Table 4.

Protein expression profile analysis by FACS

Cells obtained from embryos and stained with Epha1 antibodies 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. They were incubated for 1 hour with primary antibodies combining: rabbit anti-Sox2 (Abcam #ab92494) (1:200) and mouse anti-T/Bra (Santa Cruz #sc-166962) (1:100), or rabbit anti-Sox2 (1:200) and mouse anti-Tbx6 (Santa Cruz #517027) (1:100). After primary incubation, cells were washed twice with the kit's Perm buffer and incubated for another hour with secondary antibodies donkey anti-rabbit A568 (Thermo Fisher Scientific #A10042) (1:1500) and donkey anti-mouse Alexa Fluor® 488 (Abcam #ab150105) (1:1500). After two washes in Perm Buffer, cells were resuspended in 300 µl of PBS for FACS analysis. Data acquisition was performed in a flow cytometer LSR Fortessa X20 (BD Biosciences). Expression levels were assessed after removing debris on the basis of forward scatter-area (FSC-A) and side scatter signals (SSC-A). Doublets were discriminated using FSC-A and forward scatter–width (FSC-W) signals. 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 total Epha1-positive cells histogram (A647 channel). Quadrant gates were established according to fluorescence levels detected by the control samples processed without primary antibodies. Flow cytometry data were analyzed using FlowJo™ 10 (BD, Biosciences) software. Quadrant averages were calculated using at least 3 independent experiments and one-way analysis of variance ANOVA was used to determine statistical significance.

***In situ* hybridization**

Whole mount *in situ* hybridization was performed using *in vitro* transcribed digoxigenin-labeled antisense RNA probes as previously described (Kanzler et al., 1998). Briefly, embryos were dissected out in PBS and fixed in 4% paraformaldehyde (PFA) at 4°C overnight. Embryos were washed in PBS containing 0.1 % Tween-20 (PBT), dehydrated with methanol and rehydrated with PBT. They were then treated with proteinase K (10 µg/ml in PBT) at room temperature for a time that depended on the embryo stage (4, 7.5 or 9 minutes for E8.5, E9.5 or E10.5 embryos, respectively). The reaction was stopped with glycine (2 mg/ml in PBT) and embryos were postfixed with 4% PFA, 0,2% glutaraldehyde. Hybridization was performed at 65°C overnight in hybridization solution [50% formamide, 1.3x SSC (3 M NaCl, 300 mM sodium citrate, pH 5.5), 5 mM EDTA, 0.2% Tween-20, 50 µg/ml yeast tRNA, 100 µg/ml heparin] containing the RNA probe, followed by three washes at 65°C in hybridization solution without RNA probe, tRNA and heparin. Embryos were then washed in TBST (23 mM Tris-HCl, pH 8.0, 140 mM NaCl , 2.7 mM KCl, 0.1% Tween-20), equilibrated with MABT (100 mM maleic acid, 150 mM NaCl, 0.1% Tween-20, pH 7.5), blocked with MABT/Block [MABT containing 1% blocking reagent (Roche #11096176001) and 10% Sheep Serum] for 2-3 hours and incubated with a 1:2000 dilution of alkaline phosphatase-conjugated anti-digoxigenin antibody (Roche #11093274910) in MABT/Block at 4°C overnight. Embryos were washed extensively with MABT at room temperature for 24 hours, equilibrated in NTMT (100 mM Tris-HCl, pH 9.5, 50 mM $MgCl_2$, 100mM NaCl, 0,1% Tween-20) and developed at room temperature with NBT/BCIP (Roche #11681451001) diluted in NTMT. Reactions were stopped with PBT, fixed in 4% PFA and stored in PBT.

Probes for *Efna1*, *Epha1*, *Ngfr*, *Cldn9*, *Nkd2*, *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 5.

To section whole mount-stained embryos, these were included in gelatin/albumin (0.45% gelatin, 270 mg/ml bovine serum albumin, 180 mg/ml sucrose in PBS, jellified with 1.75% glutaraldehyde). Sections were cut at 35 µm

with a vibratome and mounted with an aqueous mounting solution (Aquatex®, Merck #108562).

Mouse ES cell culture and differentiation

CJ7 mouse embryonic stem (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 AA solution (Sigma #M-7145), 2 mM L-glutamine, 1% EmbryoMax Nucleosides (Millipore #ES-008-D), 100 U/ml Penicillin and 100 µg/ml Streptomycin (Sigma #P7539), 0.1 mM β-mercaptoethanol and 1000 U/ml LIF (Millipore #ESG1107)] on mitomycinC-inactivated primary mouse embryo fibroblasts. To start differentiation, ES cells were derived into NMP-like cells 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 #59417C) and seeded at a density of 5.000 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 supplemented with 1x N-2 (Gibco #LS17502048) and 1x 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. After differentiation cells were detached with Accutase solution and processed for FACS analysis as described above.

AUTHOR CONTRIBUTION

Conceptualization: L. dL. and M.M.; Methodology: L.dL. and M.M.; Investigation: L.dL. and A.N.; Writing-Original Draft: L. dL. and M.M.; Funding Acquisition: M.M.; Supervision: M.M.

