Gene Regulatory Networks of Epidermal and Neural Fate Choice in a Chordate

Anthony Leon, Lucie Subirana, Kevin Magre, Ildefonso Cases, Juan Tena, Manuel Irimia, Jose Luis Gomez-Skarmeta, Hector Escriva, Stephanie Bertrand

Gene Regulatory Networks of Epidermal and Neural Fate Choice in a Chordate

Anthony Leon,[1] Lucie Subirana,[1] Kevin Magre,[1] Ildefonso Cases,[2] Juan J. Tena,[2] Manuel Irimia [ID],[3,4,5] Jose Luis Gomez-Skarmeta,[2] Hector Escriva,[*,1] and Stéphanie Bertrand [ID][*,1]

[1]Sorbonne Université, CNRS, Biologie Intégrative des Organismes Marins, BIOM, Laboratoire Arago, Avenue Pierre Fabre, F-66650 Banyuls-sur-Mer, France

[2]Centro Andaluz de Biología del Desarrollo (CABD), CSIC-Universidad Pablo de Olavide-Junta de Andalucía, Sevilla, Spain

[3]Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Barcelona, Spain

[4]Universitat Pompeu Fabra (UPF), Barcelona 08003, Spain

[5]ICREA, Barcelona, Spain

*Corresponding authors: E-mails: stephanie.bertrand@obs-banyuls.fr; hescriva@obs-banyuls.fr.

Associate editor: Li Liu

Abstract

Neurons are a highly specialized cell type only found in metazoans. They can be scattered throughout the body or grouped together, forming ganglia or nerve cords. During embryogenesis, centralized nervous systems develop from the ectoderm, which also forms the epidermis. How pluripotent ectodermal cells are directed toward neural or epidermal fates, and to which extent this process is shared among different animal lineages, are still open questions. Here, by using micromere explants, we were able to define in silico the putative gene regulatory networks (GRNs) underlying the first steps of the epidermis and the central nervous system formation in the cephalochordate amphioxus. We propose that although the signal triggering neural induction in amphioxus (i.e., Nodal) is different from vertebrates, the main transcription factors implicated in this process are conserved. Moreover, our data reveal that transcription factors of the neural program seem to not only activate neural genes but also to potentially have direct inputs into the epidermal GRN, suggesting that the Nodal signal might also contribute to neural fate commitment by repressing the epidermal program. Our functional data on whole embryos support this result and highlight the complex interactions among the transcription factors activated by the signaling pathways that drive ectodermal cell fate choice in chordates.

Key words: EvoDevo, gene regulatory network, cell fate, amphioxus, neural induction.

Introduction

The ectoderm of vertebrates gives rise to both the neuroectoderm, which later forms the central nervous system (CNS), and the epidermis, which constitutes an epithelial barrier protecting the embryo from the environment (Gilbert 2000). How the ectodermal cells are directed toward these two fates is a major question in developmental biology. In the default model for neural induction in vertebrates, ectodermal cells become neural by default, whereas they acquire epidermal fate after bone morphogenetic protein (BMP) signal activation (Hemmati-Brivanlou and Melton 1997). In the whole embryo, the ventral region of the ectoderm becomes epidermal because it receives BMP signals (De Robertis and Kuroda 2004; Stern 2006; Ozair et al. 2013). In the dorsal region, the organizer produces molecules antagonizing this BMP signal (such as Noggin, Chordin, or Follistatin), allowing the dorsal ectoderm to acquire a neural fate (De Robertis and Kuroda 2004). However, some other instructive signals might be required depending on the species (Streit et al. 2000;

Wilson et al. 2000; Linker and Stern 2004; Delaune et al. 2005), and the gene regulatory networks (GRNs) switched on by these signals are still not completely unraveled. Important insights into this question have been brought by genome-wide approaches applied to cultivated stem cells (embryonic or induced) (Sankar et al. 2016; Omrani et al. 2018; Li et al. 2019), but such a system still does not allow to apprehend the full picture of the process in the whole embryo.

Vertebrates belong to the chordate phylum, which also includes their sister group, the tunicates (Delsuc et al. 2006), and the cephalochordates clade (Bourlat et al. 2006) (i.e., lancelet or amphioxus). Contrary to vertebrates, tunicate and cephalochordate epidermis have neurogenic potential and form the so-called epidermal sensory cells (Torrence and Cloney 1982; Wicht and Lacalli 2005). In cephalochordates, induction of the neural plate is triggered as in vertebrates by a dorsal organizer that is absent in tunicates (Garcia-Fernandez et al. 2007; Yu et al. 2007; Le Petillon et al. 2017). By using micromere/ectodermal

explants and graft experiments combined with pharmacological treatments, we previously showed that this organizer acts through the activation of the Nodal/activin signaling pathway and that BMP signal inhibition is not sufficient to trigger neural induction (Le Petillon et al. 2017). Consistently, micromere explants in amphioxus stay in an undifferentiated state when BMP signaling is blocked, develop into epidermis when raised in seawater, and form neural plate tissue when treated with activin. Here, we took advantage of this explant-based system to decipher the putative GRNs underlying ectodermal cell specification in cephalochordates.

Results and Discussion

Neural Induction Potentially Implies Both the Activation of the Neural GRN and the Repression of the Epidermal GRN

We previously showed that amphioxus micromere explants, produced by microdissection at the eight-cell stage, form epidermis when raised in seawater, whereas they develop into neural tissue when exposed to activin. In order to analyze the gene expression and gene expression regulation at a global scale during the process of the epidermis and nervous system development, we generated RNA-seq and ATAC-seq data from such explants collected at the time at which the whole embryo reaches the gastrula stage (G5 stage, 11 h postfertilization [hpf] at 19°C), or at the time at which the whole embryo reaches the neurula stage (N2 stage, 19 hpf at 19°C).

Differentially expressed genes with higher expression in activin-treated explants ($N = 878$, Log2(foldchange) > 2, Padj < 0.05, supplementary table S1, Supplementary Material online) at the G5 stage include Nodal/TGFβ/BMP signaling pathway genes (Vg1, Chordin, Cerberus, Lefty, Pitx, Ski), typical proneural genes (Elav, Neurogenin, Insm) and orthologues of genes coding for proteins implicated in axon guidance or neuronal migration (Sema6A, Auts2, Slit1) and on the other hand, differentially expressed genes with higher expression in untreated explants ($N = 3,059$, |Log2(foldchange)| > 2, Padj < 0.05, supplementary table S1, Supplementary Material online) comprise orthologues of genes coding for proteins associated with the cytoskeleton, like two Keratin, as well as genes orthologous to Vim, ActB, or Tekt3 and orthologues of Bin1, Eml1, Ky, Lasp1, Mid1, Speg, and Tpm3. Tfap2 (AP2) and Fhl2 (Dral), which are transcription factor/cofactor genes expressed specifically in the presumptive epidermis of amphioxus embryos (Schubert et al. 1998; Meulemans and Bronner-Fraser 2002), are also highly overexpressed in untreated explants. The unique amphioxus Dlx gene, which is expressed in the presumptive ectoderm before it touches the mesendoderm at the beginning of gastrulation, and whose expression gets restricted to the presumptive epidermis when both germ layers contact each other (Holland et al. 1996), was also overexpressed in untreated explants.

We used ATAC-seq to define accessible chromatin regions representing potential regulatory regions. Among the more

accessible regions ($N = 1,237$, Padj < 0.05, supplementary table S2, Supplementary Material online) in activin-treated explants at G5, 218 correspond to peaks associated with upregulated genes and 36 to peaks in the vicinity of genes underexpressed compared with untreated explants. Similarly, among the regions more accessible in control embryos ($N = 2,658$, Padj < 0.05, supplementary table S2, Supplementary Material online), 1,299 were associated with the genes overexpressed in untreated explants, and 29 with genes that are more expressed in activin-treated explants. Overall, the majority of more accessible peaks hence correspond to putative regulatory regions of genes that are overexpressed or for which there is no significant expression difference between activin-treated and untreated explants.

Combining the RNA-seq and ATAC-seq data, we selected the following for further analysis: 1) genes upregulated in activin-treated explants showing nearby genomic regions that were more accessible in activin-treated explants, hereafter called "all neural (G5)" genes ($N = 158$) and 2) genes upregulated in control explants showing nearby genomic regions that were more accessible in untreated explants, hereafter called "all epidermal (G5)" genes ($N = 942$) (fig. 1).

Gene Ontology (GO) term enrichment of "all neural (G5)" genes shows that activin treatment affected several signaling pathways and the expression of transcriptional regulators (fig. 2a). Moreover, activating the Nodal/activin pathway clearly directs the cells toward the neural program (fig. 2a). On the other hand, analysis of "all epidermal (G5)" genes shows that they are enriched in terms associated with metabolism and exocytosis, suggesting that control explants are more advanced in the differentiation process compared with activin-treated explants (fig. 2b). The enrichment in terms associated with chromatin modification in both groups of genes indicates that neural and epidermal fate acquisition are linked to different genomic structural remodeling processes. Altogether, these data are consistent with in situ hybridization assays of marker genes and confirm that activin treatment induces ectodermal pluripotent cell commitment to neural fate, whereas untreated ectodermal explants develop into epidermis (Le Petillon et al. 2017).

