A single cell transcriptome atlas of the developing zebrafish hindbrain

Monica Tambalo[1], Richard Mitter[2] and David G. Wilkinson[1*]

[1] Neural Development Laboratory,
The Francis Crick Institute,
1 Midland Road,
London NW1 1AT, UK.

[2] Bioinformatics and Biostatistics,
The Francis Crick Institute,
1 Midland Road,
London NW1 1AT, UK.

* Corresponding author
Email: david.wilkinson@crick.ac.uk

Key words: hindbrain segmentation, dorsoventral patterning, neurogenesis, single cell RNA sequencing, hindbrain boundary, fgf signaling

Abstract

Segmentation of the vertebrate hindbrain leads to the formation of rhombomeres, each with a distinct anteroposterior identity. Specialised boundary cells form at segment borders that act as a source or regulator of neuronal differentiation. In zebrafish, there is spatial patterning of neurogenesis in which non-neurogenic zones form at boundaries and segment centres, in part mediated by Fgf20 signaling. To further understand the control of neurogenesis, we have carried out single cell RNA sequencing of the zebrafish hindbrain at three different stages of patterning. Analyses of the data reveal known and novel markers of distinct hindbrain segments, of cell types along the dorsoventral axis, and of the transition of progenitors to neuronal differentiation. We find major shifts in the transcriptome of progenitors and of differentiating cells between the different stages analysed. Supervised clustering with markers of boundary cells and segment centres, together with RNA-seq analysis of Fgf-regulated genes, has revealed new candidate regulators of cell differentiation in the hindbrain. These data provide a valuable resource for functional investigations of the patterning of neurogenesis and the transition of progenitors to neuronal differentiation.

Introduction

Development of the central nervous system (CNS) requires precise regulation of the differentiation of neuronal and glial cell types from neural progenitor cells. This is achieved through a network of cell-cell signaling and transcription factors that inhibit or promote cell differentiation and specify cell type along the dorsoventral (D-V) and anteroposterior (A-P) axes of the neuroepithelium. Cell specification along the D-V axis involves localised sources of Shh, BMP and Wnt signals that act in a concentration-dependent manner to regulate expression of specific transcription factors (Dessaud et al., 2008; Dessaud et al., 2007; Hikasa and Sokol, 2013; Ikeya et al., 1997; Lee and Jessell, 1999; Liem et al., 1997; Panhuysen et al., 2004; Timmer et al., 2002; Ulloa and Marti, 2010). This positional information is integrated with patterning along the anteroposterior axis, which regulates expression of transcription factors that specify regional identity within the brain and spinal cord (Alexander et al., 2009). Differentiation is also under temporal regulation, with distinct neuronal or glial cell types arising at different times (Guillemot, 2007). It is essential that a pool of progenitor cells is maintained as a source of later-differentiating cells, and this is achieved by multiple mechanisms that inhibit differentiation.

The switch of progenitor cells to neuronal differentiation involves the sustained high-level expression of proneural transcription factors that initiate a cascade of gene expression leading to expression of terminal neuronal markers (Bertrand et al., 2002). The expression and function of proneural genes is antagonised by intrinsic factors, as well as by extrinsic signals such as Notch ligands and Fgfs that inhibit differentiation (Fisher and Caudy, 1998; Gonzalez-Quevedo et al., 2010; Kageyama et al., 2005; Ortega et al., 1998; Vaccarino et al., 1999; Zheng et al., 2004). In some regions of the developing CNS, neurogenesis occurs widely in the neuroepithelium, and lateral inhibition due to expression of Notch ligands by differentiating neurons ensures that progenitor cells are maintained (Pierfelice et al., 2011). In other regions, there is a patterning of neurogenesis, for example due to spatially-resticted expression along the anteroposterior or D-V axis of Hes/Her genes that inhibit neuronal differentiation (Bae et al., 2005; Geling et al., 2003). Studies of the vertebrate hindbrain have revealed further mechanisms that regulate the patterning of neuronal differentiation.

The hindbrain is an important component of the CNS which includes neurons that innervate cranial muscles, that relay sensory inputs, and control breathing, the heart and gastrointestinal systems. At early stages, the neuroepithelium of the hindbrain is subdivided to form seven rhombomeres (r1-r7), each expressing a distinct set of transcription factors, including egr2 (krox20), mafB, vhnf1 and hox genes, that underlie segmentation and anteroposterior identity (Alexander et al., 2009). A similar but different set of neurons is generated in each rhombomere (Clarke and Lumsden, 1993; Lumsden, 2004; Lumsden and Keynes, 1989); for example, in mouse the Vth, VIIth and IXth branchiomotor nerves form in r2+r3, r4+r5 and r5+r6, respectively. There is a partial understanding of mechanisms that link A-P identity to neuronal cell type specification in the hindbrain (Narita and Rijli, 2009).

Boundary formation has a crucial role in the organisation of neurons and neurogenesis in the hindbrain. Through a combination of cell identity regulation (Addison et al., 2018; Wang et al., 2017) and Eph-ephrin mediated cell segregation (Batlle and Wilkinson, 2012; Cayuso et al., 2015; Fagotto, 2014), each rhombomere is demarcated by sharp borders and has a homogeneous segmental identity. Specialised boundary cells form at each rhombomere border (Guthrie and Lumsden, 1991), which express specific molecular markers (Cheng et al., 2004; Cooke et al., 2005; Heyman et al., 1995; Letelier et al., 2018; Xu et al., 1995). These boundary cells are induced by Eph receptor signaling that leads to an increase in mechanical tension and activation of Taz (Cayuso et al., 2019). In

the chick hindbrain, boundary cells have a lower rate of proliferation (Guthrie et al., 1991) and are Sox2-expressing neural stem cells that are a source of neurogenesis (Peretz et al., 2016). A different situation occurs in the zebrafish hindbrain, in which expression of proneural transcription factors is initially widespread, and later becomes confined to zones flanking hindbrain boundary cells (Amoyel et al., 2005; Cheng et al., 2004). Notch activation promoted by *rfng* expression inhibits neurogenesis at early stages in boundary cells (Cheng et al., 2004). In addition, there is increased proliferation and inhibition of neurogenesis in boundary cells by activation of the Yap/Taz pathway downstream of mechanical tension (Voltes et al., 2019). At late stages (after 40 hpf) proliferation declines and neurogenesis starts to occur in boundary progenitors (Voltes et al., 2019), similar to the situation in chick (Peretz et al., 2016). Neurogenesis is inhibited at segment centres by fgf20-expressing neurons that act on the adjacent neuroepithelium (Gonzalez-Quevedo et al., 2010). The clustering of fgf20-expressing neurons at segment centres is maintained by semaphorin-mediated chemorepulsion from boundary cells (Terriente et al., 2012). In addition to suppressing neuronal differentiation, Fgf signaling may switch progenitors at the segment centre to glial differentiation (Esain et al., 2010). The zebrafish hindbrain thus has a precise organisation of signaling sources that underlies a stereotyped pattern of neurogenic and non-neurogenic zones, and the positioning of neurons within each segment.

We set out to identify further potential regulators of neurogenesis during hindbrain segmentation by single cell RNA sequencing (scRNA-seq) to identify genes specifically expressed in distinct progenitors and differentiating cells, prior to and during the patterning of neurogenesis. Analyses of the transcriptome of single cells revealed known genes and new markers of distinct hindbrain segments, of cell types along the D-V axis, and of the transition of progenitors to neuronal differentiation. We also find temporal changes in gene expression, both in progenitors and differentiating cells, at the different stages analysed. By carrying out supervised clustering, we have identified further genes specifically expressed in hindbrain boundary cells and segment centres. These findings are compared with bulk RNA-seq analyses following loss and gain of Fgf signaling to identify potential regulators expressed in segment centres.

RESULTS

Single-cell profiling of the developing zebrafish hindbrain and surrounding tissues

To further understand the progressive patterning of neurogenesis of the developing zebrafish hindbrain, we analysed the transcriptome of single cells at three developmental stages (Fig. 1A, B): 16 hpf (prior to patterning of neurogenesis), 24 hpf (beginning of neurogenic patterning) and 44 hpf (pattern of neurogenic and non-neurogenic zones fully established). For each stage, we micro-dissected the hindbrain territory from around 40 embryos, which were pooled. After enzymatic digestion and mechanical dissociation, the single-cell suspension was loaded into the droplet-based scRNA-seq platform 10X Genomics Chromium (Fig. 1C). In total, 9026 cells were sequenced (2929 at 16 hpf, 2568 at 24 hpf and 3529 at 44 hpf), with an average number of UMIs of 6916 and 1703 median genes per cell (Fig. S1).