ACKNOWLEDGMENTS

We would like to thank the IGC animal house, Flow Cytometry and Genomics Facilities for their expert services and assistance, and Daniel Sobral from the IGC Bioinformatics Unit for assistance with RNA-seq analysis. We would also to acknowledge all members of the Mallo laboratory for helpful discussions and comments of the course of the project and Rita Aires and Ana Casaca for reading the manuscript. This work has been supported by grants PTDC/BEX-BID/0899/2014 (FCT, Portugal) and SCML-MC-60-2014 (from Santa Casa da Misericordia de Lisboa, Portugal).

REFERENCES

Abdelkhalek H Ben, Beckers A, Schuster-gossler K, Pavlova MN, Burkhardt H, Lickert H, Rossant J, Reinhardt R, Schalkwyk LC, Herrmann BG, Ceolin M, Rivera-pomar R, Gossler A. 2004. The mouse homeobox gene Not is required for caudal notochord development and affected by the truncate mutation Hanaa. *Genes Dev* **18**:1725–1736

Abu-Abed S, Dollé P, Metzger D, Beckett B, Chambon P, 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

Aires R, de Lemos L, Nóvoa A, Jurberg AD, Mascrez B, Duboule D, Mallo M. 2019. Tail Bud Progenitor Activity Relies on a Network Comprising Gdf11, Lin28, and Hox13 Genes. *Dev Cell* **48**:383–395

Aires R, Dias A, Mallo M. 2018. Deconstructing the molecular mechanisms shaping the vertebrate body plan. *Curr Opin Cell Biol* **55**:81–86

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

Ang S-L, Wierda A, Wong D, Stevens KA, Cascio S, Rossant J, Zaret KS. 1993. The formation and maintenance of the definitive endoderm lineage in the mouse: involvement of HNF3/forkhead proteins. *Development* **119**:1301–1315

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. 2000. Gene Ontology: tool for the unification of biology. *Nat Genet* **25**:25–29

Attardi A, Fulton T, Florescu M, Shah G, Muresan L, Lenz MO, Lancaster C, Huisken J, van Oudenaarden A, Steventon B. 2018. Neuromesodermal progenitors are a conserved source of spinal cord with divergent growth dynamics. *Development* **145**:dev166728

Aubert J, Stavridis MP, Tweedie S, Reilly MO, Vierlinger K, Li M, Ghazal P, Pratt T, Mason JO, Roy D, Smith A. 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 U S A* **100**:11836–11841

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

Boulet AM, Capecchi MR. 2012. Signaling by FGF4 and FGF8 is required for axial elongation of the mouse embryo. *Dev Biol* **371**:235–245

Brown JM, Storey KG. 2000. A region of the vertebrate neural plate in which neighbouring cells can adopt neural or epidermal fates. *Curr Biol* **10**:869–872

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

Cambray N, Wilson V. 2007. Two distinct sources for a population of maturing axial progenitors. *Development* **134**:2829–2840

Chawengsaksophak K, de Graaff W, Rossant J, Deschamps J, Beck F. 2004. Cdx2 is essential for axial elongation in mouse development. *Proc Natl Acad Sci U S A* **101**:7641–7645

Duffy SL, Steiner KA, Tam PPL, Boyd AW. 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

Dush MK, Martin GR. 1992. Analysis of mouse Evx genes: Evx-1 displays graded expression in the primitive streak. *Dev Biol* **151**:273–287

Echelard Y, Epstein DJ, St-Jacques B, Shen L, Mohler J, McMahon JA, McMahon AP. 1993. Sonic hedgehog, a member of a family of putative signaling molecules, is implicated in the regulation of CNS polarity. *Cell* **75**:1417–1430

Gouti M, Delile J, Stamataki D, Wymeersch FJ, Huang Y, Kleinjung J, Wilson V, Briscoe J. 2017. A gene regulatory network balances neural and mesoderm specification during vertebrate trunk development. *Dev Cell* **41**:1–19

Gouti M, Tsakiridis A, Wymeersch FJ, Huang Y, Kleinjung J, Wilson V, 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

Gradwohl G, Fode C, Guillemot F. 1996. Restricted expression of a novel murine atonal-related bHLH protein in undifferentiated neural precursors. *Dev Biol* **180**:227–241

Greco TL, Takada S, Newhouse MM, McMahon JA, McMahon AP, Camper SA. 1996. Analysis of the vestigial tail mutation demonstrates that Wnt-3a gene dosage regulates mouse axial development. *Genes Dev* **10**:313–324

Harrison SM, Houzelstein D, Dunwoodie SL, Beddington RSP. 2000. Sp5, a new member of the Sp1 family, is dynamically expressed during development and genetically interacts with Brachyury. *Dev Biol* **227**:358–372

Henrique D, Abranches E, Verrier L, Storey KG. 2015. Neuromesodermal progenitors and the making of the spinal cord. *Development* **142**:2864–2875

Herrmann BG, Labeit S, Poustka A, King TR, Lehrach H. 1990. Cloning of the T gene required in mesoderm formation in the mouse. *Nature* **343**:617–622

Javali A, Misra A, Leonavicius K, Acharya D, Vyas B, Sambasivan R. 2017. Co-expression of Tbx6 and Sox2 identifies a novel transient neuromesoderm progenitor cell state. *Development* **144**:4522–4529

Jeong Y, Epstein DJ. 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

Jurberg AD, Aires R, Varela-Lasheras I, Nóvoa A, Mallo M. 2013. Switching axial progenitors from producing trunk to tail tissues in vertebrate embryos. *Dev Cell*