In order to understand the regulatory logic downstream of Nodal, we searched for putative transcription factor binding sites (TFBS) in the differentially accessible genomic regions of "all neural (G5)" and "all epidermal (G5)" genes (fig. 1). We first used HOMER to find enriched motifs that were then compared with known TFBS in vertebrates. Although we used a vertebrate database, it has been shown that transcription factor (TF) binding specificities are highly conserved in animals, even between species as evolutionarily distant as human and Drosophila (Nitta et al. 2015). We found that for the "all neural (G5)" genes, the more significantly enriched motifs, which are also the more abundant, correspond to putative binding sites for transcription factors of the SOX family (supplementary table S3 and fig. S1, Supplementary Material online). On the other hand, we observed an enrichment in the "all epidermal (G5)" peaks in
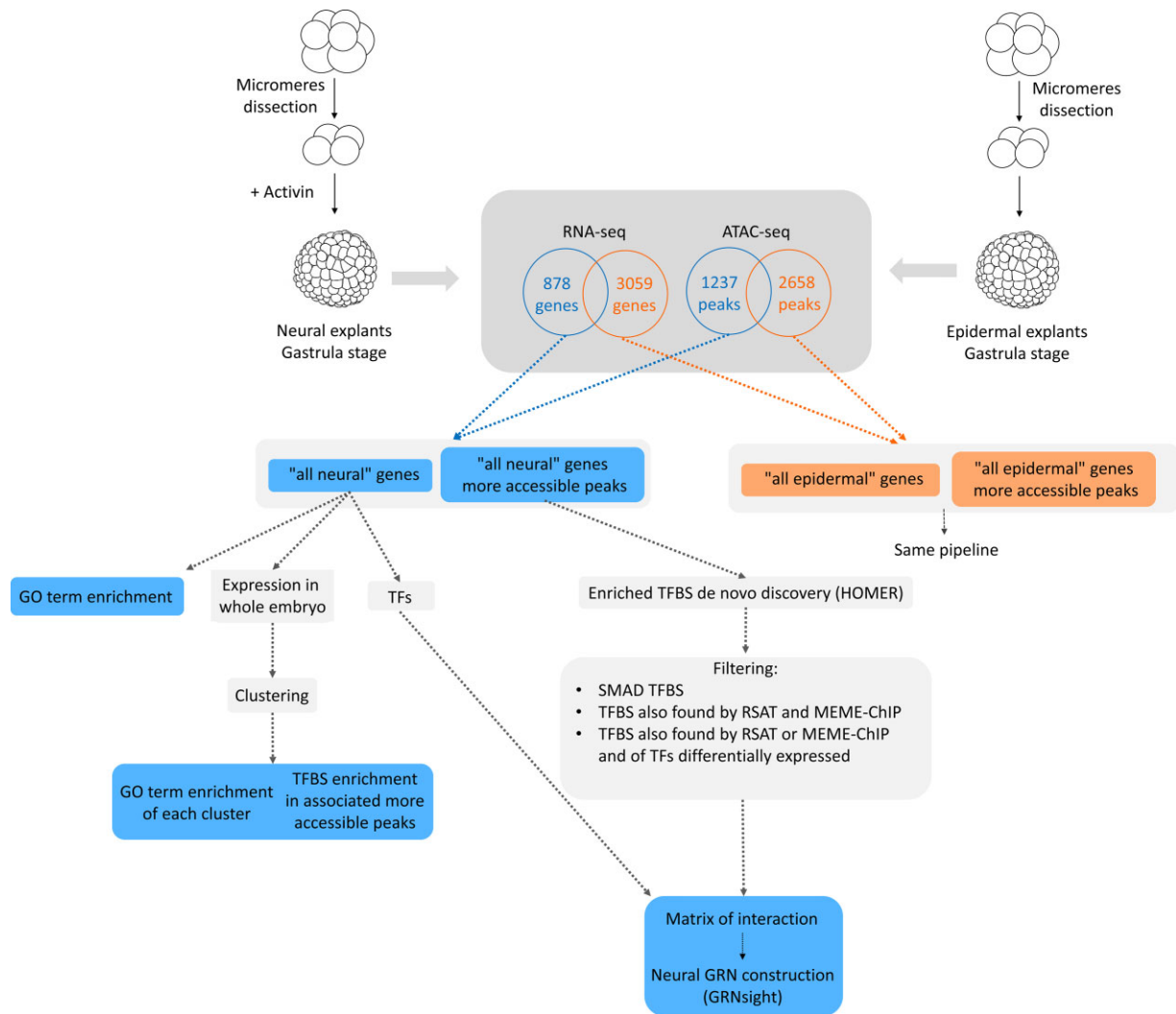
Fig. 1. Pipeline used for parallel analysis of ATAC-seq and RNA-seq data and for the putative GRN construction. The different steps presented correspond to the analysis and putative GRN construction at the G5 stage. The numbers in the circles correspond to differentially expressed genes (RNA-seq, blue for gene overexpressed in activin-treated explants, orange for genes overexpressed in untreated explants) or to differentially accessible chromatin regions (ATAC-seq, blue for regions more accessible in activin-treated explants, orange for regions more accessible in untreated explants). The detailed pipeline is presented on the left for the "all neural (G5)" genes (genes overexpressed in activin-treated explants and showing more accessible peaks in activin-treated explants).

potential binding sites for GRHL, SMAD, KLF, and DLX families as the most significant (supplementary table S4 and fig. S2, Supplementary Material online). In parallel, we also ran RSAT and MEME-ChIP in order to define more robustly the putative enriched TFBS (supplementary tables S5–S10, Supplementary Material online). We then used these data to generate an in silico putative GRN for "all neural (G5)" and "all epidermal (G5)" genes (see Materials and Methods and fig. 1) by constructing a matrix of interaction between TF belonging to the "all neural (G5)" or "all epidermal (G5)" gene sets and TF with potential TFBS in the differentially accessible regions associated to these genes.

In the neural putative GRN, the Nodal pathway coeffector Smad4 controls most of the genes, among which *SoxB1*. *SoxB1* and *Pou3* are placed just downstream in the cascade, whereas *Neurogenin* is placed below, receiving inputs from *SoxB1*, *SoxC*, *SoxD*, *SoxE*, and *FoxA* but not directly from *Smad4* (fig. 2c). Most of the downstream genes are genes

expressed in the amphioxus organizer (*FoxA*, *Gli*, *Gsc*, *Vent*, *Zic*) (Shimeld 1997; Neidert et al. 2000; Kozmik et al. 2001; Gostling and Shimeld 2003; Shimeld et al. 2007), or in the presumptive neural plate (*Snail*, *Otx*, *Nkx2-1*, *Nkx6*) (Williams and Holland 1996; Langeland et al. 1998; Venkatesh et al. 1999) at gastrula stages. SOXB1 family members are known in metazoans as playing a major role during nervous system development. In the cnidarian *Nematostella*, *SoxB(2)*, which is orthologous to SOXB1 and SOXB2 family members in bilaterians, is a marker of neural progenitor cells (Richards and Rentzsch 2014). In *Drosophila*, *SoxNeuro* controls both the specification of neuroblasts and the differentiation of neural and glial cells (Ferrero et al. 2014). In the vertebrate *Xenopus*, overexpression studies in ectodermal explants showed that *Sox1*, *Sox2*, and *Sox3* are sufficient for neural induction and that *Sox2* cooperates with *Oct91* (also known as *Oct4* or *Pou5f1*) to maintain a neural progenitor identity
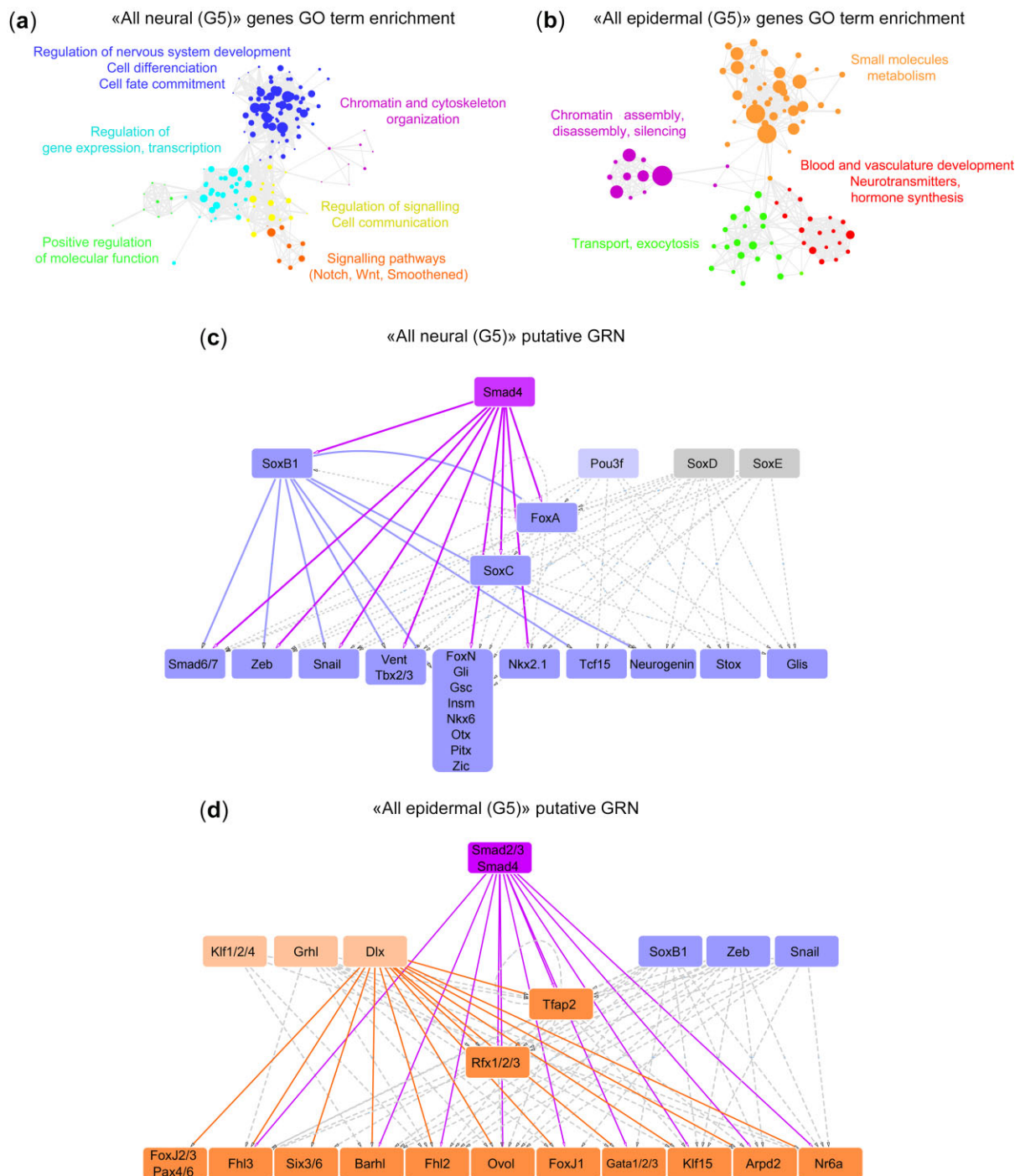
(a) «All neural (G5)» genes GO term enrichment

(b) «All epidermal (G5)» genes GO term enrichment

(c) «All neural (G5)» putative GRN

(d) «All epidermal (G5)» putative GRN

FIG. 2. GO term enrichment and putative GRN for "all neural (G5)" and "all epidermal (G5)" genes. (a) GO term enrichment visualization for the "all neural (G5)" gene list. (b) GO term enrichment visualization for the "all epidermal (G5)" gene list. For (a) and (b), GO terms that were not connected to others are not represented. (c) Putative GRN constructed using the "all neural (G5)" gene RNA-seq and ATAC-seq data. (d) Putative GRN constructed using the "all epidermal (G5)" gene RNA-seq and ATAC-seq data. In (c) and (d), the SMAD transcription factors are in purple, the transcription factors that belong to the "all neural (G5)" genes are in blue, the transcription factors that are only overexpressed in activin-treated explants are in light blue, the transcription factors whose expression is not different between untreated and activin-treated explants are in gray, the transcription factors that belong to the "all epidermal (G5)" genes are in orange, the transcription factors that are only overexpressed in untreated explants are in light orange. Edges between SMAD, SoxB1, Dlx, and the other nodes are highlighted in color, the others are in gray. Transcription factors that have the same combination of edges are grouped into unique nodes.