Seurat unsupervised clustering was used to classify cell population identity (Butler et al., 2018a; Stuart et al., 2018) after aggregating the data from all stages (Fig. S2). Cluster projection onto UMAP plots (Becht et al., 2018; McInnes et al., 2018) revealed a tight group of cells with some substructure, and a number of peripheral clusters (Fig. S2A). Since the dissections included tissues adjacent to the hindbrain, it is likely that the clusters correspond to distinct tissue types. We therefore used tissue marker genes to assign cluster identity. The progenitor marker sox3 and neuronal gene elavl3 are found to mark complementary parts of the main group of cells and together define the hindbrain territory (Fig. S2B, C). This group of cells has a substructure due to changes in transcriptome within and between different stages that will be analysed below. sox3 also marks a peripheral cluster of hindbrain cells that co-express shh (Fig. S2D) and therefore derive from the floor plate. The expression of marker genes reveals that other clusters correspond to tissues found next to the hindbrain, as follows: neural crest (*foxd3*, *twist1a*), head mesenchyme and mesendoderm (*colec12, col9a2*), vasculature (*sox7*), pharyngeal arches (*foxi1*), epidermis (*krt17*), otic vesicle (*eya2*), and otic and cranial ganglia (*neurod1*) (Fig. S2A, D). Based on this analysis, we bioinformatically recovered hindbrain cells for each stage: 1678 cells at 16 hpf, 1722 cells at 24 hpf and 2729 cells at 44 hpf (Table S1).

Overall changes in hindbrain tissue composition

We used an unsupervised graph-based clustering approach to analyse the transcriptome data at 16 hpf, 24 hpf and 44 hpf. Data sets were visualized with UMAP dimensionality reduction, and this revealed unique features that reflect the greatest transcriptomic

differences between cell types at each developmental stage (Fig. 2A, Fig. 3A, Fig. 4A). Analysis of the top 30 significantly enriched genes per each cluster, and expression of known molecular markers, enabled each cluster to be identified. The 16 hpf hindbrain is mainly constituted of progenitors (91% of total hindbrain cells), and cells at different stages of neurogenesis (neurod4, elevated neurog1 expression) account for 6% of hindbrain cells (cluster C6 in Fig. 2A). Progenitors remain the most abundant hindbrain cell type at 24 hpf (71% of hindbrain cells), while 28% of cells express markers of different stages of neurogenesis and late differentiation (C4, 6, 9, 10, 11, 12 in Fig. 3A). By 44 hpf, the proportion of progenitor cells has further diminished to 40%, with 55% of the cells expressing markers of neurogenesis and late stages of neuronal differentiation (C0, 3, 5, 6, 7, 8, 9, 10, 11, 12, 14 in Fig. 4A). The clustering of cells by transcriptomic differences changes at the three stages. At 16 hpf, clustering is mainly driven by segmental and D-V identity (Fig. 2A), whereas at 24 hpf and 44 hpf cells are clustered by D-V identity and differentiation state (Fig. 3A; Fig. 4A). This change reflects the greater proportion of cells undergoing differentiation at the later stages, with an increasing number of neuronal sub-types by 44 hpf (Fig. 4A). Below, we present more detailed analyses of each of these features that reveal known genes and novel markers of segmental identity, D-V identity and differentiation state. An annotated list including information on any previous studies of these genes is presented in Table S2.

Transcriptional signatures of hindbrain segments

The expression of known markers enables the identity of all clusters (C0-C9) at 16 hpf to be deduced (Fig. 2A). At this stage, the main features that drive clustering of hindbrain cells are segmental identity and D-V identity of progenitors, and one cluster of cells (C6) undergoing neurogenesis. We display the genes that distinguish the different clusters in a heatmap of the top 30 differentially-expressed genes (Fig. 2B; Fig. S3) and show the expression level of selected genes in UMAP projection plots that relate them to the Seurat analysis (Fig. 2C). Genes specifically expressed in different hindbrain rhombomeres (r), or in dorsal, medial or ventral domains, are listed in Fig. 2D and Fig. 2E, respectively. The single cell gene expression strongly correlates with in situ hybridisation data deposited in ZFIN. Information on any previous studies of these genes is presented in Table S2.1.

UMAP projection plots with dorsal and ventral marker genes (Fig. 2C, E) reveal the relationship between D-V identity and the clustering of cells in Seurat. For example, *zic2b* expression marks the dorsal part of all hindbrain segments, and *neurog1* marks ventral progenitors as well as differentiating neurons (Fig. 2A, 2C). For some hindbrain segments

(r2, r3, r4, r7) but not others (r1, r5, r6), cells with distinct D-V identity segregate into discrete clusters; this presumably reflects the quantitative difference in transcriptome in relation to the threshold for assigning cells to different clusters. Indeed, increasing cluster resolution further subdivides the hindbrain territory in a total of 19 clusters (Fig. S4A), with increased segregation into dorsal, medial and ventral populations (Fig. S4B). Roof plate cells (C8) in the dorsal-most neuroepithelium form a discrete cluster, expressing markers including *bmp5* and *nog1* (Fig. 2B), that is adjacent to cells expressing dorsal markers (Fig. 2A). As also seen in the aggregated data (Fig. S2), floor plate cells form a cluster (C9) that is distant in UMAP from other hindbrain cells.

The clustering of cells based on segmental identity is revealed in projection plots of selected marker genes that are expressed in different sets of segments (Fig. 2C): *eng2a* (MHB-r1), *hoxa2* (r2-r5), *egr2b* (r3, r5), *mafba* (r5, r6), *hoxa3* (r5-r7), and *hoxd4a* (r7). Cells from r2, r3 and r4 co-cluster in C0-C1, where C0 cells are ventral and C1 cells are dorsal (Fig. 2A). Seurat analysis did not discriminate r2 and r4 cells, suggesting strong transcriptional similarities, including *egfl6.1*, *fabp7a* and *sfrp5* expression (Fig. 2D). Cells from r3 are included in C0-C1, but form a discrete group that is marked, for example, by *egr2b* expression (Fig. 2A, C, D). This clustering of r2, r3 and r4 cells reflects that genes including *hoxa2b*, *sfrp5* and *sp8a* are expressed in all three segments, whereas *egr2b*, *epha4a*, *sema3fb* and other markers are expressed in r3 cells (Fig. 2C, D). After increasing cluster resolution, r3 becomes segregated from r2 and r4 (Fig. S4A). Consistent with previous studies, r3 cells are adjacent to r5 cells (C3), reflecting that they express some genes in common: in addition to the extensively-studied *egr2b* and *epha4a* genes, they express *timp2a*, *aldocb*, *smea3fb* and *myo1cb* (Fig. 2D). r5 also shares transcriptional similarities with r6, which forms an adjacent cluster (C4), including *mafba* (Fig. 2C), *cryba2b*, *crygn2*, *lim2.1*, *col15a1b* and *gas6*. However, r7 cells (C2 ventral and C5 dorsal) do not cluster adjacent to r6 cells, reflecting that although some genes are expressed in both segments (for example, *hoxa3a*, *hoxb3a* and *tox3*), many other genes are expressed in one or the other, for example, *hoxd4a* (Fig. 2C), *fabp7a*, *lratb*, *rbp5*, *rhbdl3* and *sp8a* in r7 (Fig. 2D). r1 and midbrain-hindbrain boundary (MHB) cells which express known markers (*eng2a/b* (Fig. 2C), *fgf8a, cnpy1* and *pax2a*) are found to cluster together in C7. As summarised in Table S2.1, these analyses have identified genes not previously described to have segmental expression in the hindbrain; these include *myo1cb* and *timp2b* in r3 and r5. In addition, we found genes for which expression data is available, but have not been tested functionally in the hindbrain; these include *sp8a* (strong in r4 and r7, weak in r2 and r3), *sfrp5* (r2-r4), and *wnt7aa* (r3-r7).

We further analysed the transcriptome data using PlotClusterTree in Seurat as this better represents the similarity between clusters than UMAP distance. Analysis of the 16 hpf data at higher cluster resolution (Fig. S4C) segregates cells with distinct segmental identity, D-V identity and differentiation state. Cells from r2, r3 and r4 are found to be closely related and are further subdivided based on D-V rather than segmental identity: ventral r3 and ventral r2+r4 form adjacent branches, and dorsal r3 and dorsal r2+r4 form adjacent branches. This suggests that dorsoventral identity underlies greater transcriptomic similarities between cell clusters than segmental identity within this population. The tree analysis reveals further clusters of r2+r4 cells (3 and 6) that in heat maps are found to have higher expression of genes related to cell proliferation. The tree analysis suggests that r5, r6 and most r7 cells are closely related and subdivides them into sequential and discrete branches, each further subdivided into dorsal and ventral populations. Some cells classified as r7 (cluster 15) form a separate branch; however, we find that these do not express *hoxd4a*, and may correspond to spinal cord cells caudal to the hindbrain. Finally, the MHB-r1, roof plate, differentiating neurons, and floor plate form separate branches.

Dorsoventral signatures of progenitors and differentiating neurons

D-V positional information regulated by BMP, Wnt and Shh signaling is a key feature of the developing neuroepithelium that underlies specification of neuronal cell types. Extensive molecular characterization has been carried out in the spinal cord (Delile et al., 2019; Gouti et al., 2015), but less widely for the hindbrain. At all stages analysed, progenitors were clustered based on their D-V identity, reflecting that D-V patterning is established early and maintained during hindbrain neurogenesis. Seurat analysis at 16 hpf segregates cells into dorsal and ventral progenitors, as well as roof plate and floor plate (Fig. 2A). However, UMAP projection plots with known markers (listed in Fig. 2E), and increasing cluster resolution (Fig. S4), reveals that these are further subdivided into dorsal, medial and ventral domains. Seurat analysis at 24 hpf and 44 hpf clusters cells into dorsal, medial and ventral populations, plus roof plate and floor plate (Fig. 3A, Fig. 4A). In addition, progenitor cells are further segregated based on expression of proliferation markers. Selected genes that mark these different populations are presented in UMAP projection plots (Fig. 3C, Fig. 4C), in dot plots of relative expression levels in progenitors and differentiating neurons (Fig. 3D, E; Fig. 4D), and in situ hybridisation analyses (Fig. 3F-H; Fig. 4E-H). We describe the 24 hpf and 44 hpf data in more detail below.