25:451–462

Kanzler B, Kuschert SJ, Liu YH, Mallo M. 1998. Hoxa-2 restricts the chondrogenic domain and inhibits bone formation during development of the branchial area. *Development* **125**:2587–2597

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2 : accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**:R36

Koch F, Scholze M, Wittler L, Schifferl D, Sudheer S, Grote P, Timmermann B, Macura K, Herrmann BG. 2017. Antagonistic Activities of Sox2 and Brachyury Control the Fate Choice of Neuro-Mesodermal Progenitors. *Dev Cell* **42**:514–526

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**:357–359

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**:550

McGrew MJ, Sherman A, Lillico SG, Ellard FM, Radcliffe PA, Gilhooley HJ, Mitrophanous KA, Cambray N, Wilson V, Sang H. 2008. Localised axial progenitor cell populations in the avian tail bud are not committed to a posterior Hox identity. *Development* **135**:2289–99

McKay IJ, Lewis J, Lumsden A. 1996. The role of FGF-3 in early inner ear development: An analysis in normal and kreisler mutant mice. *Dev Biol* **174**:370–378

McPherron AC, Lawle AM, Lee S-J. 1999. Regulation of anterior/posterior patterning of the axial skeleton by growth/differentiation factor 11. *Nat Genet* **22**:260–264

Murphy P, Hill RE. 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

Naiche L a, Holder N, Lewandoski M. 2011. FGF4 and FGF8 comprise the wavefront activity that controls somitogenesis. *Proc Natl Acad Sci U S A* **108**:4018–4023

Naruse-Nakajima C, Asano M, Iwakura Y. 2001. Involvement of EphA2 in the formation of the tail notochord via interaction with ephrinA1. *Mech Dev* **102**:95–105

Picelli S, Faridani OR, Björklund ÅK, Winberg G, Sagasser S, Sandberg R. 2014. Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc* **9**:171–181

Rodrigues-Pinto R, Berry A, Piper-Hanley K, Hanley N, Richardson SM, Hoyland JA. 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

Sakai Y, Meno C, Fujii H, Nishino J, Shiratori H, Saijoh Y, Rossant J, 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

Savory JGA, Mansfield M, Rijli FM, Lohnes D. 2011. Cdx mediates neural tube closure through transcriptional regulation of the planar cell polarity gene *Ptk7*. *Development* **138**:1361–1370

Soriano P. 1999. Generalized lacZ expression with the ROSA26 Cre reporter strain. *Nat Genet* **21**:70–71

Srinivas S, Watanabe T, Lin C-S, William CM, Tanabe Y, Jessell TM, Costantini F. 2001. Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus. *BMC Dev Biol* **1**:4

Stern CD, Charité J, Deschamps J, Duboule D, Durston AJ, Kmita M, Nicolas JF, Palmeirim I, Smith JC, Wolpert L. 2006. Head-tail patterning of the vertebrate embryo: One, two or many unresolved problems? *Int J Dev Biol* **50**:3–15

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

Swiatek PJ, 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

Takada S, Stark KL, Shea MJ, Vassileva G, McMahon JA, McMahon AP. 1994.

Wnt-3a regulates somite and tailbud formation in the mouse embryo. *Genes Dev* **8**:174–189

Takeichi M, Okubo K, Uchida T, Ohtsuki T, Chisaka O, Takebayashi H, Ikenaka K, Kawamoto S, Nabeshima Y. 2002. Non-overlapping expression of Olig3 and Olig2 in the embryonic neural tube. *Mech Dev* **113**:169–174

Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. 2013. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol* **31**:46–53

Turner D a, Hayward PC, Baillie-Johnson P, Rué P, Broome R, Faunes F, 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

Tzouanacou E, Wegener A, Wymeersch FJ, Wilson V, Nicolas JF. 2009. Redefining the Progression of Lineage Segregations during Mammalian Embryogenesis by Clonal Analysis. *Dev Cell* **17**:365–376

van Rooijen C, Simmini S, Bialecka M, Neijts R, van de Ven C, Beck F, Deschamps J. 2012. Evolutionarily conserved requirement of Cdx for post-occipital tissue emergence. *Development* **139**:2576–2583

Verrier L, Davidson L, Gierliński M, Dady A, Storey KG. 2018. Neural differentiation, selection and transcriptomic profiling of human neuromesodermal progenitor-like cells in vitro. *Development* **145**:dev166215

Walther C, Gruss P. 1991. Pax-6, a murine paired box gene, is expressed in the developing CNS. *Development* **113**:1435–1449

Wilson V, Olivera-Martínez I, Storey KG. 2009. Stem cells, signals and vertebrate body axis extension. *Development* **136**:1591–1604

Wymeersch FJ, Huang Y, Blin G, Cambray N, Wilkie R, Wong FCK, Wilson V. 2016. Position-dependent plasticity of distinct progenitor types in the primitive streak. *Elife* **5**:e10042

Yamaguchi TP, Bradley a, McMahon a P, Jones S. 1999. A Wnt5a pathway underlies outgrowth of multiple structures in the vertebrate embryo. *Development* **126**:1211–1223