(Archer et al. 2011). However, SOXB1 members are also recognized as factors implicated in pluripotency maintenance or induction. Hence, Sox2 was described as a factor able to transform somatic cells into induced

pluripotent stem cells in vertebrates when combined with Oct4 (Pou5f1), Klf4, and Myc (Takahashi and Yamanaka 2006). The position of SoxB1 at the top of the neural putative GRN at G5 stage indicates that it could be

the main TF controlling the early phase of neural development in amphioxus, although it might also be implicated in maintaining the presumptive neural plate cells in a progenitor state.

In the epidermal putative GRN, Smad2/3-Smad4 are predicted to regulate the most downstream genes. *Grhl*, *Klf1/2/4*, and *Dlx*, which are orthologues of proepidermal genes in vertebrates (Dirksen et al. 1994; Pera et al. 1999; Segre et al. 1999; Hwang et al. 2011; Qiao et al. 2012; Kotkamp et al. 2014; Miles et al. 2017; Li et al. 2019), are placed at the top of the cascade. *Tfap2* (AP2), which is orthologous to *Tfap2c* known to encode an epidermal induction factor (Qiao et al. 2012; Li et al. 2019), is placed below. Among the downstream TFs, *Rfx1/2/4* and *FoxJ1* are orthologues of known ciliogenesis regulators in vertebrates (Thomas et al. 2010) consistent with the appearance of cilia in ectodermal cells of the amphioxus embryo during gastrulation. Although *Pax4/6* and *Six3/6* are referred to as typical neural genes, in amphioxus, they are expressed in the anterior epidermis together with *Arpd2* (Glardon et al. 1998; Kozmik et al. 2007; Albuixech-Crespo et al. 2017), and their presence in the epidermal putative GRN might reflect the fact that the epidermal explants adopt an anterior fate, consistent with the absence of the organizer. Finally, concerning Gata1/2/3, it has been shown that Gata3 interacts with Tfap2a, Tfap2c, and Grhl2 during embryonic stem cells differentiation into committed surface ectoderm in vertebrates (Li et al. 2019). Ovol orthologues in vertebrates are also known as regulators of epidermal fate determination (Lee et al. 2014) and can drive differentiation of dermal fibroblasts into keratinocytes in combination with other epidermal factors such as Klf4 or Grhl3 (Chen et al. 2014). Moreover, Ovol factors are known to inhibit epithelial–mesenchymal transition (EMT) (Saxena et al. 2020) and neural induction (Zhang et al. 2013).

It is to be noted that some genes that are overexpressed in activin-treated explants compared with untreated explants code for TFs that have binding sites in more accessible chromatin regions of some "all epidermal (G5)" TFs. This is the case of SoxB1, Zeb, and Snail that are placed at the top of the epidermal putative GRN (fig. 2d). Interestingly, the Snail and Zeb vertebrate orthologues are known transcriptional repressors acting during EMT to repress epithelial genes (Lamouille et al. 2014), and their position in both the neural and epidermal putative GRNs suggests they might be in part acting on neural development through repression of the epidermal putative GRN in amphioxus. Altogether our data suggest that the first step of neural induction in cephalochordates involves, downstream of SMAD signaling, the same key master genes described in vertebrates but also in other bilaterians. In addition, this primary phase is predicted in silico to be potentially associated with an active repression of conserved epidermal genes by TFs of the neural program, in accordance with the fact that there are more genes overexpressed in epidermal explants that genes overexpressed after activin treatment.

To analyze later stages of neural tissue formation after neural induction, we undertook the same comparative

RNA-seq and ATAC-seq approach on the explants raised until the N2 stage. Among the overexpressed genes in activin-treated explants ($N = 1{,}850$, Log2(foldchange) > 2, Padj < 0.05), we found actors of Nodal/BMP, Wnt, and FGF (Fibroblast Growth Factor) signaling pathways (*Chrd*, *Fgf9/16/20*, *Fzd4*, *Lefty*, *Vg1*), markers of differentiated neurons (genes encoding proteins associated with synapses or neuron projections, neurotransmitter receptor genes, or axon guidance genes such as, e.g., orthologues of *Chrna6*, *Chrnb2*, *Cntnap1*, *Glra2*, *Htr4*, *Slit*) and several transcription factors known to be expressed in neural tissues in amphioxus and/or in vertebrates (supplementary table S11, Supplementary Material online). Among the latter, *Fezf* and *Otx* are markers of the cerebral vesicle in cephalochordates (i.e., anterior CNS) (Williams and Holland 1996; Irimia et al. 2010), which suggests that the Nodal signal has a double function: it first induces neural fate acquisition and then it anteriorizes the formed neural tissue, in line with previous data obtained using whole embryos (Onai et al. 2012; Le Petillon et al. 2017). The genes that are overexpressed in untreated explants compared with activin-treated explants ($N = 3{,}636$, |Log2(foldchange)| > 2, Padj < 0.05) include again keratin and intermediate filament coding genes as well as genes implicated in cilia formation and function (dynein heavy and light chain genes, orthologues of *Armc4*, *Bbs5*, *Crocc*, *Lca5*) together with genes implicated in various metabolic pathways (supplementary table S11, Supplementary Material online). Concerning the ATAC-seq data, genomic regions that are more accessible in activin-treated explants ($N = 681$) were enriched in motifs that are putative SOX family binding sites, whereas no enrichment in SMAD factor binding sites was detected (supplementary tables S12 and S13, Supplementary Material online). On the other hand, the putative regulatory regions more accessible in untreated explants ($N = 6{,}572$) were enriched in motifs that correspond to similar TFBS as observed for the gastrula stage (supplementary tables S12 and S14, Supplementary Material online). We then focused on the "all neural (N2)" ($N = 99$) and "all epidermal (N2)" genes ($N = 1{,}491$). The "all neural (N2)" genes are enriched in GO terms associated with CNS development, transcription, and signaling pathways but not in chromatin remodeling terms contrary to the earlier stage (fig. 3a). Similarly, "all epidermal (N2)" genes are enriched in terms related to metabolism, transport, and development, but not in chromatin silencing associated terms (fig. 3b).

To reconstruct the neurula stage putative GRNs, we used the same strategy as above, but we also included genes with predicted TFBS that were part of the gastrula stage putative GRNs (supplementary tables S15–S20, Supplementary Material online). In the neural putative GRN, Neurogenin and SoxB1 are placed at the top of the cascade and have putative binding sites in almost all the ATAC-seq peak regions of the downstream nodes (fig. 3c). In the epidermal putative GRN, we found at the top of the cascade Dlx as the master regulator of most "all epidermal (N2)" TF genes. Interestingly, although SMAD is not inferred to play an important role in activating the GRN downstream of Nodal signaling at this
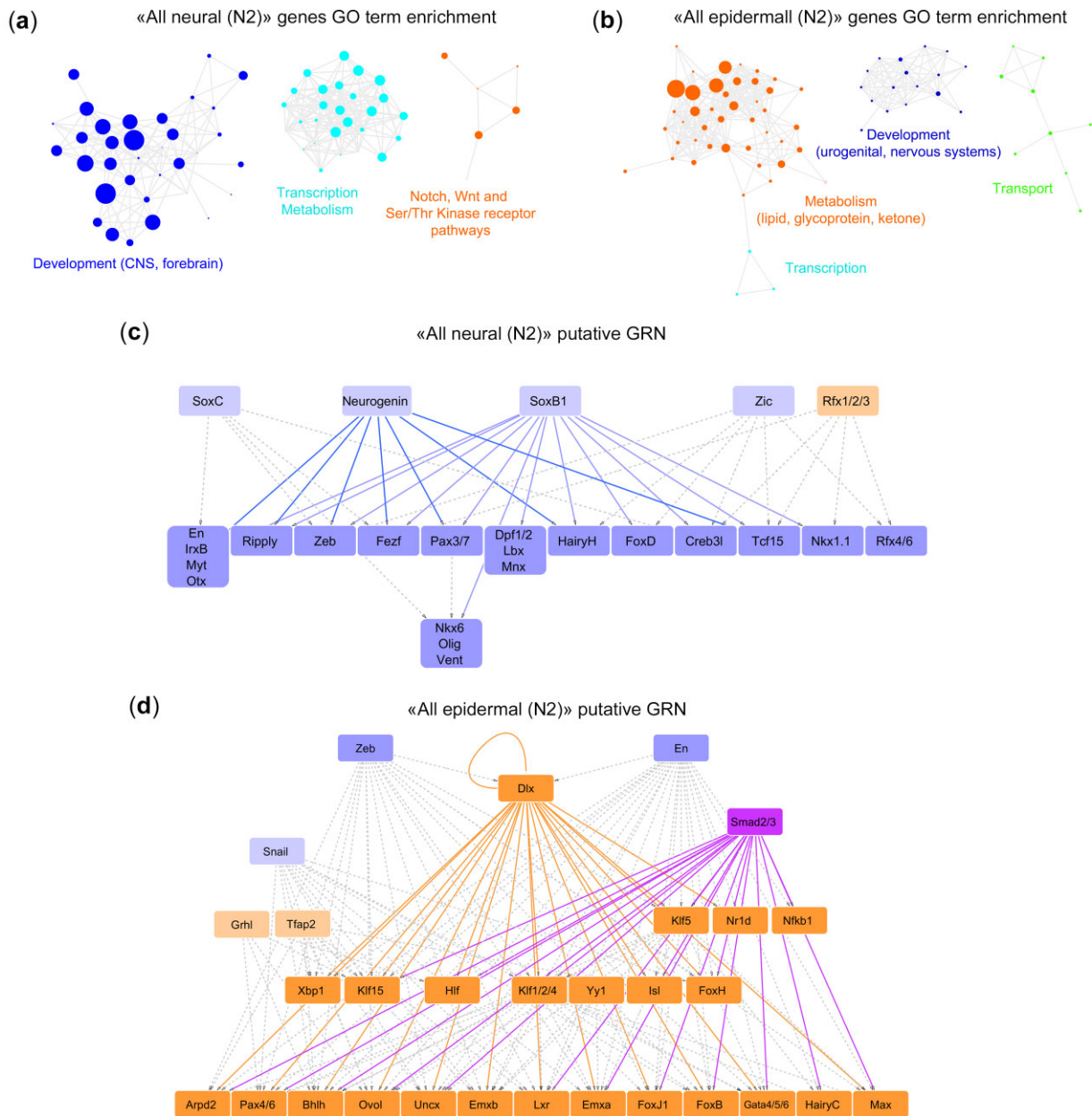
Fig. 3. GO term enrichment and putative GRN for "all neural (N2)" and "all epidermal (N2)" genes. (a) GO term enrichment visualization for the "all neural (N2)" gene list. (b) GO term enrichment visualization for the "all epidermal (N2)" gene list. For (a) and (b), GO terms that were not connected to others are not presented. (c) Putative GRN constructed using the "all neural (N2)" gene RNA-seq and ATAC-seq data. (d) Putative GRN constructed using the "all epidermal (N2)" gene RNA-seq and ATAC-seq data. In (c) and (d), edges between Smad2/3, SoxB1a, Neurogenin, Dlx, and the other nodes are highlighted in color, the others are in gray. Color code is as in fig. 2.

stage, it seems to still be an actor of the epidermal GRN. Moreover, the "all neural" TFs Zeb, Snail, and En were inferred to also be important repressors of epidermal genes at this later stage (fig. 3d), consistent with their known transcriptional repressor activity in vertebrates (Hanks et al. 1998; Lamouille et al. 2014).