At 24 hpf, clusters C0, and part of C1 and C5 are found to express known markers of dorsal progenitors, including *zic2b* (Fig. 3A, C), other *zic* genes (Elsen et al., 2008; Grinblat and Sive, 2001), *msx1b/3* (Miyake et al., 2012), and *olig3/4* (Tiso et al., 2009) (Fig. 3B, D). In addition, these cells express novel markers including *casz1*, *cdon*, *fzd10*, *myca*, and *pdgfaa* (Fig. 2B, D; Fig. S5; Table S2.2). C1 is distinguished from C0 and C5 by expression of proliferation markers, including *cdca8* (Fig. 3A, C). Expression of dorsal markers including *zic2b* is also detected in dorsal differentiating neurons (C11, Fig. 3A, C). Dorsal progenitors in C0 and C1 and dorsal differentiating neurons (C11) express the proneural gene *atoh1a* (Fig. 3C, E) (Elsen et al., 2009), which we verified by in situ hybridisation (Fig. 3F, F'). Medial progenitors are found in a subset of cells in C1, C3 and C5, sharing a few dorsally- (e.g. *zic* genes) and ventrally-expressed (e.g. *foxb1a, pax6a*) factors, while uniquely expressing markers including *gsx1*, *pax7a/b, ptf1a* and *lbx1b* (Fig.3C, D; Fig. S5; Table S2.2). This analysis further shows that the proneural gene *ascl1a* is expressed medially in hindbrain progenitors (C1, C5) and differentiating neurons (C10), with expression overlapping with *neurod4* (Fig. 3C, E; in situ hybridisation in Fig. 3G, G'). Ventral progenitors are subdivided into multiple clusters (C7, C2, C3, C8). Cells in C7 and C3 express higher levels of factors involved in the cell cycle, for instance *mcm* genes (*mcm2-6*), while ventral-restricted genes are enriched in C2 (e.g. *dbx1a/b*) and C8 (e.g. *irx3a*) (Fig. S5). Overall, we found a ventral progenitor signature in which they express a unique set of transcription factors: *sox21a*, *foxb1a*, *sp8a* and *dbx1a/1b*. These ventral progenitors and differentiating neurons express the proneural gene, *neurog1* (Fig. 3C, E), which we verified by in situ hybridisation (Fig. 3H, H'). In addition, these cells express several signalling modulators: *sfrp5* (soluble inhibitor of Wnt signalling), *cyp26b1* (RA degradation), *scube2* (Shh long-range signalling), and *sulf2b* (heparan sulfate proteoglycans) (Fig. 3D), which may contribute to modulation of Wnt, RA and Shh levels that underlie neuronal cell type specification (Dessaud et al., 2008; Lara-Ramirez et al., 2013; Lupo et al., 2006; Ulloa and Marti, 2010). Analysis at 44 hpf also clusters progenitor cells based on D-V identity marked by *zic* genes, *ptf1a, lbx1b, dbx1a*, and the proneural genes *ascl1* and *neurog1* (Fig. 4A, B, C; Fig. S6; Fig. S7). The major feature that has emerged by this stage is differentiation to form a number of neuronal cell types that is described below.

Characterization of neuronal complexity

Different neuronal subtypes are progressively generated from the distinct D-V progenitor domains. At 16 hpf, Seurat analysis identifies a single cluster (C6) expressing markers of neurogenesis (Fig. 2A), and at 24 hpf and 44 hpf identifies distinct clusters that express

early and late markers of neuronal differentiation (Fig. 3A, Fig. 4A). To determine whether the transcriptome of differentiating cells is similar or different at 16, 24 and 44 hpf, we aggregated the data from all stages and carried out Seurat analysis. Unsupervised clustering identifies 12 clusters and separates progenitors (C0, C2, C3, C4), progenitors and glia (C1), neurons at different stages of differentiation (C5, C6, C7, C8, C9, C10), and the floor plate (C11) (Fig. 5A; Fig. S8). When cells are labelled by their developmental stage (Fig. 5B), we found that some cells at different stages of neuronal differentiation at 16 hpf overlap with cells at 24 hpf and 44 hpf. Interestingly, they express the Activin-binding protein *fstl1a* (Fig. 5D) as well as transcriptional regulators including *ebf2* (Fig. S8). Likewise, there is some overlap of neurogenesis and neuronal cell types at 24 hpf with differentiating cells at 44 hpf. However, most of the differentiating cells at 24 hpf and 44 hpf are segregated from cells at the earlier stages, consistent with the generation of new neuronal cell types. There are also shifts in the transcriptome of progenitor cells which will be discussed below.

To characterise the neuronal complexity at 44 hpf (Fig. 4A-D), we classified neuronal sub-types based on Hernandez-Miranda and colleagues (Hernandez-Miranda et al., 2017; Lu et al., 2015). Dorsal progenitors expressing *atoh1a* (C9 in Fig. 4A; Fig. 4C) generate dA1 excitatory interneurons (C0) in the hindbrain, a heterogeneous population that functions in sensory information processing (Hernandez-Miranda et al., 2017). *barhl1a*, *barhl2* (Fig. 4C; in situ hybridisation in Fig. 4E; Table S2.3), *lhx2b* and *lhx9* are among their known markers, and in addition we find *alcama*, *bcl11ba* (BAF Chromatin Remodelling Complex), *pdzrn3b* and *scrt1b* (Fig. 4D). Noradrenergic neuron (NAN) development (C10, C11) is marked by expression of *tfapa2a* (Fig. 4C; Table S2.3) which is important for activation of key NA enzymes (Holzschuh et al., 2003; Kim et al., 2001). These cells also express the transcription factors *dmbx1a, lhx1a, lhx5* and *lbx2*. Interestingly, the two clusters of NAN cells are distinguished by expression of several transcription factors, including *lmx1bb, tlx2, phox2a* and *phox2bb* (Fig. 4D; Fig. S6; Table S2.3). Another class of neurons found in the hindbrain are GABAergic inhibitory interneurons (dB4), here clustered in C3 (Fig. 4A). These cells express *pax2, lhx1* and *lhx5* (Fig. 4D, F; Fig. S6; Table S2.3) which may constitute a transcription factor code (Burrill et al., 1997; Gross et al., 2002; Muller et al., 2002; Pillai et al., 2007). A subset of these cells coexpress *otpa/b* (Fig. 4B-D; in situ hybridisation in Fig. 4G; Table S2.3), transcription factors involved in dopaminergic neuron specification (Fernandes et al., 2013), suggesting heterogeneity at this stage. More ventrally (C5), neurons are marked by *tal1* (Fig.4B-D; in situ hybridisation in Fig. 4H) and *gata2a/3* expression (Fig.4D; Table S2.3), resembling ventral neurons identified in the

spinal cord (Andrzejczuk et al., 2018). A further cluster of ventral neurons is C14, which express *vsx1, tal1* and *foxn4* (Fig. 4C, D; in situ hybridisation of *tal1* in Fig. 4H; Table S2.3), defining this domain as V2 interneurons. *vsx1*-expressing cells in the hindbrain and spinal cord have been defined as non-apical progenitors, able to generate one excitatory (V2a) and one inhibitory (V2b) interneuron, and proposed to be a pool important for rapid generation of the sensory-locomotor circuit (McIntosh et al., 2017); their molecular signature is reported in Fig. 4D. Motor neurons can be identified in C12 (*isl1, isl2, phox2a*), and in the hindbrain *lhx4, nkx6.1* and *tbx3a* are further transcription factors expressed in these cells (Fig.4D; Table S2.3). A further neuronal cluster (C6) expresses a specific combination of genes (e.g. *aldocb, calm1b, camk2n1a, rbfox1*; Fig. 4C, D; Table S2.3), but could not be classified. C15 consists of lateral line neuromast cells that were present in the dissected tissue and had not been removed bioinformatically. Our transcriptome atlas thus gives new insights into factors expressed in different neuronal cell types in the hindbrain.