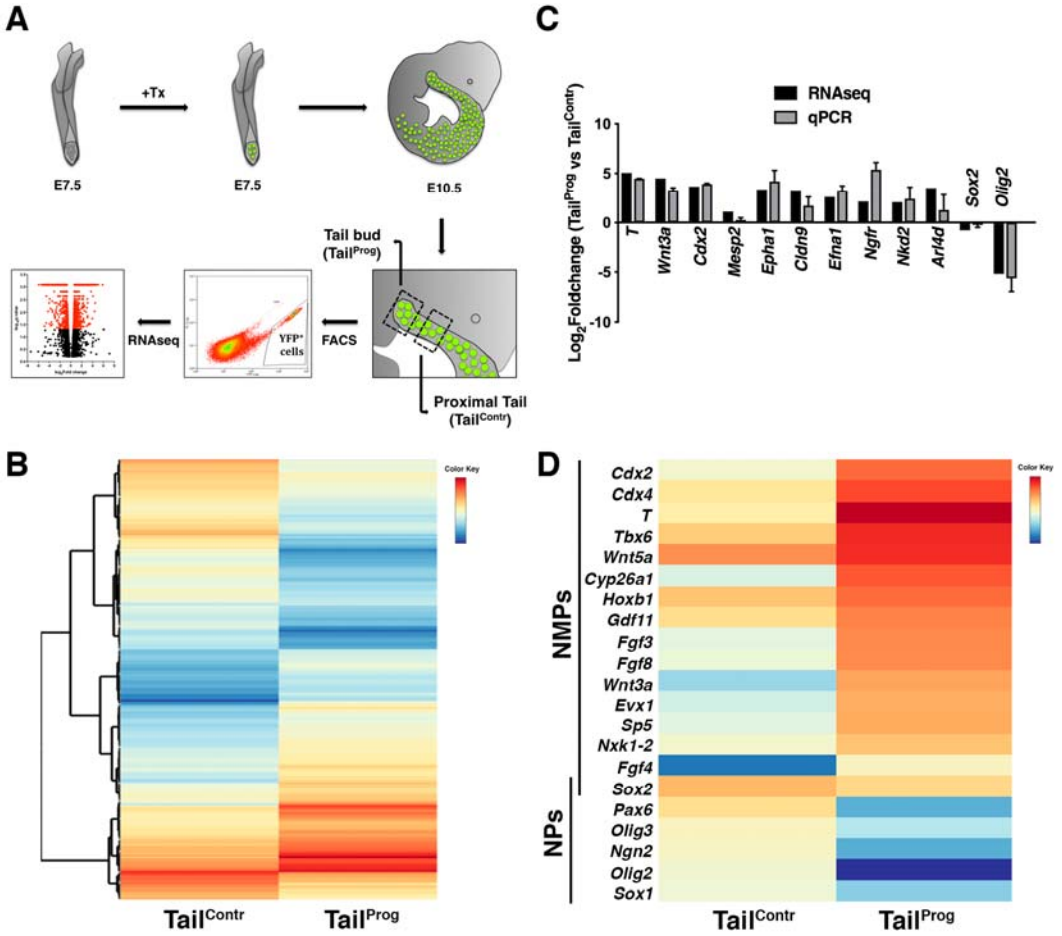

Figure 1. Labeling and Isolation of YFP-positive cells for transcriptomic analysis by RNA-seq. **A)** Strategy for isolating YFP-positive cells dissected from the tail bud and proximal tail regions of E10.5 *Cdx2P*-cre^{ERT}::*ROSA26-YFP-R* embryos, YFP-positive cells were collected by FACS for each population to perform RNA sequencing analysis (n=2). **B)** Heat map showing differentially expressed genes between Tail^{Prog} and Tail^{Contr}. **C)** Validation of RNA-seq data by RT-qPCR. Comparison of 12 differentially expressed genes values determined by RNA-seq and RT-qPCR. The error bars represent the standard deviation error of three independent replicates. **D)** Heat map displaying several differentially expressed genes associated with neuro-mesodermal progenitors (NMPs) or neural progenitors (NPs) identities between Tail^{Prog} and Tail^{Contr}. The color key bar represents the average of normalized counts on logarithmic scale. Red and blue reflect high and low gene expression levels, respectively.