Validation in the Whole Embryo of the Constructed Putative GRNs

In order to better characterize the biological significance of the putative GRNs defined above on a time scale and at the level of the whole embryo, we studied the dynamics of gene expression through normal embryonic development. We clustered the "all neural (G5)" genes based on normalized expression data previously obtained in whole embryos (Marletaz et al. 2018) from the fertilized egg stage to the late neurula stage (N5, 27 hpf at 19°C). For the genes belonging to each profile, we analyzed the GO term enrichment and the enrichment in TFBS of the corresponding ATAC-seq peak regions showing more accessibility in activin-treated explants (fig. 4). We observed that the only "all neural (G5)" gene cluster with SMAD TFBS enrichment corresponds to genes with a maximum expression at

the G3 stage, which is enriched in GO terms associated with transcription regulation. The genes belonging to the three clusters with a maximum expression at later stages are enriched in GO terms associated with CNS development, and the associated ATAC-seq peak regions are enriched in TFBS for SOX family TFs, which are downstream of SMAD signal in the neural gastrula putative GRN. These data suggest that the putative GRN constructed using ectodermal explants data is mirrored at the level of the whole embryo, with neural induction downstream of Nodal signaling pathway starting at the beginning of gastrulation, when the ectoderm first touches the mesendoderm.

We hence analyzed the expression pattern of some members of the epidermal and neural putative gastrula GRNs by in situ hybridization in embryos from the eight-cell stage up to the N4 stage (fig. 5). All the epidermal factors (*Dlx, Fhl2, FoxJ1, Grhl, Klf1/2/4, Nr6a, Ovol, Rfx1/2/3,* and *Tfap2*) are expressed in the presumptive epidermis region at G4 and G5 stages, consistent with our putative GRN. *Dlx, Fhl2,* and *Tfap2* maintain an epidermal

expression until the N4 stage. On the other hand, *Nr6a* becomes expressed in the cerebral vesicle (fig. 5, arrowhead) and the endoderm during neurulation. The expression of *FoxJ1, Grhl, Ovol,* and *Rfx1/2/3* fades in the epidermis after gastrulation. *Klf1/2/4* expression is then detected in the anterior endoderm at the N4 stage (fig. 5, arrowhead), whereas at this stage. *Rfx1/2/3* is expressed in the anterior CNS (fig. 5, arrowhead). Maternal *Rfx1/2/3* expression is also detected by in situ hybridization in the whole embryo from the eight-cell stage. From G0 to G2, the zygotic expression is restricted to the presumptive ectoderm, as also observed for *Dlx* and *Ovol*. Concerning the neural genes, *Neurogenin, Pou3fl, Snail, Stox, SoxB1a,* and *Tcf15* are expressed in the presumptive neural plate at G4 and G5 with *Pou3fl* being also expressed all around the blastopore, *SoxB1a* being additionally expressed at G4 in the presumptive epidermis and *Stox* presenting an expression in the mesendoderm layer below the presumptive neural plate region. At G5, *Zeb,* whose expression is detected in the dorsal mesendoderm at G4, is expressed in both the dorsal mesendoderm and the presumptive neural plate. *Zeb* expression is then
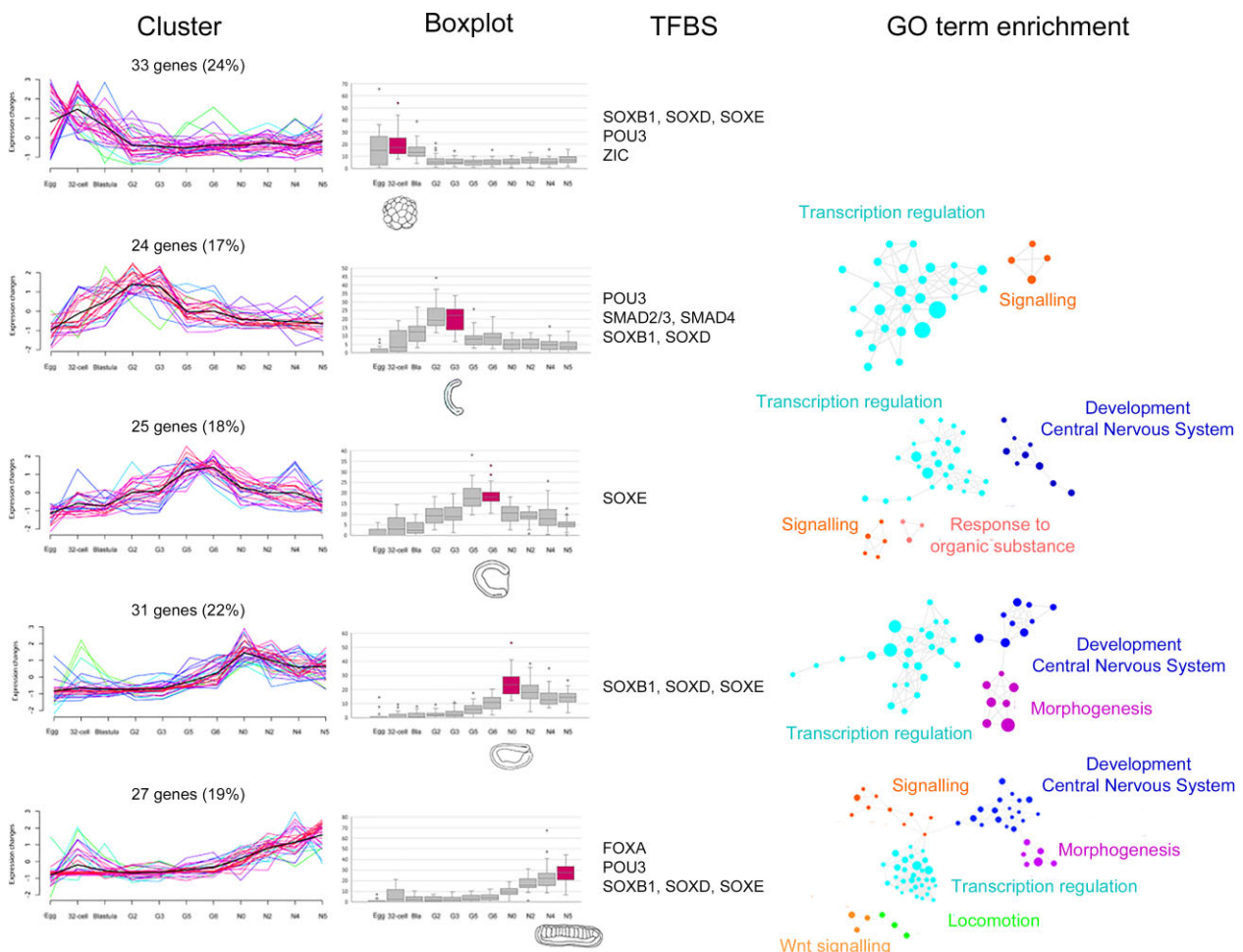
FIG. 4. Analysis of "all neural (G5)" genes expression in whole embryos. Mfuzz clustering graphs are presented on the left. The color code reflects "Membership" values calculated by Mfuzz (magenta for high values and green for low values of Membership score). Corresponding box plots (center line, mean; box limits, upper and lower quartiles; points, outliers; box in magenta = box with the highest mean expression), enrichment in TFBS of the ATAC-seq peaks with differential accessibility corresponding to the core genes of each cluster, and GO term enrichment of these core genes are presented.