Transcriptional shift of hindbrain progenitors

In addition to finding temporal differences in expression of neurogenic markers, Seurat analysis of the aggregated data found that 16, 24 and 44 hpf progenitors are in largely distinct clusters in UMAP space (Fig. 5A, B). Analysis of the top 30 significant enriched genes per cluster highlights transcriptional similarities and differences between progenitors (Fig. S8). Genes enriched in both dorsal and ventral progenitors at 16 hpf (C0 and C2) and 24 hpf (C3 and C4) include *cldn5a, fsta* and proliferative markers such as *pcna* (Fig. 5C; Fig. S8; Table S2.4). Gene ontology terms associated with the top 30 genes enriched in these progenitors highlight their proliferative property (Fig. 5E). A drastic reduction in proliferation has taken place by 44 hpf. As examples, we show that *mki67, nusap1, ccnd1* and *cdca8* are widely expressed in the early hindbrain, whereas they are restricted to a small proportion of dorsal progenitors and *vsx1*-expressing cells at 44 hpf (Fig. 5D; Fig. S9). In addition, genes associated with cell cycle arrest (*cdkn1ca, cdkn1cb*) and the Notch signalling pathway have increased expression at 44 hpf (Fig. 5E). Glial cells become apparent at 44 hpf in the medio-ventral progenitor pool marked by *fabp7a* (C1), and we find they also express *atp1b4* and *atp1a1b* (Fig. 5D; Fig. S8; Table S2.4). Furthermore, *miR9* loci are detected only at 44 hpf (*miR9.1 CR848047.1, miR9.3 CU929451.2, miR9.6 CU467822.1*) (Fig. 5C, D; Table S2.4), when they are known to play a key role in the timing of neurogenesis (Coolen et al., 2013; Coolen et al., 2012). Overall, this analysis highlights that there are significant temporal changes in gene expression in progenitors between 24 hpf and 44 hpf in the developing hindbrain (Fig. 5F).

Boundary cell and segment centre progenitors

During hindbrain development in zebrafish, proneural gene expression becomes confined to zones flanking the segment boundaries, with low expression in hindbrain boundary cells and also in rhombomere centres (Amoyel et al., 2005; Cheng et al., 2004; Gonzalez-Quevedo et al., 2010). The progenitors at these locations are classified as non-neurogenic since they have low expression of proneural genes required for neuronal differentiation, though this has not been directly shown by lineage analysis (Fig.6A). The inhibition of neurogenesis has been shown to involve Notch activation (Cheng et al., 2004) and Yap/Taz nuclear translocation (Voltes et al., 2019) at boundaries, and Fgf20 signaling at segment centres (Gonzalez-Quevedo et al., 2010). These distinct progenitor populations were not identified by unsupervised clustering, because this is dominated by the large differences in the transcriptome during D-V patterning and differentiation. We therefore used supervised clustering with known markers to reveal the transcriptional signature of the neurogenic and non-neurogenic cell populations.

We bioinformatically isolated 24 hpf ventral progenitors and used *rfng* (boundary), *etv5b* (segment centre), and *neurog1* and *neurod4* (neuronal differentiation) to drive clustering. Three clusters were obtained that are divided into eight sub-clusters (Fig. 6B): C2 corresponds to boundary cells (Fig. 6C), C1, C3 and C6 to segment centres (Fig. 6G), and C0, C4, C5 and C7 to neurogenic cells (Fig. 6Q). The neurogenic cells form a continuum in which there is increasing expression of proneural genes and decreased expression of a proliferation marker *mki67* (Fig. 6Q). We found that boundary cells that express *rfng* (C2; Fig. 6D) also express some previously known markers (Fig. S10; Table S2.5): *rasgef1ba* (Letelier et al., 2018) and the Rho GTPase *rac3b* (Fig.6E; Letelier et al., 2018). In addition, we find new genes with expression enriched at boundaries including *rnd2, prdm8*, *gsx* and *grasp*. We noticed that the BMP inhibitor *follistatin 1b* (Fig.2F; Dal-Pra et al., 2006) is enriched both in segment centres and boundary cells (Fig. S10) and in situ hybridisation analysis confirmed the increased expression at boundaries (Fig. 6F). Thus, we identified a distinct set of factors present in boundary cells with potential functional implications.

At each segment centre, cells respond to Fgf20 signalling and upregulate the Fgf-direct target *etv5b* (Esain et al., 2010; Gonzalez-Quevedo et al., 2010) which we used to drive clustering of 24 hpf progenitors. *etv5b*-expressing cells are found in three adjacent clusters, C1, C3 and C6. In C1 there is transcriptional overlap of *etv5b* with *neurog1,*

ascl1b.1 and *neurod4* (Fig. 6G, Q; Table S2.5) while proneural genes are expressed at a low level in C6 and not detected in C3 cells. The overlapping expression in C1 and C6 likely reflects that at 24 hpf, *etv5b* is expressed in stripes located at the centre of each segment (Fig. 6H; Table S2.5) but neurogenic gene expression has yet to be fully down-regulated (Fig.3H; Gonzalez-Quevedo et al., 2010). Many of the genes co-expressed with *etv5b* (Fig. S10) have an unknown expression pattern, but based on previous work (Gonzalez-Quevedo et al., 2010) we reasoned that all segment centre marker genes would be under Fgf20 control. We therefore performed a bulk RNA-seq experiment comparing wild-type dissected hindbrain to f*gf20a* mutant tissue (Gonzalez-Quevedo et al., 2010) (Fig. S11). *metrnl*, which is present in C1 and C6 (Fig. 6G), was among the downregulated genes in f*gf20a*-/- mutants, and in situ hybridisation reveals is expressed in segment centres (Fig. 6K). The f*gf20a*-/- RNA-seq screen also found *fsta*, which is present in all clusters (Fig. 6G), and in situ hybridisation suggested has complex expression that includes segment centre cells (Fig. 6N). However, *etv5b* was not found in this screen, which likely reflects that it has a complex expression pattern in the hindbrain, otic vesicle and cranial ganglia (Table S3). We therefore also profiled transgenic hindbrains expressing heat-shock induced constitutively activated FgfR1 (Tg(*hsp70:ca-fgfr1*)) and compared to heat-shocked counterparts (Table S4). This screen found *etv5b, metrnl* and *fsta* among the top genes induced by Fgf signalling (Fig. S12). In situ hybridization confirmed that Fgf20 signalling is both necessary and sufficient for expression of *etv5b, metrnl* and *fsta* in segment centres (Fig. 6H-P). *metrnl* encodes a cytokine with an unknown receptor. Since the related meteorin gene (metrn) has been implicated in gliogenesis in other contexts (Lee et al., 2010; Nishino et al., 2004), it is a candidate to promote glial cell differentiation that occurs at segment centres. Interestingly, *fsta* is also expressed by boundary cells, and thus correlates with non-neurogenic progenitors. Overall, we found a limited number of genes that are exclusively expressed by boundary or centre progenitors, while the majority of transcripts are expressed in the two cell populations (Fig. S10), suggesting similarities in their transcriptome.

At 44 hpf, neurogenic zones are fully refined but *rfng* and other boundary cell markers are no longer detected. We therefore only used *etv5b* and *neurog1+neurod4* to drive clustering. At this stage, *etv5b*-expressing cells segregate together in three adjacent subclusters (C2, C6, C7) and the overlap with neurogenic genes has greatly decreased (Fig. 6R-U). *metrn* and *metrnl* are expressed in a similar pattern to *slc1a2b, atp1a1b* and other glial markers, further suggesting that the Metrn family could play a role in hindbrain gliogenesis. Neurogenic cells segregate into two clusters that are further subdivided: C4,

C5 and C0 have a gradient of *neurog1* and *neuroD4* expression, suggestive of the progression of neuronal differentiation, while C3 and C1 express only *neuroD4*, suggestive of late differentiation (Fig. 6R). These latter cells present a unique signature (Fig. 6U) which includes the expression of *fstl1a* (Fig. 5D; Fig. 6T), the transcription factors *scrt1a*, *scrt2* (Fig. 6T ; Fig. 7C, E) and *nhlh2* (Fig. 6T; Table S2.6).

Transcription factors temporally regulating hindbrain neurogenesis

To illustrate developmental insights that can be extracted from the single cell RNA-Seq data, we focused on transcription factors (TFs) (AnimalTFDB3.0 database; Zhang et al., 2012) and inferred their potential contribution to hindbrain neurogenesis. We used the aggregated data set (Fig. 5A) and performed pseudotime analysis using Monocle v3.0.2 (Qiu et al., 2017; Trapnell et al., 2014b), which orders cells uniquely on the similarity of their global TF expression profiles. This created a pseudotime trajectory with three discrete cell states (Fig. S14). The root of the trajectory was defined as the state containing the majority of the 16 hpf progenitor cells. The three states are characterized by the expression of: *sox2*, *egr2b*, *mafba*, *zic* genes, *pax6a/b* and *zbtb16a/b* among others for the progenitor state; *sox3*, *neurog1*, *atoh1a*, *dbx1a*, *gsx1*, *lbx1b* are in the intermediate differentiation state; *atoh1b*, *neurod4*, *isl1*, *vsx1*, *tal1*, *pax2a* and other neuronal transcription factors have high expression level in the final state (Fig. S14). Along the trajectory, cells are ordered largely based on developmental stage of origin and state of differentiation (Fig. 7A, B). 16 hpf and 24 hpf progenitors are found mainly at the start of the trajectory, followed by 44 hpf progenitors. 16 hpf differentiating cells present a TFs expression pattern that mostly resemble 24 hpf progenitors, with the exception of few cells found at the end of the trajectory, while 24 hpf and 44 hpf differentiating cells highly overlap (Fig. 7B). These data further suggest transcriptional changes in early versus late hindbrain progenitors.