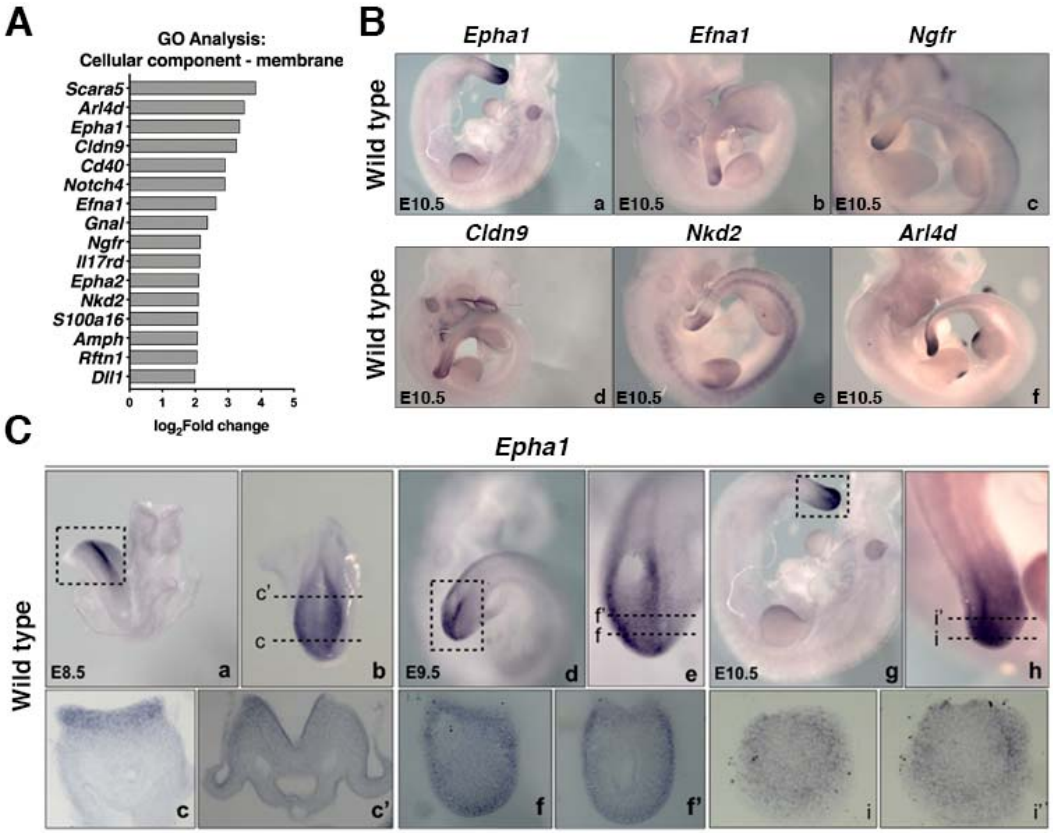

Figure 2. Identification of a cell-surface marker for axial progenitors. **A)** Differentially expressed genes that encode for plasma membrane proteins according to gene ontology (GO) classification based on cellular component. **B)** mRNA expression patterns of *Epha1*, *Efna1*, *Ngfr*, *Cldn9*, *Nkd2* and *Arl4d* in wild type E10.5 embryos by whole-mount *in situ* hybridization **C)** Expression pattern of *Epha1* in wild type embryos at E8.5 (a-c'), E9.5 (d-f') and E10.5 (g-i'). (c-c'), (f-f') and (i-i') show vibratome sections through the regions shown in b, e and h.