FIG. 5. Expression patterns of genes of the epidermal and neural putative gastrula stage GRNs analyzed by in situ hybridization. The expression of epidermal genes (*Rfx1/2/3*, *Dlx*, *Ovol*, *Grhl*, *Klf1/2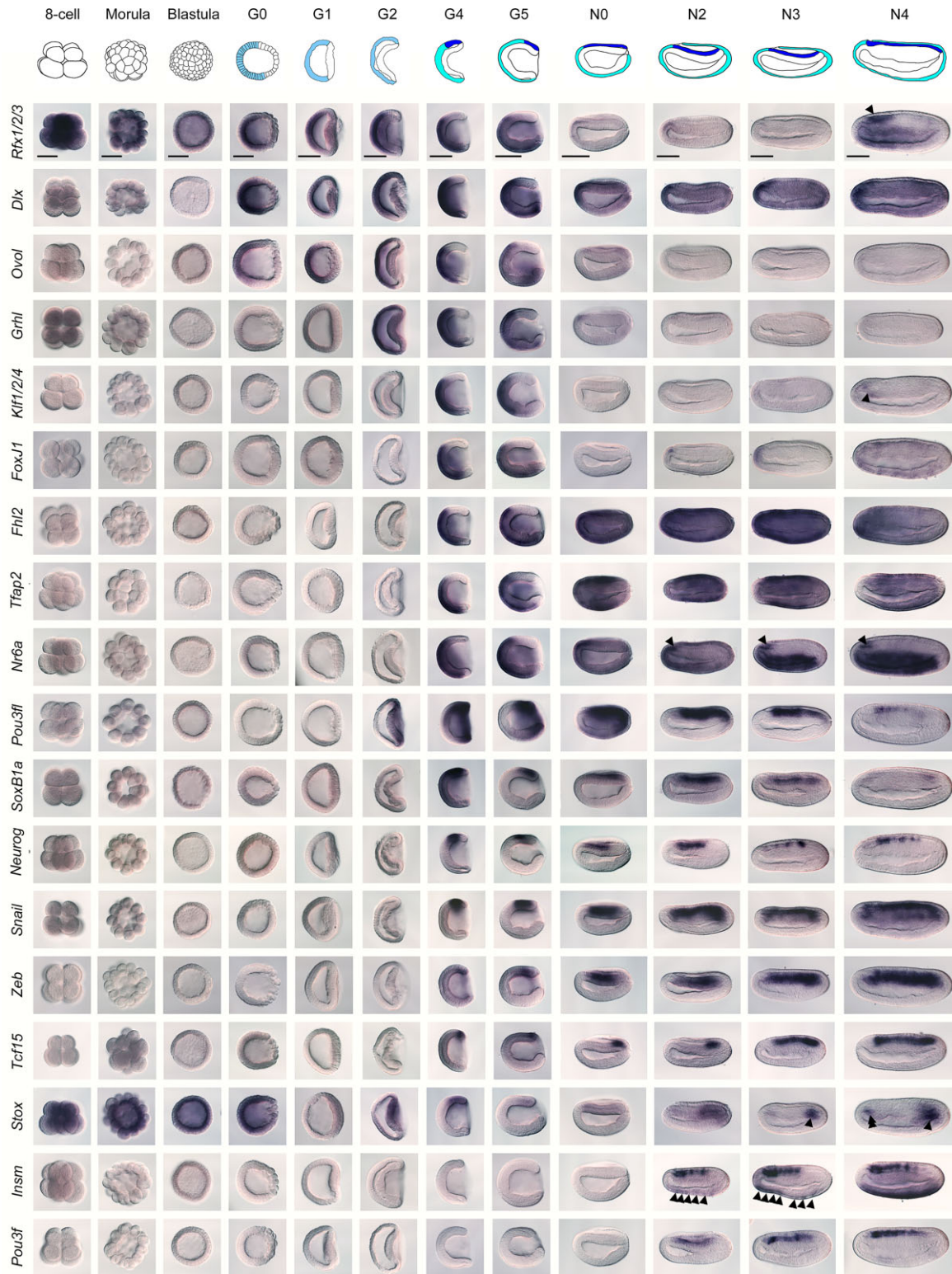/4*, *FoxJ1*, *Fhl2*, *Tfap2*) and neural genes (*Pou3fl*, *SoxB1a*, *Neurogenin*, *Snail*, *Zeb*, *Tcf15*, *Stox*, *Insm*, *Pou3f*) were analyzed at eight-cell, morula, blastula, G0, G1, G2, G4, G5, N0, N2, N3, and N4 stages. Side views are presented with anterior to the left and dorsal to the top for gastrula and neurula stages. In the drawings, the presumptive ectoderm is highlighted in light blue, the epidermis territory in turquoise, and the forming CNS in dark blue. At N4 stage, *Rfx1/2/3* is expressed in the cerebral vesicle (arrowhead) and *Klf1/2/4* in the anterior endoderm (arrowhead). *Nr6a* is expressed during neurulation in the cerebral vesicle region (arrowheads). *Stox* is expressed at N3 and N4 stages in the posterior endoderm (arrowheads) and in the anterior endoderm at N4 stage (double arrowhead). *Insm* is expressed in the neural plate/tube as well as in the epidermal sensory neurons at neurula stages (arrowheads). Scale bar = 50 μm.

detected in the neural plate/tube and the axial mesoderm during neurulation, quite similarly to *Snail*. The expression of *Pou3f* and *Insm* is first detected in the neural plate at the N2 stage, *Insm* being also expressed from N2 to N4 in ventral epidermal neurons (fig. 5, arrowheads). Only two neural genes show an expression detected by in situ hybridization before G4. *Pou3fl* expression is first observed in the mesendoderm at G2. *Stox* is maternally expressed from the eight-cell stage in the whole embryo. At G0, expression is still observed in the presumptive ectoderm but fades at G1. Then, at G2, *Stox* is transiently expressed in the dorsal mesendoderm. During neurulation, the neural expression of *Stox* is lost, and expression starts in the posterior endoderm (fig. 5, arrowhead) with an additional domain in the anterior endoderm observed at the N4 stage (fig. 5, double arrowhead). These data are consistent with our predicted GRN for neural and epidermal early determination step and show that several epidermal genes start to be expressed in the ectoderm before neural genes and before the mesendoderm touches the ectoderm at G0 and G1 stages.

In order to further validate the in silico predictions, we tested if neural induction could be disturbed by interfering in vivo with the function of the main TF nodes. For the neural program, we tested the function of SoxB1 as it is at the top of the cascade induced by activin and has the largest number of associated TFBS in differentially accessible ATAC-seq peaks between treated vs. untreated explants at the gastrula stage (supplementary figs. S1 and S3, Supplementary Material online). There are three SOXB1 family members in amphioxus that are the result of specific lineage duplications, named *SoxB1a*, *SoxB1b*, and *SoxB1c* (Meulemans and Bronner-Fraser 2007). *SoxB1a* and *SoxB1b* are both expressed in the ectoderm during gastrulation and then at a high level in the neural plate (Holland et al. 2000; Meulemans and Bronner-Fraser 2007) (fig. 5). However, *SoxB1b* is also expressed

maternally and in the whole embryo during cleavage stages and then in the vegetal region of the blastula (Meulemans and Bronner-Fraser 2007). *SoxB1c* is the latest to be expressed and its expression is not restricted to neural tissues (Meulemans and Bronner-Fraser 2007). We, therefore, chose to test the function of *SoxB1a* because it is expressed early but not maternally and could be the SOXB1 family member driving neural induction. Moreover, we reasoned that the putative TFBS recognized by SoxB1a and SoxB1b might be identical, given the high sequence similarity of their DNA binding domains (90% in the HMG-box). Given that no efficient knock-out/knock-down methods exist for the amphioxus model species *Branchiostoma lanceolatum*, we first used a chimera between SoxB1a and the Engrailed repressor domain (SoxB1a–EN) (Jaynes and O'Farrell 1991), an approach that we previously successfully used for other TFs (Aldea et al. 2019). When we injected mRNA coding for such a constitutive repressor form of SoxB1a, development was blocked and embryos failed to terminate gastrulation. However, neural and epidermal fates were acquired by dorsal and ventral ectoderm, respectively, as indicated by the expression of *Elav*, *Neurogenin*, and *K1* (*Keratin1*) (fig. 6). SOXB1 family members are known in many metazoans as factors implicated in the maintenance of pluripotency and proliferation. The developmental arrest observed in injected embryos might be due to a similar property of SOXB1 TFs in amphioxus. Interestingly, the injection of mRNA of the constitutive activator form (a SoxB1a-VP16 chimera) (Sadowski et al. 1988), or of the native form, of SoxB1a did not induce any developmental defect, in line with the fact that SOXB1 putative binding sites might already be occupied at early developmental stages by maternally expressed SoxB1b. Hence, the constitutive repressor form of SoxB1a is not able to block the engagement in the neural fate, and the constitutively active form is not able to induce the formation of neural

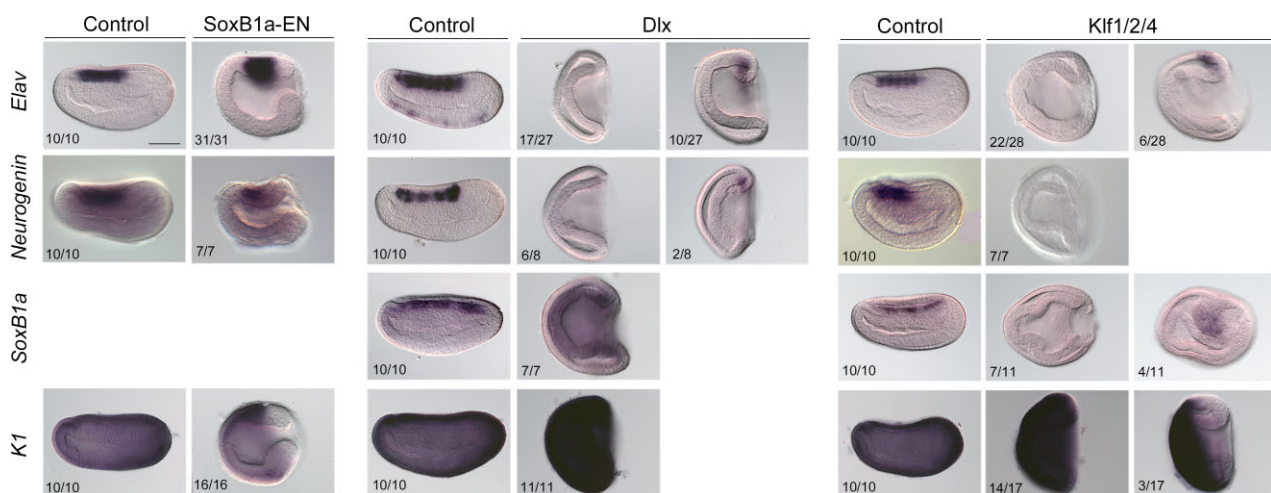**FIG. 6.** Functional analysis of putative GRN nodes. In situ hybridization for *Elav*, *Neurogenin*, *SoxB1a*, and *K1* were undertaken on control embryos and on embryos injected with *SoxB1a-EN*, *Dlx*, or *Klf1/2/4* mRNA and fixed at the N2 stage (19 hpf at 19°C). Side views with anterior to the left and dorsal to the top. Scale bar = 50 μm. The number of embryos showing the same expression pattern is indicated in each picture.

tissues from non-neural ectoderm. Neural induction, if we refer to the proneural putative GRN we constructed, would then directly rely on SMAD activity downstream of Nodal. The position of SoxB1 in the neural putative GRN constructed with gastrula stage data, based on the presence of corresponding putative TFBS in the regulatory regions of genes that are overexpressed after Nodal pathway activation, might reflect a pioneering role of this TF at earlier stages.

Next, given that neural induction seems to involve active repression of the epidermal program from the in silico GRN, we tested if TFs of the epidermal putative GRN, whose expression is downregulated by the activin treatment, were able to block neural induction. We first chose Dlx because it is at the top of the gastrula and neurula putative GRNs, it is the TF with more TFBS in the ATAC-seq peak regions more accessible in untreated explants at the gastrula stage (fig. 2d, supplementary figs. S2 and S4, Supplementary Material online), and it is among the first epidermal genes to be expressed in the ectoderm (fig. 5). Moreover, although there are at least six DLX paralogues in vertebrates, organized in three bigene clusters and showing diverse functions, there is a unique gene in amphioxus, which is considered an ancestral state in chordates (Debiais-Thibaud et al. 2013). In tunicates, there are three DLX genes, with two of them also forming a cluster (Debiais-Thibaud et al. 2013), and one of them, namely Dlx.b, specifies the epidermal fate together with the orthologue of Tfap2 in ascidians (Imai et al. 2017). Another reason we chose Dlx is that at the G5 stage, although this gene is underexpressed in activin-treated explants, it has putative regulatory regions that are more accessible in treated explants compared with control explants and that present SMAD and SOXB1 putative TFBS (supplementary table S21, Supplementary Material online).