To identify the temporal cascade of TFs that may be involved in neurogenic cell-fate decision, we mapped TFs that significantly varied in their pseudo-temporal expression pattern, and clustered them according to their expression dynamic (Fig. 7C). This analysis highlights multiple discrete shifts in TF expression occurring during hindbrain neurogenesis. Seven distinct patterns were identified, where the first has high expression at the beginning of pseudotime, and the others present a progressive shift until reaching a peak of expression of neuronal markers at the end of differentiation. The first group (G1) includes egr2b and mafba, which are genes involved in segmental identity of progenitors that are rapidly down-regulated at the onset of differentiation. In the next group (G2) are

genes expressed in progenitors but not down-regulated until later in pseudotime. These genes have been implicated in the maintenance of the progenitor fate and/or inhibition of neurogenesis. Among them, *zic* and her genes promote neural progenitor identity and inhibit differentation (Bae et al., 2005; Coolen et al., 2012; Nyholm et al., 2007; Scholpp et al., 2009), *id* genes encode negative regulators of proneural bHLH proteins and are abundant in multipotent cells (Ellis et al., 1990; Garrell and Modolell, 1990; Ling et al., 2014) and *zbtb16a* (*plzfa*) inhibits neurogenesis and the encoded protein is degraded in order for neuronal differentiation to progress (Sobieszczuk et al., 2010). The following group of genes (G3) with shifting expression in pseudotime are: *sox3* which has initial constant expression followed by a drop in differentiated cells; *neurog1* (reviewed by Bertrand et al., 2002); *prdm12b,* a regulator of V1 interneuron fate decision (Thélie et al., 2015; Zannino et al., 2014); and *foxp4* that is progressively expressed during neuronal differentiation and promotes detachment of differentiating cells from the neuroepithelium (Rousso et al., 2012). *atoh1b* and *neurod4* are found in the next step of the cascade (G4) together with *ebf2*, a factor that acts downstream of proneural genes and necessary for initiation of migration toward the mantle layer and neuronal differentiation (Garcia-Dominguez et al., 2003). In the next group, a subset of genes initiates expression that then increases late in pseudotime (G5). They include *zbtb20* that functions during corticogenesis as temporal regulator for the generation of layer-specific neuronal subtypes (Tonchev et al., 2016), and the less studied *uncx, nhlh2, lhx4* and *sox12*. Furthermore, members of the zebrafish scratch family (*scrt1a/b/2*) has a similar dynamic pattern and they show enrichment within the neurogenic zone with some dorso-ventral differences: *scrt1a* and *scrt1b* are expressed ventrally and dorsally (Fig. 7D-E) while *scrt2* is only found ventrally (Fig. 7F). These genes have been implicated in the onset of neuronal migration (Itoh et al., 2013; Paul et al., 2014). Followed by a group of TFs with a later onset of expression that does not decline (G6), these factors are implicated in neuronal specification (*otpa, tal1, pax2a*). The final group of genes with an onset of expression late in pseudotime (G7) also encode regulators of neuronal identity (*isl1/2a, gata3, lhx1a/5/9*; Fig. S14).

To further explore TFs role in hindbrain neurogenesis we used a complementary approach that does not relay on pseudotemporal ordering. A genetic regulatory network (GRN) was created using GENIE3 (Huynh-Thu et al., 2010), which uses a Random Forest machine-learning algorithm to predict the strength of putative regulatory links between a target gene and the expression pattern of input genes (i.e. transcription factors). Since there have been extensive studies of gene regulation during hindbrain segmentation (Parker and

Krumlauf, 2017), we tested whether GENIE3 finds known interactions. We analysed the transciptome data from 16 hpf and focussed on a module that includes regulators of segmentation and A-P identity (Fig. S15). We find potential interactions between egr2b, mafba, hox genes and epha4a, which include seven interactions that have been verified in vivo (asterisks in Fig. S15). A GRN was produced for each individual stage (16 hpf Table S5.1, 24 hpf Table S5.2, 44 hpf Table S5.3), and we present findings for 44 hpf since these are more relevant for late steps of neurogenesis. To focus on the predictions with higher significance, we applied a threshold of >0.025 of important measure (IM) and these interactions were analysed in Cytoscape (Shannon et al., 2003) (Table S5). This cut-off recovered 4637 total interactions that constitute a valuable resource to guide future *in vivo* functional validations. Given the complexity of the network we extracted a submodule to exemplify its predictive potential. We interrogated the network to specifically predict the role of *scrt* genes during neurogenesis, and extracted their closest neighbours (Fig. 7G). This network module predicts interconnections between genes in G5a, G5a and G4. *scrt1a* and *scrt2* are found in a feedback loop with *nhlh2*, and upstream of neurogenic factors (*neurod4, elavl3, otpa/b*, and *pax2a*), while *scrt1b* is connected to *atoh1a/b, atoh8*, and *barhl1a/b*.

DISCUSSION

The single cell transcriptome atlas that we present here is a resource for further investigation of mechanisms that regulate neurogenesis and other aspects of hindbrain development. We analysed the transcriptome of hindbrain cells prior to (16 hpf), during (24 hpf) and after (44 hpf) the patterning of neurogenesis to form discrete neurogenic and non-neurogenic zones within segments. We used unbiased methods to cluster cells based on transcriptional differences, and identified genes that mark distinct hindbrain segments, cell types along the D-V axis, and neuronal differentiation. By comparing our findings with previous studies, we have created an annotated list of genes that indicates which are previously known and which are novel markers, as also highlighted in the relevant Results section.

Seurat analysis at 16 hpf clustered cells based on segment-specific gene expression and gave a global picture of differences in the transcriptome of distinct segments. The organisation of clusters from r2 to r6 suggests that neighbouring segments have a similar transcriptome, but with a significant difference between odd- and even-numbered segments. This is consistent with previous studies showing nested expression of *hox* genes that regulate anterior-posterior identity (reviewed by Alexander et al., 2009; Tümpel et

al., 2009), and the role of *egr2* in regulating gene expression in r3 and r5 that confers distinct properties from r2, r4 and r6 (Voiculescu et al., 2001). In contrast, r7 cells do not cluster adjacent to r6 cells, suggestive of a distinct identity which may reflect that it is a transitional zone to the anterior spinal cord.

We find major differences in gene expression in differentiating neurons at 16 hpf and 24 hpf compared with 44 hpf, as expected from the generation of distinct neuronal cell types at different times. Our analyses reveal new genes that are co-expressed with known markers of neuronal cell types that form along the D-V axis. In addition to transcription factors, these include modulators of the Shh, RA and Wnt pathways. Interestingly, many differentiating neurons at all stages express *fstl1a*, suggesting a potential role of BMP inhibition. The generation of different neuronal cell types at 44 hpf compared with 16 hpf and 24 hpf is accompanied by changes in gene expression in progenitor cells at these stages, including proliferation markers and miR9 microRNAs. By carrying out pseudotime analysis, we inferred progressive changes in gene expression during the differentiation of progenitor cells to neurons. These data suggest a cascade in which genes that define segmental identity are rapidly down-regulated, followed by factors that maintain progenitor cells, in turn followed by upregulation of genes required for neuronal migration and transcription factors that define neuronal identity. We also analysed transcription factor expression using an algorithm to predict gene regulatory networks. We focussed on *scrt* family genes that regulate neuronal migration, and this found potential relationships with proneural factors and regulators of neuronal identity. We envisage that investigators can interrogate the network for other TFs of interest to guide biological hypotheses and phenotypic screening of specific mutants.

One motivation for this study was to find genes that mark the distinct neurogenic and non-neurogenic zones that are established in the zebrafish hindbrain. These features are not found in the unbiased analysis, as this is dominated by the greatest transcriptomic differences. We therefore used known markers of hindbrain boundary cells, neurogenic cells and segment centres to drive clustering of the progenitor population. In addition, we carried out RNA-seq analyses after manipulation of Fgf pathway activation which inhibits neurogenesis at segment centres. These analyses identified novel signaling factors, most notably follistatin and meteorin family members expressed in boundary cells and/or segment centres that are candidates to inhibit neurogenesis or promote gliogenesis. The single cell transcriptome data will enable investigators to extract information on other specific cell populations by this approach.

MATERIALS AND METHODS

Maintenance of zebrafish strains and husbandry

Zebrafish embryos were raised at 28.5°C or 25°C depending on the required stage (Westerfield, 2007). Embryos were staged according to hour post fertilization (hpf) and morphological criteria (Kimmel et al., 1995). The zebrafish work was carried out under a UK Home Office Licence under the Animals (Scientific Procedures) Act 1986 and underwent full ethical review.

Mutant strains and heat shock treatment

fgf20a (dob) mutant embryos (Whitehead et al., 2005) were obtained from homozygous mutant in-crosses. Transgenic Tg(*hsp70:ca-fgfr1*) embryos are heterozygotes from outcrosses (Gonzalez-Quevedo et al., 2010; Marques et al., 2008). To induce constitutively active Fgfr1, Tg(*hsp70:ca-fgfr1*) embryos at 22 hpf were heat shocked for 30 min at 38.5°C and then incubated for 2 h at 28.5°C. Since around 50% of the embryos are carrying the transgene, controls and treated embryos were collected from the same heat-shocked clutch, avoiding any issue with differences in genomic background and changes in gene expression due to the heat shock treatment. After mRNA extraction, qPCR was performed to identify properly dissected tissues and discriminate between controls and *fgfr1* over-expressing tissues.