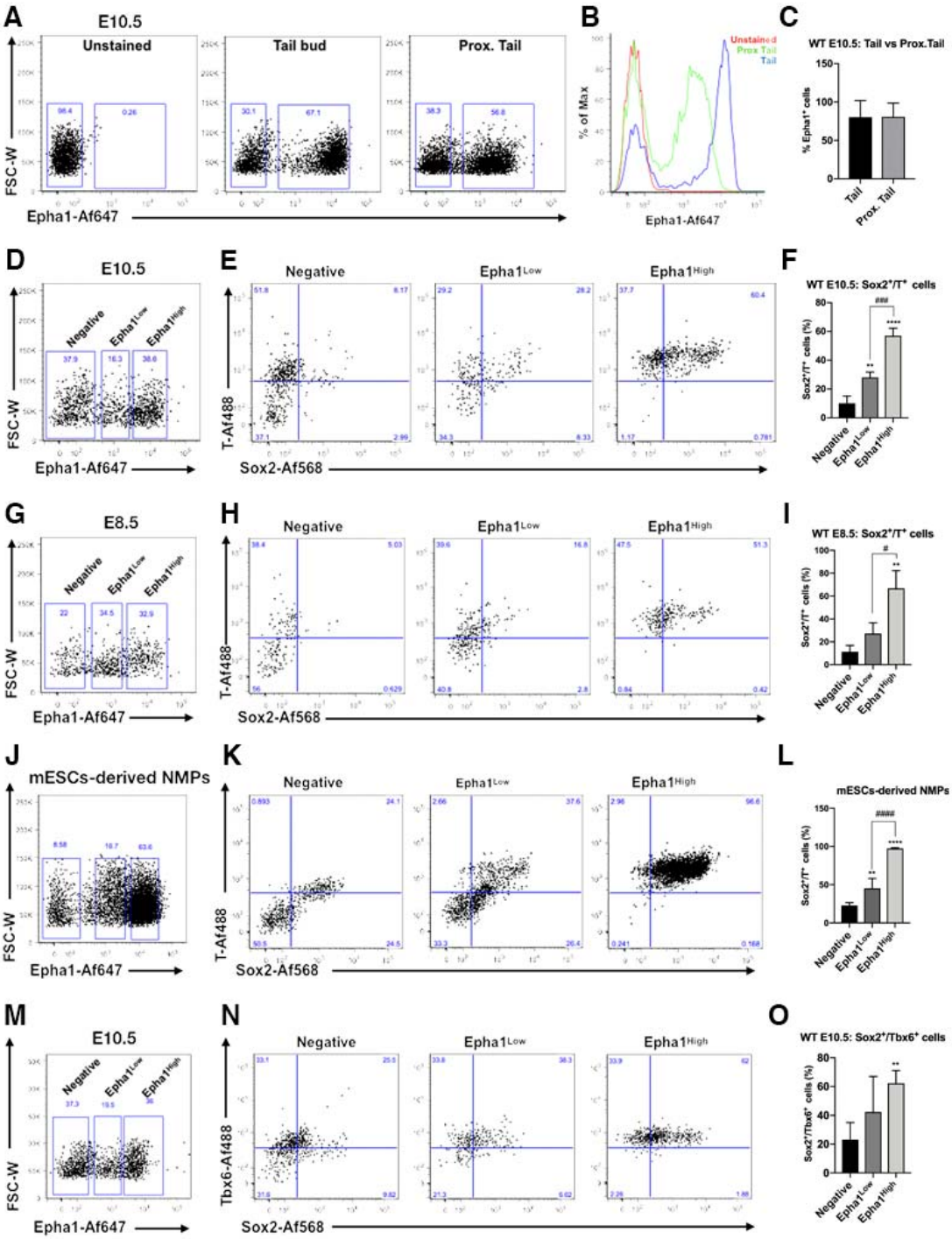

Figure 3. Flow cytometry analysis of Epha1 subpopulations. **A)** FACS analysis showing fluorescence spread into Epha1-Af647 channel of cells from the tail bud and proximal tail regions of 10.5 wild type embryos. An unstained control from the tail bud is also shown. **B)** Single parameter histogram showing the different fluorescent intensity of Epha1-positive cells from the tail bud (blue) and proximal tail (green) cells from E10.5 wild type embryos. The unstained control is shown in red. **C)** Quantification of Epha1-positive cells (%) found in the tail bud and proximal tail of E10.5 wild type embryos. **D)** Gating strategy to identify Epha1 cell subsets from the tail bud of E10.5 wild type

embryos. **E)** FACS profiles showing Sox2 and T expression in Epha1-negative, Epha1^{Low} and Epha1^{High} cell populations gated in D. **F)** Quantification of double positive cells for Sox2/T (%) between the different Epha1 cell compartments. **G)** Epha1 cell compartments from the posterior region of E8.5 wild type embryos. **H)** FACS profiles showing Sox2 and T expression in the different Epha1 cell subsets gated in G. **I)** Quantification of double positive cells for Sox2/T (%) between the different Epha1 compartments of the posterior region of E8.5 wild type embryos. **J)** Epha1-cell subsets in mESCs-derived NMPs. **K)** FACS profiles showing Sox2 and T expression in Epha1 cell populations gated in J. **L)** Quantification of double positive cells for Sox2/T (%) between the different Epha1 cell subsets from mESCs-derived NMPs. **M)** Gating strategy to identify Epha1 cell subsets from the tail bud of E10.5 wild type embryos. **N)** FACS profiles showing Sox2 and Tbx6 expression on the different Epha1 cell populations gated in M. **O)** Quantification of double positive cells for Sox2/Tbx6 (%) in the different subsets of Epha1 cells from tail bud of E10.5 wild type embryos. Quadrant averages were calculated using at least 3 independent experiments and one-way analysis of variance ANOVA was used to determine statistical significance. (**$p < 0,01$ and ****$p < 0,0001$ vs Negative; $^{##}p < 0,01$ and $^{####}p < 0,0001$ vs Epha1^{Low}). Error bars indicate the standard deviation (SD).