We additionally analyzed the function of Klf1/2/4 because it is also at the top of the gastrula cascade (fig. 2d), is transiently expressed in the presumptive epidermis at G4 and G5 stages (fig. 5), and because its vertebrate orthologues are known as the main regulators of epidermis formation and/or as cell reprograming factors (Bialkowska et al. 2017). Embryos injected with mRNA coding for both Dlx and Klf1/2/4 failed to terminate gastrulation, and the whole ectoderm became epidermal as indicated by K1 expression and the absence of Elav, Neurogenin, or SoxB1a expression (fig. 6). Collectively, these functional data indicate that, in the whole embryo, upstream TFs of the epidermal putative GRN are able to reprogram the dorsal ectoderm toward an epidermal fate in line with the fact that neural induction would require a repression of the epidermal program. This proposition is also in accordance with previous data showing that ectoderm regions expressing Neurogenin at the gastrula stage, and that are hence engaged in the neural program, can be redirected toward an epidermal fate in several conditions: when Nodal activation is not continued long enough, when Nodal signal is activated but FGF signal inhibited, or when both Nodal and BMP signals are activated (Le Petillon et al. 2017).

Conclusion

In vertebrates, the default model for neural induction is mainly based on experiments that address the function of several signaling pathways, such as BMP, FGF, and Wnt, and their interactions (Rogers et al. 2009). However, these signals also participate in many developmental processes that take place simultaneously, and their mode of action is not an ON/OFF system. This renders the study of this process very complex. Here, we went beyond the analysis of the function of signaling pathways to better understand the evolution of the first step of CNS formation in chordates by assessing the genetic programs they control during ectodermal cell fate commitment in cephalochordates. Using a global approach, we inferred the epidermal and neural fate acquisition putative GRNs in amphioxus. Our data suggest that the program activated by Nodal during early development might involve TFs that are individually known to be important actors of neural fate commitment in vertebrates, but also more generally in bilaterians, such as SOXB1 and other SOX family members (Wegner 2011), POU family members (Latchman 1999), and bHLH genes (Baker and Brown 2018) such as Neurogenin. More interestingly, the interconnectivity we observed between the neural and the epidermal putative GRNs, with genes of the neural GRN predicted to repress the epidermal GRN, but not the other way around, along with our functional results, strongly suggest that the epidermal program is initiated before neural induction and that epidermis is the default fate of ectoderm in the embryo with neural induction relying on a direct inhibition of the epidermal GRN. This makes sense, evolutionarily speaking at both the developmental time and the evolutionary time points of view. First, considering the development of embryos that form by external fertilization, which was probably the ancestral mode in metazoans, having an external germ layer that rapidly acquires its protective barrier capabilities might have procured a better fitness. Secondly, although the epidermis has long been considered in vertebrates as non-neural ectoderm, this concept is in part biased by the fact that vertebrates have lost the neurogenic capability of the epidermis that is present in both cephalochordates and tunicates (Holland 2005). Indeed, in both species, sensory neurons differentiate within the BMP-positive region of the ectoderm that corresponds to the presumptive epidermis. If we consider the epidermis of cephalochordates comparable to the ectoderm of cnidarians that also forms several populations of neurons by using conserved actors of the SOXB1, POU4F, and bHLH families (Richards and Rentzsch 2014; Rentzsch et al. 2017; Tourniere et al. 2020), the emergence of condensed nervous systems in bilaterians would have been possible thanks to the acquisition of a signal able to block the epidermal program in a large region of the embryo that already had neurogenic potential. The nature of this signal might be different in distinct animal lineages, although the link with dorsoventral patterning and BMP/antiBMP, if not ancestral, might have been recruited several times (Lowe et al. 2006; Yaguchi et al. 2010; Lambert et al. 2016; Su et al. 2019;

Lyons et al. 2020). The question regarding the homology of nerve cords in diverse bilaterian clades has often been addressed by comparing molecules/genes important for their patterning (Arendt and Nubler-Jung 1999; Lowe et al. 2003, 2006; Holland et al. 2013; Kaul-Strehlow et al. 2017; Martin-Duran et al. 2018), but we believe that deciphering the early actors of neural and epidermal fate commitment, as well as their functional connections, will shed more informative light on how CNSs have evolved in metazoans.

Materials and Methods

Embryos and Explants

Branchiostoma lanceolatum embryos were obtained by in vitro fertilization in Petri dishes filled with filtered seawater. Gametes were obtained by heat stimulation of ripe adults collected at the Racou beach near Argelès-sur-Mer, France (latitude 42° 32′ 53″ N and longitude 3° 03′ 27″ E), with a specific permission granted by the Prefect of Region Provence Alpes Côte d'Azur (Fuentes et al. 2007). *Branchiostoma lanceolatum* is not a protected species. Microdissection at the eight-cell stage was undertaken as previously described to separate the four micromeres from the four macromeres (Le Petillon et al. 2020). Micromere explants were transferred to a Petri dish filled with seawater or to a Petri dish filled with seawater plus activin (50 ng/ml, human recombinant protein, R&D) and raised at 19°C. Explants were then collected and either frozen for RNA extraction or directly processed for ATAC-seq. Staging is as proposed in Bertrand et al. (2021) and Carvalho et al. (2021).

RNA-seq and ATAC-seq

RNA was extracted using the RNeasy Plus Micro Kit (Qiagen) following the manufacturer's instructions. For ATAC-seq library construction, 80–100 explants were transferred in a 1.5 ml tube, in duplicates. We then followed the method described in Magri et al. (2020).

Sequencing of the RNA-seq and ATAC-seq libraries was undertaken at the Centre for Genomic Regulation (CRG, Barcelona) Genomics Unit platform.

ATAC-seq data analysis was performed using the nf-core/atacseq pipeline (v1.0.0) (Ewels et al. 2020), which runs Nextflow (v19.10.0) (Di Tommaso et al. 2017), for quality controls, read alignment against Bl71nemr assembly (Marletaz et al. 2018), filtering, data visualization, peak calling, read count, and differential accessibility analysis.

For the RNA-seq data analysis, we used nf-core/rnaseq pipeline (v1.4) (Ewels et al. 2020) for read alignment, read count, and quality controls of the results. After this, we performed a differential gene expression analysis using the DESeq2 R library (v1.30.1) (Love et al. 2014).

GO-term Enrichment

The list of gene names was converted using DAVID 6.8 (Jiao et al. 2012). Singular enrichment analysis was undertaken using agriGO v.2 (Tian et al. 2017). The analysis output was then sent to REVIGO (Supek et al. 2011) for visualization using SimRel as the semantic similarity measure. The interactive graphs were customized using Cytoscape 3.8.2 (Shannon et al. 2003).

TFBS Prediction

Motif analysis on sequences corresponding to ATAC-seq peak regions was undertaken using three methods: HOMER (v4.11, default parameters, vertebrate set of motifs) (Heinz et al. 2010), MEME-ChIP (classic mode and JASPAR CORE [2018] VERTEBRATES as motifs database) (Machanick and Bailey 2011), and RSAT (peak-motifs tool, motifs compared with JASPAR CORE [2020] VERTEBRATES database) (Nguyen et al. 2018).

GRN Construction

Four GRNs were constructed using data of "all neural (G5)" genes, "all epidermal (G5)" genes, "all neural (N2)" genes, and "all epidermal (N2)" genes (fig. 1). For each set of genes, we considered the matrix of HOMER predicted TFBS for each ATAC-seq peak region. We restricted the considered TFBS to 1) TFBS predicted by HOMER, MEME-ChIP, and RSAT; 2) TFBS predicted by HOMER, one of the other algorithms (MEME-ChIP or RSAT) and that correspond to TFs showing differential expression between treated and untreated explants; 3) TFBS of "all neural" and "all epidermal" TFs at the corresponding stage; and 4) SMAD binding sites. From the matrix of interaction (TFs vs. TFBS), a GRN was constructed using GRNsight (Dahlquist et al. 2016). The GRN was then imported in Cytoscape 3.8.2 for graphical representation (Shannon et al. 2003).

Clustering of Whole-Embryo RNA-seq Data

RNA-seq data on whole embryos published in (Marletaz et al. 2018) (available under Gene Expression Omnibus accession GSE106430) were normalized and clustered using Mfuzz (Kumar and Futschik 2007), implemented in R, choosing the fuzzy c-mean soft clustering method. The fuzzifier parameter m was directly estimated from the data. The α-threshold was set to the default value. The core genes for each cluster were used for GO-term enrichment analysis and for TFBS enrichment analysis of the corresponding differentially accessible ATAC-seq peak sequences using analysis of motif enrichment (McLeay and Bailey 2010).

In situ Hybridization of GRN Nodes

Wild-type embryos were fixed at the eight-cell, morula, blastula, G0, G1, G2, G4, G5, N0, N2, N3, and N4 stages in PFA 4% in MOPS buffer (0.1 M MOPS pH 7.5; 2 mM MgSO$_4$; 1 mM EGTA; 0.5 M NaCl). In situ hybridizations were performed as described previously (Somorjai et al. 2008). The accession number/sequences used for probe synthesis are given in supplementary table S22, Supplementary Material online.

Microinjection

All the vectors for mRNA synthesis were constructed using the pCS2+ expression vector backbone. The constitutive repressor form of SoxB1a (SoxB1a-EN) was created by fuzing the coding sequence of the repressor domain of the Engrailed protein to the N-terminal side of the full-length sequence. The constitutively active form of SoxB1a was created by fusion with the coding sequence of the VP16 domain. The vectors were linearized and in vitro transcription was performed using the mMESSAGE mMACHINE SP6 Transcription Kit.

Microinjections of mRNA were carried out as described in Hirsinger et al. (2015). Embryos were fixed at the desired developmental stage in PFA 4%-MOPS buffer for in situ hybridization. In situ hybridizations were undertaken as described previously (Somorjai et al. 2008) using Neurogenin, K1, Elav, and SoxB1a probes synthesized using the DIG-labeling Kit (Roche).

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Author Contributions

Conceptualization: H.E., S.B.
Methodology: H.E., S.B., M.I., J-L.G.-S., J.J.T.
Investigation: A.L., L.S., K.M., I.C., S.B., H.E.
Writing—original draft: S.B., H.E.
Writing—review and editing: A.L., L.S., I.C., M.I., J.J.T., S.B., H.E.
Supervision: H.E., S.B., J.-L.G.-S., J.J.T.

Data Availability

Sequences used for probe synthesis and for mRNA synthesis are available in Genbank (see supplementary table S22, Supplementary Material online). RNA-seq and ATAC-seq data are available under Gene Expression Omnibus (GEO) accession GSE166498.