Whole-mount in situ hybridization

For whole-mount in situ hybridization, embryos or explants were fixed in 4% PFA overnight at 4°C, or 4 h at room temperature, and kept in methanol at -20°C prior to processing. Some probes have been previously described: *neurog1* and *neurod4* (Alexander et al., 2009; Gonzalez-Quevedo et al., 2010), *pax2* (Krauss et al., 1991), *rfng* (Cheng et al., 2004), *etv5b* (cb805, ZFIN), *metrnl* (MPMGp609H2240Q8, RZPD), *sox3* (EST clone: IMAGp998H108974Q). Additional probes were generated from cDNA of 20-44 hpf embryos. A forward primer was used together with a reverse primer with a T7 promoter site (5'gaaatTAATACGACTCACTATAGg3') for amplification; see Table 1. Digoxigenin-UTP labelled riboprobes were synthesised and *in situ* hybridization performed as previously described (Xu et al., 1994). After BCIP/NBT colour development, embryos were re-fixed for 30 min, cleared in 70% glycerol/PBS, and mounted to view the dorsal or lateral side. For each gene at least two independent replicates were performed using more then 30 embryos each time. For transverse sections, embryos were extensively washed in PBST prior to mounting in 4% agarose/water. Embryos were sectioned using a Vibratome (Lecia

VT1000 S), generating transverse sections of a thickness of 40 µm. Imaging was carried out with a Zeiss Axioplan2 with an Axiocam HRc camera.

Hindbrain dissection

Embryos at the desired stage were decorionated and de-yolked in DMEM with high Glucose, no Glutamine, no Calcium (11530556, Gibco); hindbrains were micro-dissected using 0.33 mm micro-fine sterile needles. Dissected tissues were kept in DMEM until further processed. For RNA-seq a single hindbrain tissue was collected in an individual tube and the quality of the dissection evaluated by qPCR (data not shown). For scRNA-seq, around 40 tissues per stage were pooled and immediately processed for cell dissociation.

RNA extraction, cDNA preparation and qPCR

RNA was isolated using Quick-RNA Microprep kit (Zymo Research) and eluted in 15 µl (Lan et al., 2009). To evaluate the quality of dissection 3 µl of RNA was reverse transcribed using SuperScript™ III Reverse Transcriptase (ThermoFisher Scientific), and the remainder stored at -80°C until processed. Primers for target genes were designed with PrimerQuest (IDT). qPCR was performed using QuantStudio 3 (ThermoFisher Scientific) with SYBR green Platinum™ SYBR™ Green qPCR SuperMix-UDG (ThermoFisher Scientific) master mix. The ΔΔCt method was used to calculate gene expression (Livak and Schmittgen, 2001). ß-actin was used as reference gene. Primers used are listed in Table 2. Samples without contamination were processed for RNA-seq.

Library Preparation and RNA-sequencing

Libraries for the *fgf20a-/-* experiment were prepared with the Ovation® RNA-Seq System V2 (7102, NuGEN) for cDNA amplification, followed by NexteraXT (Illumina) for library preparation. These libraries were sequenced on the HiSeq 2000 (Illumina), with paired-end 75 bp reads. Libraries for the constitutive active Fgfr1 experiment were prepared with the Clontech SMARTer kit (634926, TaKaRa) for cDNA amplification, followed by NexteraXT (Illumina) for library preparation. These libraries were sequenced on the HiSeq 4000 (Illumina), with single ended 75 bp reads.

Sequence alignment and analysis of differentially expressed genes

The quality of the samples was assessed using FastQC. Reads were aligned against zebrafish genome GRCz10 and Ensembl release 86 transcript annotations using STAR v2.5.1b (Dobin et al., 2013) via the transcript quantification software RSEM v1.2.31 (Li and Dewey, 2011). Gene-level counts were rounded to integers and subsequently used for

differential expression analysis with DESeq2 v1.20.0 (Anders and Huber, 2010) using default settings. Differential expression results were thresholded for significance based on an FDR<=0.01, a fold-change of +/- 2 and a minimum normalized count of >30 in all contributing samples from at least one of the replicate groups being compared. Heatmaps were created using rlog transformed count data, scaled across samples using a z-score.

Preparation of single cells from zebrafish hindbrain

Around 40 hindbrain tissues per stage (16 hpf, 24 hpf, 44 hpf) were dissected as described above. The samples were incubated with FACS max cell dissociation solution (T200100, Amsbio) supplemented with 1mg/ml Papain (10108014001, Sigma) for 25 min at 37°C and resuspended once during incubation. Cells were then transferred to HBSS (no calcium, no magnesium, no phenol red; 11140035, ThermoFisher Scientific) supplemented with 5%FBS, Rock inhibitor (Y-27632, Stem Cell Technologies) and 1X non-essential amino acids (11140035, ThermoFisher Scientific). Cells were further disaggregated by pipetting and filtered several times using 20 µm strainers (130-101-812, Miltenyi Biotech GmbH). To access quality live/cell death, cell size and number of clumps were measured. Samples with a viability above 65% were used for single cell sequencing. During protocol optimization, qPCR was carried out to check that gene expression levels are similar in dissociated cells and the intact hindbrain.

10X Genomics single-cell library preparation

A suspension of 10,000 single cells was loaded onto the 10X Genomics Single Cell 3' Chip. cDNA synthesis and library construction were performed according to the manufacturers protocol for the Chromium Single Cell 3' v2 protocol (PN-120233, 10X Genomics). cDNA amplification involved 12 PCR cycles. Samples were sequenced on Illumina HiSeq 4000 using 100 bp paired-end runs.

Bioinformatic analysis of scRNA-seq data

The 10X Cell Ranger software was used to de-multiplex Illumina BCL output, create fastq files and generate single cell feature counts for each library using a transcriptome built from the zebrafish Ensembl release 89, GRCz10.

Seurat unsupervised analysis of aggregated data

Three 10X libraries representing the 16 hpf, 24 hpf and 44 hpf stages of embryonic development were aggregated using the 10X software “cellranger aggr” function, which sub-samples reads such that all libraries have the same effective sequencing depth.

Aggregated count data were further analysed using the Seurat v3.1.0 (Butler et al., 2018b) package within R v3.6.1.

Cell quality was assessed using some simple QC metrics: library size, total number of expressed genes and mitochondrial RNA content. Outlier cells were flagged if they were above/below 3 median absolute deviations (MADs) from the median for any metric in a dataset specific manner.

Data were normalised across cells using the "LogNormalize" function with a scale factor of 10,000. A set of genes highly variable across cells was identified using the "FindVariableGenes" function (selection.method = "vst", nfeatures = 2000). Data were centred and scaled using the "ScaleData" function with default parameters.

PCA analysis was performed on the scaled data using the variant genes. Significant principle components were identified by manual inspection of the top loading genes and by plotting the standard deviations of the top 100 components.

The first 30 principal components were used to create a Shared Nearest Neighbour (SNN) graph using the "FindNeighbours" function (k.param=20). This was used to find clusters of cells showing similar expression using the FindClusters function (resolution = 0.8).

The Uniform Manifold Approximation and Projection (UMAP) dimensional reduction technique was used to visualize data from the first 30 principal components in 2 dimensional space ("RunUMAP function). Graphing of the output enabled visualization of cell cluster identity and marker gene expression.

Visual inspection of hindbrain and non-hindbrain marker genes suggested some clusters were constituted by contaminant non-hindbrain cells; see Supplementary File1 for a list of valid hindbrain cells. A new iteration of the analysis was then performed as above, this time excluding contaminant cells from the aggregated data prior to normalisation, variable gene selection, data scaling and dimension reduction (PC1-30) and cluster identification (resolution = 0.8).

Biomarkers of each cluster were identified using Wilcoxon rank sum tests using Seurat's "FindAllMarkers" function. It was stipulated that genes must be present in 10% of the cells in a cluster and show a logFC of at least 0.25 to be considered for testing. Only positive markers were reported. The expression profile of top markers ranked by average logFC were visualised as heatmaps and dotplots of the scaled data. Cluster identity was determined using visual inspection focusing on the expression of known marker genes.

Seurat unsupervised analysis of individual stages

Count data for individual stages were loaded directly into Seurat from the 10X results files separately, without aggregation. Downstream analysis was conducted as for the aggregated dataset. For each stage dataset the first 30 principal components were used for cluster identification. Differing resolutions were passed to the "FindClusters" function based on how well the resultant clusters corroborated known marker gene expression: 16 hpf (resolution=0.7), 24 hpf (resolution=1.2), 44 hpf (resolution=1.0). The 16 hpf data were further analysed at higher resolution and also using PlotClusterTree in Seurat.

Seurat supervised clustering of ventral progenitors from individual stages

For each stage, cells identified as being ventral progenitors in the aggregate analysis were subset and subjected to supervised clustering using custom sets of marker genes to drive PCA analysis, cluster identification and UMAP dimensional reduction. For 24 hpf ventral progenitor cells, the genes used were *rfng* (boundary), *etv5b* (segment centre) and *neurog1*, *neurod4* (neuronal differentiation). For 44 hpf ventral progenitor cells, the list was restricted to *etv5b*, *neurog1* and *neurod4*.