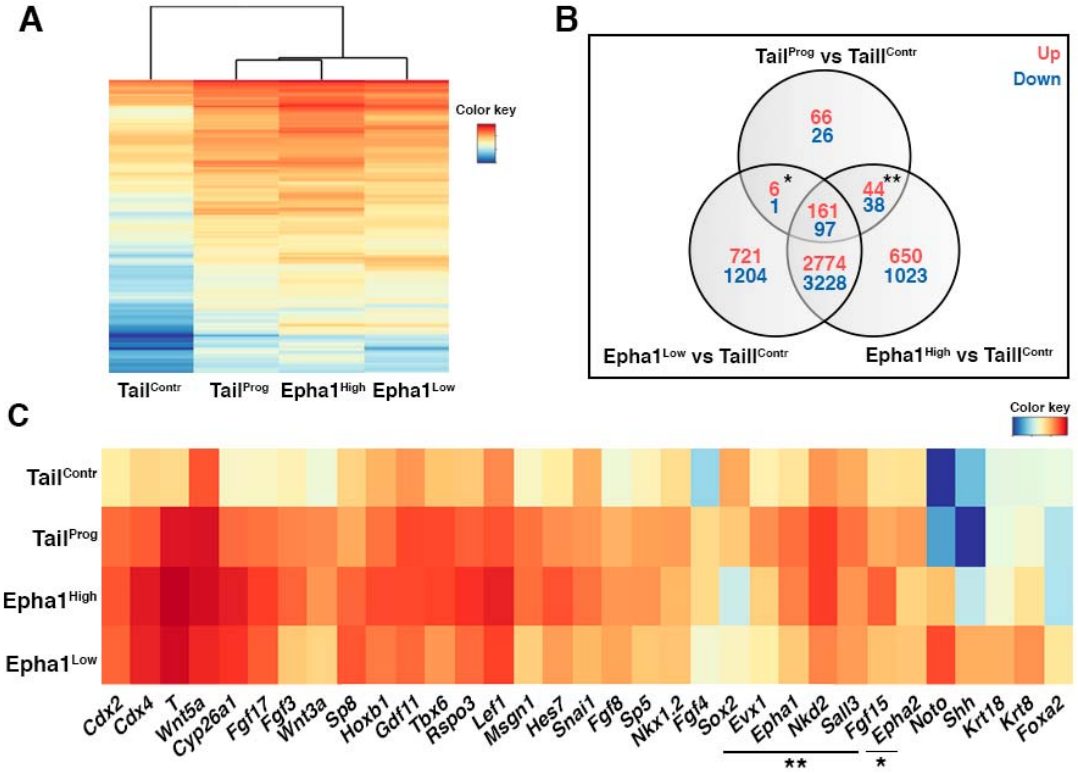

Figure 4. Transcriptome analysis of Tail^{Prog}, Epha1^{High} and Epha1^{Low} subpopulations from the tail bud of wild type E10.5 embryos. A) Heat map of common significantly differentially expressed genes between Tail^{Contr} and Tail^{Prog}, Epha1^{High}, Epha1^{Low}. Clustering analysis was derived using Euclidean distance and Ward's method. **B)** Venn diagram showing the shared and the distinct expressed genes in Tail^{Prog}, Epha1^{High} and Epha1^{Low} compared to Tail^{Contr}. Significantly up-regulated genes are shown in red while significantly down-regulated genes are shown in blue. **C**) Heat map showing relevant differentially expressed genes associated with axial progenitor, mesoderm and notochord progenitor identity amongst Tail^{Contr}, Tail^{Prog}, Epha1^{High} and Epha1^{Low}. (*) Transcript included in the differentially expressed genes group shared only by Tail^{Prog} and Epha1^{Low}; (**) Transcripts included in the differentially expressed genes group shared only by Tail^{Prog} and Epha1^{High}. The color key bar represents the average of normalized counts on logarithmic scale. Red and blue refer to high and low gene expression levels, respectively.

Table 1. FACS data of Epha1 subpopulations. Percentage of positive cells in the respective Epha1-cell compartments. The values show the mean ± SD from at least three independent experiments.

Sample	Compartment	Percentage of cells (mean±SD)			
		Sox2-/T-	**Sox2-/T+**	**Sox2+/T+**	**Sox2+/T-**
E10.5	Epha1^{Neg}	17,78±16,82	70,63±16,48	10,12±5,00	1,50±1,43
	Epha1^{Low}	27,50±5,91	36,95±10,96	28,03±3,75	3,28±4,44
	Epha1^{High}	3,53±2,09	39,33±5,05	56,93±5,33	0,26±0,45
Sample	**Compartment**	**Percentage of cells (mean±SD)**			
		Sox2-/T-	**Sox2-/T+**	**Sox2+/T+**	**Sox2+/T-**
E8.5	Epha1^{Neg}	47,87±10,36	37,47±2,53	11,31±5,45	3,40±5,35
	Epha1^{Low}	35,23±8,21	31,33±11,90	27,20±9,46	6,25±6,45
	Epha1^{High}	5,84±8,20	22,60±22,40	66,67±15,55	4,91±8,14
Sample	**Compartment**	**Percentage of cells (mean±SD)**			
		Sox2-/T-	**Sox2-/T+**	**Sox2+/T+**	**Sox2+/T-**
mESCs-derived NMPs	Epha1^{Neg}	48,85±7,76	0,91±0,43	22,90±3,65	21,93±18,02
	Epha1^{Low}	28,93±7,65	3,07±1,02	45,33±12,80	22,68±4,94
	Epha1^{High}	0,14±0,13	2,20±0,89	97,45±1,04	0,21±0,08
Sample	**Compartment**	**Percentage of cells (mean±SD)**			
		Sox2-/Tbx6-	**Sox2-/Tbx6+**	**Sox2+/Tbx6+**	**Sox2+/Tbx6-**
E10.5	Epha1^{Neg}	21,11±20,18	53,06±29,15	23,04±11,90	4,80±4,81
	Epha1^{Low}	24,27±16,25	25,83±11,22	42,30±24,68	7,73±8,10
	Epha1^{High}	10,63±9,54	24,62±6,68	62,30±8,78	2,43±1,91

Table 2. Differentially expressed genes shared by Epha1^{High}, Epha1^{Low} and Tail^{Prog} relative to Tail^{Contr}.