References

Albuixech-Crespo B, Lopez-Blanch L, Burguera D, Maeso I, Sanchez-Arrones L, Moreno-Bravo JA, Somorjai I, Pascual-Anaya J, Puelles E, Bovolenta P, et al. 2017. Molecular regionalization of the developing amphioxus neural tube challenges major partitions of the vertebrate brain. *PLoS Biol.* **15**:e2001573.

Aldea D, Subirana L, Keime C, Meister L, Maeso I, Marcellini S, Gomez-Skarmeta JL, Bertrand S, Escriva H. 2019. Genetic regulation of amphioxus somitogenesis informs the evolution of the vertebrate head mesoderm. *Nat Ecol Evol.* **3**:1233–1240.

Archer TC, Jin J, Casey ES. 2011. Interaction of Sox1, Sox2, Sox3 and Oct4 during primary neurogenesis. *Dev Biol.* **350**:429–440.

Arendt D, Nubler-Jung K. 1999. Comparison of early nerve cord development in insects and vertebrates. *Development* **126**: 2309–2325.

Baker NE, Brown NL. 2018. All in the family: proneural bHLH genes and neuronal diversity. *Development* **145**:dev159426.

Bertrand S, Carvalho JE, Dauga D, Matentzoglu N, Daric V, Yu JK, Schubert M, Escriva H. 2021. The ontology of the amphioxus anatomy and life cycle (AMPHX). *Front Cell Dev Biol.* **9**:668025.

Bialkowska AB, Yang VW, Mallipattu SK. 2017. Krüppel-like factors in mammalian stem cells and development. *Development* **144**: 737–754.

Bourlat SJ, Juliusdottir T, Lowe CJ, Freeman R, Aronowicz J, Kirschner M, Lander ES, Thorndyke M, Nakano H, Kohn AB, et al. 2006. Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. *Nature* **444**:85–88.

Carvalho JE, Lahaye F, Yong LW, Croce JC, Escriva H, Yu JK, Schubert M. 2021. An updated staging system for cephalochordate development: one table suits them all. *Front Cell Dev Biol.* **9**:668006.

Chen Y, Mistry DS, Sen GL. 2014. Highly rapid and efficient conversion of human fibroblasts to keratinocyte-like cells. *J Invest Dermatol.* **134**:335–344.

Dahlquist KD, Dionisio JDN, Fitzpatrick BG, Anguiano NA, Varshneya A, Southwick BJ, Samdarshi M. 2016. GRNsight: a web application and service for visualizing models of small- to medium-scale gene regulatory networks. *PeerJ Comput Sci.* **2**:e85.

Debiais-Thibaud M, Metcalfe CJ, Pollack J, Germon I, Ekker M, Depew M, Laurenti P, Borday-Birraux V, Casane D. 2013. Heterogeneous conservation of Dlx paralog co-expression in jawed vertebrates. *PLoS One* **8**:e68182.

Delaune E, Lemaire P, Kodjabachian L. 2005. Neural induction in *Xenopus* requires early FGF signalling in addition to BMP inhibition. *Development* **132**:299–310.

Delsuc F, Brinkmann H, Chourrout D, Philippe H. 2006. Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature* **439**:965–968.

De Robertis EM, Kuroda H. 2004. Dorsal–ventral patterning and neural induction in *Xenopus* embryos. *Annu Rev Cell Dev Biol.* **20**:285–308.

Dirksen ML, Morasso MI, Sargent TD, Jamrich M. 1994. Differential expression of a distal-less homeobox gene Xdll-2 in ectodermal cell lineages. *Mech Dev.* **46**:63–70.

Di Tommaso P, Chatzou M, Floden EW, Barja PP, Palumbo E, Notredame C. 2017. Nextflow enables reproducible computational workflows. *Nat Biotechnol.* **35**:316–319.

Ewels PA, Peltzer A, Fillinger S, Patel H, Alneberg J, Wilm A, Garcia MU, Di Tommaso P, Nahnsen S. 2020. The nf-core framework for community-curated bioinformatics pipelines. *Nat Biotechnol.* **38**:276–278.

Ferrero E, Fischer B, Russell S. 2014. SoxNeuro orchestrates central nervous system specification and differentiation in Drosophila and is only partially redundant with Dichaete. *Genome Biol.* **15**:R74.

Fuentes M, Benito E, Bertrand S, Paris M, Mignardot A, Godoy L, Jimenez-Delgado S, Oliveri D, Candiani S, Hirsinger E, et al. 2007. Insights into spawning behavior and development of the European amphioxus (*Branchiostoma lanceolatum*). *J Exp Zool B Mol Dev Evol.* **308B**:484–493.

Garcia-Fernandez J, D'Aniello S, Escriva H. 2007. Organizing chordates with an organizer. *Bioessays* **29**:619–624.

Gilbert SF. 2000. *Developmental biology.* Sunderland (MA): Sinauer Associates.

Glardon S, Holland LZ, Gehring WJ, Holland ND. 1998. Isolation and developmental expression of the amphioxus Pax-6 gene (AmphiPax-6): insights into eye and photoreceptor evolution. *Development* **125**:2701–2710.

Gostling NJ, Shimeld SM. 2003. Protochordate Zic genes define primitive somite compartments and highlight molecular changes underlying neural crest evolution. *Evol Dev.* **5**:136–144.

Hanks MC, Loomis CA, Harris E, Tong CX, Anson-Cartwright L, Auerbach A, Joyner A. 1998. Drosophila engrailed can substitute for mouse Engrailed1 function in mid-hindbrain, but not limb development. *Development* **125**:4521–4530.

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**:576–589.

Hemmati-Brivanlou A, Melton D. 1997. Vertebrate embryonic cells will become nerve cells unless told otherwise. *Cell* **88**:13–17.

Hirsinger E, Carvalho JE, Chevalier C, Lutfalla G, Nicolas JF, Peyrieras N, Schubert M. 2015. Expression of fluorescent proteins in *Branchiostoma lanceolatum* by mRNA injection into unfertilized oocytes. *J Vis Exp.* **95**:52042.

Holland LZ. 2005. Non-neural ectoderm is really neural: evolution of developmental patterning mechanisms in the non-neural ectoderm of chordates and the problem of sensory cell homologies. *J Exp Zool B Mol Dev Evol.* **304B**:304–323.

Holland LZ, Carvalho JE, Escriva H, Laudet V, Schubert M, Shimeld SM, Yu JK. 2013. Evolution of bilaterian central nervous systems: a single origin? *Evodevo* **4**:27.

Holland ND, Panganiban G, Henyey EL, Holland LZ. 1996. Sequence and developmental expression of AmphiDll, an amphioxus distal-less gene transcribed in the ectoderm, epidermis and nervous system: insights into evolution of craniate forebrain and neural crest. *Development* **122**:2911–2920.

Holland LZ, Schubert M, Holland ND, Neuman T. 2000. Evolutionary conservation of the presumptive neural plate markers AmphiSox1/2/3 and AmphiNeurogenin in the invertebrate chordate amphioxus. *Dev Biol.* **226**:18–33.

Hwang J, Kita R, Kwon HS, Choi EH, Lee SH, Udey MC, Morasso MI. 2011. Epidermal ablation of Dlx3 is linked to IL-17-associated skin inflammation. *Proc Natl Acad Sci U S A.* **108**:11566–11571.

Imai KS, Hikawa H, Kobayashi K, Satou Y. 2017. Tfap2 and Sox1/2/3 cooperatively specify ectodermal fates in ascidian embryos. *Development* **144**:33–37.

Irimia M, Pineiro C, Maeso I, Gomez-Skarmeta JL, Casares F, Garcia-Fernandez J. 2010. Conserved developmental expression of Fezf in chordates and Drosophila and the origin of the Zona Limitans Intrathalamica (ZLI) brain organizer. *Evodevo* **1**:7.

Jaynes JB, O'Farrell PH. 1991. Active repression of transcription by the engrailed homeodomain protein. *EMBO J.* **10**:1427–1433.

Jiao X, Sherman BT, Huang da W, Stephens R, Baseler MW, Lane HC, Lempicki RA. 2012. DAVID-WS: a stateful web service to facilitate gene/protein list analysis. *Bioinformatics* **28**:1805–1806.

Kaul-Strehlow S, Urata M, Praher D, Wanninger A. 2017. Neuronal patterning of the tubular collar cord is highly conserved among enteropneusts but dissimilar to the chordate neural tube. *Sci Rep.* **7**:7003.

Kotkamp K, Mossner R, Allen A, Onichtchouk D, Driever W. 2014. A Pou5f1/Oct4 dependent Klf2a, Klf2b, and Klf17 regulatory sub-network contributes to EVL and ectoderm development during zebrafish embryogenesis. *Dev Biol.* **385**:433–447.

Kozmik Z, Holland ND, Kreslova J, Oliveri D, Schubert M, Jonasova K, Holland LZ, Pestarino M, Benes V, Candiani S. 2007. Pax-Six-Eya-Dach network during amphioxus development: conservation in vitro but context specificity in vivo. *Dev Biol.* **306**:143–159.

Kozmik Z, Holland LZ, Schubert M, Lacalli TC, Kreslova J, Vlcek C, Holland ND. 2001. Characterization of Amphioxus AmphiVent, an evolutionarily conserved marker for chordate ventral mesoderm. *Genesis* **29**:172–179.

Kumar L, Futschik ME. 2007. Mfuzz: a software package for soft clustering of microarray data. *Bioinformation* **2**:5–7.

Lambert JD, Johnson AB, Hudson CN, Chan A. 2016. Dpp/BMP2-4 mediates signaling from the D-quadrant organizer in a spiralian embryo. *Curr Biol.* **26**:2003–2010.

Lamouille S, Xu J, Derynck R. 2014. Molecular mechanisms of epithelial–mesenchymal transition. *Nat Rev Mol Cell Biol.* **15**:178–196.

Langeland JA, Tomsa JM, Jackman WR Jr, Kimmel CB. 1998. An amphioxus snail gene: expression in paraxial mesoderm and neural plate suggests a conserved role in patterning the chordate embryo. *Dev Genes Evol.* **208**:569–577.

Latchman DS. 1999. POU family transcription factors in the nervous system. *J Cell Physiol.* **179**:126–133.

Lee B, Villarreal-Ponce A, Fallahi M, Ovadia J, Sun P, Yu QC, Ito S, Sinha S, Nie Q, Dai X. 2014. Transcriptional mechanisms link epithelial plasticity to adhesion and differentiation of epidermal progenitor cells. *Dev Cell* **29**:47–58.

Le Petillon Y, Bertrand S, Escriva H. 2020. Spawning induction and embryo micromanipulation protocols in the amphioxus *Branchiostoma lanceolatum*. *Methods Mol Biol.* **2047**:347–359.

Le Petillon Y, Luxardi G, Scerbo P, Cibois M, Leon A, Subirana L, Irimia M, Kodjabachian L, Escriva H, Bertrand S. 2017. Nodal–activin pathway is a conserved neural induction signal in chordates. *Nat Ecol Evol.* **1**:1192–1200.