Pseudotime analysis of aggregated dataset using Monocle3

Pseudotime analysis was conducted using the Bioconductor package Monocle v3.0.2 (Trapnell et al., 2014a). Count data from the individual stages were combined. The "preprocess_cds" function was used to normalise the data to address sequencing depth differences before PCA dimensional reduction (n=50). The three datasets were then aligned by fitting a linear model to the cells PCA co-ordinates and subtracting a "stage" effect ("align_cds" function: num_dim = 50, alignment_group = "stage"). Next, the data were subjected to UMAP dimensional reduction and cell clustering ("cluster_cells": resolution=0.001). A principal graph was plotted through the UMAP using the "learn_graph" function, representing the path through development. The graph was in

turn used to order cells through the developmental programme as pseudotime using sox3 positive 16hpf cells as the start of the programme.

Genes changing as a function of pseudotime were determined using graph-auto-correlation analysis ("graph_test" function). Selected genes listed as being transcription factors in the AnimalTFDB3.0 database were presented on a heatmap of expression over pseudotime.

GENIE3 inference of regulatory networks

The Bioconductor package GENIE3 v1.4.3 (Huynh-Thu et al., 2010) was used to infer regulatory networks of genes within cells of individual developmental stages. For each stage, an expression matrix of raw gene counts, with non-hindbrain cells removed, was constructed and passed to the GENIE3 function together with a list of zebrafish transcription factors identified in the AnimalTFB3.0 database (targets = NULL, treeMethod = "RF", K = "sqrt", nTrees = 1000) in order to create a weighted adjacency matrix. The weights describe the likelihood of a regulator-gene / target-gene link being genuine. This matrix was converted to a table of regulatory links (regulator-gene, target-gene, link-weight). Regulator/target links with weights > 0.025 (data available in Supp. File 12) were visualised as an interaction directed network within Cytoscape (Shannon et al., 2003).

ACKNOWLEDGEMENTS

We thank Andreas Sagner and Julien Delile for advice, and Qiling Xu for comments on the manuscript. We are also grateful to the Francis Crick Institute Advanced Sequencing platform and Aquatics facility for their excellent support. This work was supported by the Francis Crick Institute which receives its core funding from Cancer Research UK (FC001217), the UK Medical Research Council (FC001217), and the Wellcome Trust (FC001217).

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHOR CONTRIBUTIONS

Conceptualisation: M.T., D.G.W; Methodology: M.T., R.M., D.G.W.; Software: R.M.; Formal Analysis: R.M.; Investigation: M.T.; Writing – original draft: M.T., D.G.W.; Writing – review and editing: M.T., R.M., D.G.W.; Supervision: D.G.W.; Funding acquisition: D.G.W.

DATA AVAILABILITY

Single cell RNA sequencing and bulk RNA sequencing data have been deposited in in Gene Expression Omnibus under accession number GSE141428. Single cell RNA sequencing data are available at the Single Cell Portal https://singlecell.broadinstitute.org/single_cell/study/SCP667/a-single-cell-transcriptome-atlas-of-the-developing-zebrafish-hindbrain#study-summary, and the R analysis script developed for this paper is available at https://github.com/crickbabs/ZebrafishDevelopingHindbrainAtlas.

REFERENCES

Addison, M., et al. (2018). Cell Identity Switching Regulated by Retinoic Acid Signaling Maintains Homogeneous Segments in the Hindbrain. *Dev. Cell* **45**, 606-620 e3.

Alexander, T., et al. (2009). Hox genes and segmentation of the hindbrain and axial skeleton. *Annu. Rev. Cell Dev. Biol.* **25**, 431-56.

Amoyel, M., et al. (2005). Wnt1 regulates neurogenesis and mediates lateral inhibition of boundary cell specification in the zebrafish hindbrain. *Development* **132**, 775-85.

Anders, S., et al. (2010). Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106.

Andrzejczuk, L. A., et al. (2018). Tal1, Gata2a, and Gata3 Have Distinct Functions in the Development of V2b and Cerebrospinal Fluid-Contacting KA Spinal Neurons. *Front. Neurosci.* **12**, 170.

Bae, Y. K., et al. (2005). Patterning of proneuronal and inter-proneuronal domains by hairy- and enhancer of split-related genes in zebrafish neuroectoderm. *Development* **132**, 1375-85.

Batlle, E., et al. (2012). Molecular mechanisms of cell segregation and boundary formation in development and tumorigenesis. *Cold Spring Harb. Perspect. Biol.* **4**, a008227.

Becht, E., et al. (2018). Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.*

Bertrand, N., et al. (2002). Proneural genes and the specification of neural cell types. *Nat. Rev. Neurosci.* **3**, 517-30.

Burrill, J. D., et al. (1997). PAX2 is expressed in multiple spinal cord interneurons, including a population of EN1+ interneurons that require PAX6 for their development. *Development* **124**, 4493-4503.

Butler, A., et al. (2018a). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.*

Butler, A., et al. (2018b). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411-420.

Cayuso, J., et al. (2019). Actomyosin regulation by Eph receptor signaling couples boundary cell formation to border sharpness. *bioRxiv*, 683631.

Cayuso, J., et al. (2015). Mechanisms of boundary formation by Eph receptor and ephrin signaling. *Dev. Biol.* **401**, 122-31.

Cheng, Y. C., et al. (2004). Notch activation regulates the segregation and differentiation of rhombomere boundary cells in the zebrafish hindbrain. *Dev. Cell* **6**, 539-50.

Clarke, J. D., et al. (1993). Segmental repetition of neuronal phenotype sets in the chick embryo hindbrain. *Development* **118**, 151-62.

Cooke, J. E., et al. (2005). EphA4 is required for cell adhesion and rhombomere-boundary formation in the zebrafish. *Curr. Biol.* **15**, 536-42.

Coolen, M., et al. (2013). miR-9: a versatile regulator of neurogenesis. *Front. Cell. Neurosci.* **7**, 220.

Coolen, M., et al. (2012). miR-9 controls the timing of neurogenesis through the direct inhibition of antagonistic factors. *Dev. Cell* **22**, 1052-64.

Dal-Pra, S., et al. (2006). Noggin1 and Follistatin-like2 function redundantly to Chordin to antagonize BMP activity. *Dev Biol* **298**, 514-26.

Delile, J., et al. (2019). Single cell transcriptomics reveals spatial and temporal dynamics of gene expression in the developing mouse spinal cord. *Development* **146**.

Dessaud, E., et al. (2008). Pattern formation in the vertebrate neural tube: a sonic hedgehog morphogen-regulated transcriptional network. *Development* **135**, 2489-503.

Dessaud, E., et al. (2007). Interpretation of the sonic hedgehog morphogen gradient by a temporal adaptation mechanism. *Nature* **450**, 717-20.

Dobin, A., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21.
Ellis, H. M., et al. (1990). extramacrochaetae, a negative regulator of sensory organ development in Drosophila, defines a new class of helix-loop-helix proteins. *Cell* **61**, 27-38.
Elsen, G. E., et al. (2008). Zic1 and Zic4 regulate zebrafish roof plate specification and hindbrain ventricle morphogenesis. *Dev. Biol.* **314**, 376-92.
Elsen, G. E., et al. (2009). The autism susceptibility gene met regulates zebrafish cerebellar development and facial motor neuron migration. *Dev. Biol.* **335**, 78-92.
Esain, V., et al. (2010). FGF-receptor signalling controls neural cell diversity in the zebrafish hindbrain by regulating olig2 and sox9. *Development* **137**, 33-42.
Fagotto, F. (2014). The cellular basis of tissue separation. *Development* **141**, 3303-18.
Fernandes, A. M., et al. (2013). Orthopedia transcription factor otpa and otpb paralogous genes function during dopaminergic and neuroendocrine cell specification in larval zebrafish. *PLoS One* **8**, e75002.
Fisher, A., et al. (1998). The function of hairy-related bHLH repressor proteins in cell fate decisions. *Bioessays* **20**, 298-306.
Garcia-Dominguez, M., et al. (2003). Ebf gene function is required for coupling neuronal differentiation and cell cycle exit. *Development* **130**, 6013-25.
Garrell, J., et al. (1990). The Drosophila extramacrochaetae locus, an antagonist of proneural genes that, like these genes, encodes a helix-loop-helix protein. *Cell* **61**, 39-48.
Geling, A., et al. (2003). bHLH transcription factor Her5 links patterning to regional inhibition of neurogenesis at the midbrain-hindbrain boundary. *Development* **130**, 1591-604.
Gonzalez-Quevedo, R., et al. (2010). Neuronal regulation of the spatial patterning of neurogenesis. *Dev Cell* **18**, 136-47.
Gouti, M., et al. (2015). The route to spinal cord cell types: a tale of signals and switches. *Trends Genet.* **31**, 282-9.
Grinblat, Y., et al. (2001). zic Gene expression marks anteroposterior pattern in the presumptive neurectoderm of the zebrafish gastrula. *Dev. Dyn.* **222**, 688-93.
Gross, M. K., et al. (2002). Lbx1 specifies somatosensory association interneurons in the dorsal spinal cord. *Neuron* **34**, 535-49.
Guillemot, F. (2007). Spatial and temporal specification of neural fates by transcription factor codes. *Development* **134**, 3771-80.
Guthrie, S., et al. (1991). Patterns of cell division and interkinetic nuclear migration in the chick embryo hindbrain. *J. Neurobiol.* **22**, 742-54.
Guthrie, S., et al. (1991). Formation and regeneration of rhombomere boundaries in the developing chick hindbrain. *Development* **112**, 221-9.
Hernandez-Miranda, L. R., et al. (2017). The dorsal spinal cord and hindbrain: From developmental mechanisms to functional circuits. *Dev. Biol.* **432**, 34-42.
Heyman, I., et al. (1995). Cell and matrix specialisations of rhombomere boundaries. *Dev. Dyn.* **204**, 301-15.
Hikasa, H., et al. (2013). Wnt signaling in vertebrate axis specification. *Cold Spring Harb. Perspect. Biol.* **5**, a007955.
Holzschuh, J., et al. (2003). Noradrenergic neurons in the zebrafish hindbrain are induced by retinoic acid and require tfap2a for expression of the neurotransmitter phenotype. *Development* **130**, 5741-54.
Huynh-Thu, V. A., et al. (2010). Inferring regulatory networks from expression data using tree-based methods. *PLoS One* **5**.
Ikeya, M., et al. (1997). Wnt signalling required for expansion of neural crest and CNS progenitors. *Nature* **389**, 966-70.
Itoh, Y., et al. (2013). Scratch regulates neuronal migration onset via an epithelial-mesenchymal transition-like mechanism. *Nat. Neurosci.* **16**, 416-25.