Shared by Epha1^{High}, Epha1^{Low} and Tail^{Prog}	up-regulated	*2610528A11Rik, 2810404F17Rik, AW551984, Ablim2, Adamts17, Adamtsl2, Aplnr, Arl4d, Avpr1b, B230323A14Rik, B3gnt7, BC064078, Bhlhe40, Blnk, Btg2, Car3, Cdh3, Cdkn1a, Cdx2, Cdx4, Cited1, Cldn9, Cpa6, Cyp26a1, Cyp26c1, D230030E09Rik, Dmgdh, Dmtn, Drd1, Dusp6, Dync1i1, Efna1, Elf3, Endod1, Eogt, Epor, Erap1, Erich2, Ets2, Etv5, Evx1os, Ezr, Fabp7, Fam163b, Fermt1, Fgf17, Fgf3, Fgf8, Fgf9, Foxn1, Fstl4, Gad1, Gdf11, Gm11497, Gm16702, Gm26793, Gm6667, Gm9894, Gna14, Gnal, Gpr50, Gpr83, Hes7, Hottip, Hoxa11os, Hoxaas3, Hoxb1, Hoxc10, Hoxc13, Hoxc9, Ifitm1, Igsf21, Il17ra, Irf1, Irgm1, Itgb2l, Itgb8, Kcnab3, Kcnh5, Kitl, Klhl32, Krt19, Krt39, Krt82, Lamc2, Laptm4b, Lef1, Lefty1, Lhpp, Lmo2, Lypd6b, Lzts1, Maff, Mapkapk3, Mixl1, Mmp2, Mnx1, Mogat2, Msgn1, Myadm, Naaa, Ndrg2, Ngfr, Nkd1, Nkx1-2, Nptx2, Oasl2, Otop1, Oxtr, Pcolce2, Pcp4l1, Pde1c, Pde4b, Pea15a, Pigv, Pla2g7, Pmaip1, Pmepa1, Prickle1, Prlr, Psmb8, Rab19, Rassf9, Rell1, Rftn1, Ripk4, Rln1, Rrad, Rspo3, S100a16, Scara5, Sct, Slc16a2, Slc2a3, Slc7a7, Smc6, Smtnl2, Sp5, Spry2, Spry4, Sulf1, Sv2b, Synj2, T, Tbx6, Tcea3, Tgfb1i1, Tmprss2, Tnfrsf12a, Tpm1, Trim25, Trim36, Trip10, Tubb6, Uap1l1, Vil1, Wls, Wnt10a, Wnt3a, Wnt5a, Zic3*
	down-regulated	*A830082K12Rik, Abca9, Adamts12, Adgra2, Ahnak, Arg1, Btbd17, Ccnd2, Cdh11, Cdkn1c, Ckb, Clmp, Col1a1, Col1a2, Col2a1, Col3a1, Ddr2, Des, Dnm1, Dpep1, Ebf3, Eya2, Fam181b, Fgfr2, Foxp4, Gas1, Gdf10, Gfra1, Grem, Grrp1, Igdcc3, Igf1, Igfbp5, Ildr2, Irx3, Klf3, Klhdc10, Lsp1, Marcks, Mdfi, Meis1, Mest, Msx3, Mybpc1, Nedd9, Nell2, Neurod4, Neurog1, Neurog2, Nkx2-9, Nkx3-1, Nkx6-1, Nr2f1, Nrcam, Nrg1, Olig2, Olig3, Osr1, Pag1, Pax6, Pcdh18, Pcolce, Peg3, Pgf, Pgm5, Pik3r1, Pkdcc, Plch1, Pou3f2, Ppfibp2, Prdm12, Ptprd, Ptx3, Rdh10, Rfx4, Scrn1, Scube2, Scube3, Sec24d, Serpinf1, Sfrp1, Sfrp2, Sncaip, Sox1, Sox10, Sox5, Spon1, Synpo, Tbx3, Tfap2b, Tgfbi, Thsd4, Tle1, Tns3, Tshz1, Uaca,Unc5c*
Shared by Epha1^{High} and Tail^{Prog}	up-regulated	*4933439K11Rik, Amph, Car13, Cdhr1, Chst7, Dgkg, Dll1, Dll3, Dusp4, Edar, Egln3, Epha1, Erfe, Etv4, Evx1, Fgf15, Gm45889, Hoxa7, Hoxc6, Hoxd12, Lemd1, Mpzl2, Mylip, Myom2, Nfatc1, Nkd2, Otx1, Peli3, Pik3c2b, Ppfibp1, Ptk7, Qsox1, Rnf43, Rora, S1pr5, Sall3, Sema6a, Shisa6, Stra6, Tgm3, Trim9, Tshr, Vwa2, Zyx*
	down-regulated	*Apln, Asb4, Bmp3, Ccdc160, Colec12, Cpm, Dmd, Dnm3, Dse, Eno3, Epha7, Ets1, Fam181a, Fez1, Fli1, Fmod, Foxd2os, Foxp2, Fstl1, Hpgd, Id2, Moxd1, Ncald, Net1, Nr2f2, Nrp1, Pdgfc, Prkce, Rab3il1, Selenop, Slc38a4, Slc6a13, Snai2, Sox6, Sox9, Thbs1, Tmem100, Zfhx3*
Shared by Epha1^{Low} and Tail^{Prog}	up-regulated	*Dnah11, Epha2, Gm10800, Pax2, Slit2, Susd4*
	down-regulated	*Gypc*

Table 3. List of primers used for genotyping

Gene		Sequence 5' to 3'
Cre	Forward	CGAGTGATGAGGTTCGCAAG
	Reverse	CACCAGCTTGCATGATCT
YFP wild type Allele	Forward	CTGGCTTCTGAGGACCG
	Reverse	CAGGACAACGCCCACACA
YFP mutant Allele	Forward	AGGGCGAGGAGCTGTTCA
	Reverse	TGAAGTCGATGCCCTTCAG

Table 4. List of primers used for RT-qPCR

Gene		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
T(Bra)	Forward	ACCCAGCTCTAAGGAACCAC
	Reverse	GCTGGCGTTATGACTCACAG
Wnt3a	Forward	ATTGAATTTGGAGGAATGGT
	Reverse	CTTGAAGTACGTGTAACGTG

Table 5. List of primers used for *in situ* hybridization probes

Gene		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