Li L, Wang Y, Torkelson JL, Shankar G, Pattison JM, Zhen HH, Fang F, Duren Z, Xin J, Gaddam S, et al. 2019. TFAP2C- and p63-dependent networks sequentially rearrange chromatin landscapes to drive human epidermal lineage commitment. *Cell Stem Cell* **24**:271–284.e8.

Linker C, Stern CD. 2004. Neural induction requires BMP inhibition only as a late step, and involves signals other than FGF and Wnt antagonists. *Development* **131**:5671–5681.

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

Lowe CJ, Terasaki M, Wu M, Freeman RM Jr, Runft L, Kwan K, Haigo S, Aronowicz J, Lander E, Gruber C, et al. 2006. Dorsoventral patterning in hemichordates: insights into early chordate evolution. *PLoS Biol.* **4**:e291.

Lowe CJ, Wu M, Salic A, Evans L, Lander E, Stange-Thomann N, Gruber CE, Gerhart J, Kirschner M. 2003. Anteroposterior patterning in hemichordates and the origins of the chordate nervous system. *Cell* **113**:853–865.

Lyons DC, Perry KJ, Batzel G, Henry JQ. 2020. BMP signaling plays a role in anterior-neural/head development, but not organizer activity, in the gastropod Crepidula fornicata. *Dev Biol.* **463**:135–157.

Machanick P, Bailey TL. 2011. MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics* **27**:1696–1697.

Magri MS, Jimenez-Gancedo S, Bertrand S, Madgwick A, Escriva H, Lemaire P, Gomez-Skarmeta JL. 2020. Assaying chromatin accessibility using ATAC-seq in invertebrate chordate embryos. *Front Cell Dev Biol.* **7**:372.

Marletaz F, Firbas PN, Maeso I, Tena JJ, Bogdanovic O, Perry M, Wyatt CDR, de la Calle-Mustienes E, Bertrand S, Burguera D, et al. 2018. Amphioxus functional genomics and the origins of vertebrate gene regulation. *Nature* **564**:64–70.

Martin-Duran JM, Pang K, Borve A, Le HS, Furu A, Cannon JT, Jondelius U, Hejnol A. 2018. Convergent evolution of bilaterian nerve cords. *Nature* **553**:45–50.

McLeay RC, Bailey TL. 2010. Motif enrichment analysis: a unified framework and an evaluation on ChIP data. *BMC Bioinform* **11**:165.

Meulemans D, Bronner-Fraser M. 2002. Amphioxus and lamprey AP-2 genes: implications for neural crest evolution and migration patterns. *Development* **129**:4953–4962.

Meulemans D, Bronner-Fraser M. 2007. The amphioxus *SoxB* family: implications for the evolution of vertebrate placodes. *Int J Biol Sci.* **3**:356–364.

Miles LB, Darido C, Kaslin J, Heath JK, Jane SM, Dworkin S. 2017. Mis-expression of grainyhead-like transcription factors in zebrafish leads to defects in enveloping layer (EVL) integrity, cellular morphogenesis and axial extension. *Sci Rep.* **7**:17607.

Neidert AH, Panopoulou G, Langeland JA. 2000. Amphioxus goosecoid and the evolution of the head organizer and prechordal plate. *Evol Dev.* **2**:303–310.

Nguyen NTT, Contreras-Moreira B, Castro-Mondragon JA, Santana-Garcia W, Ossio R, Robles-Espinoza CD, Bahin M, Collombet S, Vincens P, Thieffry D, et al. 2018. RSAT 2018: regulatory sequence analysis tools 20th anniversary. *Nucleic Acids Res.* **46**:W209–W214.

Nitta KR, Jolma A, Yin Y, Morgunova E, Kivioja T, Akhtar J, Hens K, Toivonen J, Deplancke B, Furlong EE, et al. 2015. Conservation of transcription factor binding specificities across 600 million years of bilateria evolution. *Elife* **4**:e04837.

Omrani MR, Yaqubi M, Mohammadnia A. 2018. Transcription factors in regulatory and protein subnetworks during generation of neural stem cells and neurons from direct reprogramming of non-fibroblastic cell sources. *Neuroscience* **380**:63–77.

Onai T, Takai A, Setiamarga DH, Holland LZ. 2012. Essential role of Dkk3 for head formation by inhibiting Wnt/β-catenin and Nodal/Vg1 signaling pathways in the basal chordate amphioxus. *Evol Dev.* **14**:338–350.

Ozair MZ, Kintner C, Brivanlou AH. 2013. Neural induction and early patterning in vertebrates. *Wiley Interdiscip Rev Dev Biol.* **2**:479–498.

Pera E, Stein S, Kessel M. 1999. Ectodermal patterning in the avian embryo: epidermis versus neural plate. *Development* **126**:63–73.

Qiao Y, Zhu Y, Sheng N, Chen J, Tao R, Zhu Q, Zhang T, Qian C, Jing N. 2012. AP2γ regulates neural and epidermal development downstream of the BMP pathway at early stages of ectodermal patterning. *Cell Res.* **22**:1546–1561.

Rentzsch F, Layden M, Manuel M. 2017. The cellular and molecular basis of cnidarian neurogenesis. *Wiley Interdiscip Rev Dev Biol.* **6**:e257.

Richards GS, Rentzsch F. 2014. Transgenic analysis of a *SoxB* gene reveals neural progenitor cells in the cnidarian *Nematostella vectensis*. *Development* **141**:4681–4689.

Rogers CD, Moody SA, Casey ES. 2009. Neural induction and factors that stabilize a neural fate. *Birth Defects Res C Embryo Today* **87**:249–262.

Sadowski I, Ma J, Triezenberg S, Ptashne M. 1988. GAL4-VP16 is an unusually potent transcriptional activator. *Nature* **335**:563–564.

Sankar S, Yellajoshyula D, Zhang B, Teets B, Rockweiler N, Kroll KL. 2016. Gene regulatory networks in neural cell fate acquisition from genome-wide chromatin association of Geminin and Zic1. *Sci Rep.* **6**:37412.

Saxena K, Srikrishnan S, Celia-Terrassa T, Jolly MK. 2022. OVOL1/2: drivers of epithelial differentiation in development, disease, and reprogramming. *Cells Tissues Organs* **211**:183–192. doi:10.1159/000511383

Schubert M, Holland ND, Holland LZ. 1998. Amphioxus AmphiDRAL encoding a LIM-domain protein: expression in the epidermis but not in the presumptive neuroectoderm. *Mech Dev.* **76**:203–205.

Segre JA, Bauer C, Fuchs E. 1999. Klf4 is a transcription factor required for establishing the barrier function of the skin. *Nat Genet.* **22**:356–360.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**:2498–2504.

Shimeld SM. 1997. Characterisation of amphioxus HNF-3 genes: conserved expression in the notochord and floor plate. *Dev Biol.* **183**:74–85.

Shimeld SM, van den Heuvel M, Dawber R, Briscoe J. 2007. An amphioxus Gli gene reveals conservation of midline patterning and the evolution of hedgehog signalling diversity in chordates. *PLoS One* **2**:e864.

Somorjai I, Bertrand S, Camasses A, Haguenauer A, Escriva H. 2008. Evidence for stasis and not genetic piracy in developmental expression patterns of *Branchiostoma lanceolatum* and *Branchiostoma floridae*, two amphioxus species that have evolved independently over the course of 200 Myr. *Dev Genes Evol.* **218**:703–713.

Stern CD. 2006. Neural induction: 10 years on since the 'default model'. *Curr Opin Cell Biol.* **18**:692–697.

Streit A, Berliner AJ, Papanayotou C, Sirulnik A, Stern CD. 2000. Initiation of neural induction by FGF signalling before gastrulation. *Nature* **406**:74–78.

Su YH, Chen YC, Ting HC, Fan TP, Lin CY, Wang KT, Yu JK. 2019. BMP controls dorsoventral and neural patterning in indirect-developing hemichordates providing insight into a possible origin of chordates. *Proc Natl Acad Sci U S A.* **116**:12925–12932.

Supek F, Bosnjak M, Skunca N, Smuc T. 2011. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* **6**:e21800.

Takahashi K, Yamanaka S. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**:663–676.

Thomas J, Morle L, Soulavie F, Laurencon A, Sagnol S, Durand B. 2010. Transcriptional control of genes involved in ciliogenesis: a first step in making cilia. *Biol Cell* **102**:499–513.

Tian T, Liu Y, Yan H, You Q, Yi X, Du Z, Xu W, Su Z. 2017. agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res.* **45**:W122–W129.

Torrence SA, Cloney RA. 1982. Nervous system of ascidian larvae: caudal primary sensory neurons. *Zoomorphology* **99**:103–115.

Tourniere O, Dolan D, Richards GS, Sunagar K, Columbus-Shenkar YY, Moran Y, Rentzsch F. 2020. NvPOU4/Brain3 functions as a terminal selector gene in the nervous system of the cnidarian Nematostella vectensis. *Cell Rep.* **30**:4473–4489.e5.

Venkatesh TV, Holland ND, Holland LZ, Su MT, Bodmer R. 1999. Sequence and developmental expression of amphioxus AmphiNk2-1: insights into the evolutionary origin of the vertebrate thyroid gland and forebrain. *Dev Genes Evol.* **209**:254–259.

Wegner M. 2011. SOX after SOX: SOXession regulates neurogenesis. *Genes Dev.* **25**:2423–2428.

Wicht H, Lacalli TC. 2005. The nervous system of amphioxus: structure, development, and evolutionary significance. *Can J Zool.* **83**:122–150.

Williams NA, Holland PW. 1996. Old head on young shoulders. *Nature* **383**:490.

Wilson SI, Graziano E, Harland R, Jessell TM, Edlund T. 2000. An early requirement for FGF signalling in the acquisition of neural cell fate in the chick embryo. *Curr Biol.* **10**:421–429.

Yaguchi S, Yaguchi J, Angerer RC, Angerer LM, Burke RD. 2010. TGFβ signaling positions the ciliary band and patterns neurons in the sea urchin embryo. *Dev Biol.* **347**:71–81.

Yu JK, Satou Y, Holland ND, Shin IT, Kohara Y, Satoh N, Bronner-Fraser M, Holland LZ. 2007. Axial patterning in cephalochordates and the evolution of the organizer. *Nature* **445**:613–617.

Zhang T, Zhu Q, Xie Z, Chen Y, Qiao Y, Li L, Jing N. 2013. The zinc finger transcription factor Ovol2 acts downstream of the bone morphogenetic protein pathway to regulate the cell fate decision between neuroectoderm and mesendoderm. *J Biol Chem.* **288**:6166–6177.