Kageyama, R., et al. (2005). Roles of bHLH genes in neural stem cell differentiation. *Exp. Cell Res.* **306**, 343-8.
Kim, H. S., et al. (2001). Regulation of the tyrosine hydroxylase and dopamine beta-hydroxylase genes by the transcription factor AP-2. *J. Neurochem.* **76**, 280-94.
Kimmel, C. B., et al. (1995). Stages of embryonic development of the zebrafish. *Dev. Dyn.* **203**, 253-310.
Krauss, S., et al. (1991). Expression of the zebrafish paired box gene pax[zf-b] during early neurogenesis. *Development* **113**, 1193-206.
Lan, C. C., et al. (2009). Quantitative real-time RT-PCR (qRT-PCR) of zebrafish transcripts: optimization of RNA extraction, quality control considerations, and data analysis. *Cold Spring Harb Protoc* **2009**, pdb prot5314.
Lara-Ramirez, R., et al. (2013). Retinoic acid signaling in spinal cord development. *Int. J. Biochem. Cell Biol.* **45**, 1302-13.
Lee, H. S., et al. (2010). Meteorin promotes the formation of GFAP-positive glia via activation of the Jak-STAT3 pathway. *J. Cell Sci.* **123**, 1959-68.
Lee, K. J., et al. (1999). The specification of dorsal cell fates in the vertebrate central nervous system. *Annu. Rev. Neurosci.* **22**, 261-94.
Letelier, J., et al. (2018). Evolutionary emergence of the rac3b/rfng/sgca regulatory cluster refined mechanisms for hindbrain boundaries formation. *Proc. Natl. Acad. Sci. U. S. A.*
Li, B., et al. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323.
Liem, K. F., Jr., et al. (1997). A role for the roof plate and its resident TGFbeta-related proteins in neuronal patterning in the dorsal spinal cord. *Cell* **91**, 127-38.
Ling, F., et al. (2014). Id proteins: small molecules, mighty regulators. *Curr. Top. Dev. Biol.* **110**, 189-216.
Livak, K. J., et al. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402-8.
Lu, D. C., et al. (2015). Molecular and cellular development of spinal cord locomotor circuitry. *Front. Mol. Neurosci.* **8**, 25.
Lumsden, A. (2004). Segmentation and compartition in the early avian hindbrain. *Mech. Dev.* **121**, 1081-8.
Lumsden, A., et al. (1989). Segmental patterns of neuronal development in the chick hindbrain. *Nature* **337**, 424-8.
Lupo, G., et al. (2006). Mechanisms of ventral patterning in the vertebrate nervous system. *Nat. Rev. Neurosci.* **7**, 103-14.
Marques, S. R., et al. (2008). Reiterative roles for FGF signaling in the establishment of size and proportion of the zebrafish heart. *Dev Biol* **321**, 397-406.
Mcinnes, L., et al. (2018). UMAP: Uniform Manifold Approximation and Projection. *Journal of Open Source Software* **3**.
Mcintosh, R., et al. (2017). Spatial distribution and characterization of non-apical progenitors in the zebrafish embryo central nervous system. *Open Biol* **7**.
Miyake, A., et al. (2012). Neucrin, a novel secreted antagonist of canonical Wnt signaling, plays roles in developing neural tissues in zebrafish. *Mech. Dev.* **128**, 577-90.
Muller, T., et al. (2002). The homeodomain factor lbx1 distinguishes two major programs of neuronal differentiation in the dorsal spinal cord. *Neuron* **34**, 551-62.
Narita, Y., et al. (2009). Chapter 5 Hox Genes in Neural Patterning and Circuit Formation in the Mouse Hindbrain. *Hox Genes*, pp. 139-167.
Nishino, J., et al. (2004). Meteorin: a secreted protein that regulates glial cell differentiation and promotes axonal extension. *EMBO J.* **23**, 1998-2008.
Nyholm, M. K., et al. (2007). The zebrafish zic2a-zic5 gene pair acts downstream of canonical Wnt signaling to control cell proliferation in the developing tectum. *Development* **134**, 735-46.

Ortega, S., et al. (1998). Neuronal defects and delayed wound healing in mice lacking fibroblast growth factor 2. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 5672-7.
Panhuysen, M., et al. (2004). Effects of Wnt1 signaling on proliferation in the developing mid-/hindbrain region. *Mol. Cell. Neurosci.* **26**, 101-11.
Parker, H. J., et al. (2017). Segmental arithmetic: summing up the Hox gene regulatory network for hindbrain development in chordates. *Wiley Interdiscip Rev Dev Biol* **6**.
Paul, V., et al. (2014). Scratch2 modulates neurogenesis and cell migration through antagonism of bHLH proteins in the developing neocortex. *Cereb. Cortex* **24**, 754-72.
Peretz, Y., et al. (2016). A new role of hindbrain boundaries as pools of neural stem/progenitor cells regulated by Sox2. *BMC Biol.* **14**, 57.
Pierfelice, T., et al. (2011). Notch in the vertebrate nervous system: an old dog with new tricks. *Neuron* **69**, 840-55.
Pillai, A., et al. (2007). Lhx1 and Lhx5 maintain the inhibitory-neurotransmitter status of interneurons in the dorsal spinal cord. *Development* **134**, 357-66.
Qiu, X., et al. (2017). Single-cell mRNA quantification and differential analysis with Census. *Nat Methods* **14**, 309-315.
Rousso, D. L., et al. (2012). Foxp-mediated suppression of N-cadherin regulates neuroepithelial character and progenitor maintenance in the CNS. *Neuron* **74**, 314-30.
Scholpp, S., et al. (2009). Her6 regulates the neurogenetic gradient and neuronal identity in the thalamus. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 19895-900.
Shannon, P., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-504.
Sobieszczuk, D. F., et al. (2010). A feedback loop mediated by degradation of an inhibitor is required to initiate neuronal differentiation. *Genes Dev.* **24**, 206-18.
Stuart, T., et al. (2018). Comprehensive integration of single cell data. *bioRxiv*, 460147.
Terriente, J., et al. (2012). Signalling from hindbrain boundaries regulates neuronal clustering that patterns neurogenesis. *Development* **139**, 2978-87.
Thélie, A., et al. (2015). Prdm12 specifies V1 interneurons through cross-repressive interactions with Dbx1and Nkx6 genes in Xenopus. *Development* **142**, 3416-3428.
Timmer, J. R., et al. (2002). BMP signaling patterns the dorsal and intermediate neural tube via regulation of homeobox and helix-loop-helix transcription factors. *Development* **129**, 2459-72.
Tiso, N., et al. (2009). Differential expression and regulation of olig genes in zebrafish. *J. Comp. Neurol.* **515**, 378-96.
Tonchev, A. B., et al. (2016). Zbtb20 modulates the sequential generation of neuronal layers in developing cortex. *Mol. Brain* **9**, 65.
Trapnell, C., et al. (2014a). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* **32**, 381-386.
Trapnell, C., et al. (2014b). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* **32**, 381-6.
Tümpel, S., et al. (2009). Chapter 8 Hox Genes and Segmentation of the Vertebrate Hindbrain. *Hox Genes*, pp. 103-137.
Ulloa, F., et al. (2010). Wnt won the war: antagonistic role of Wnt over Shh controls dorso-ventral patterning of the vertebrate neural tube. *Dev. Dyn.* **239**, 69-76.
Vaccarino, F. M., et al. (1999). Changes in cerebral cortex size are governed by fibroblast growth factor during embryogenesis. *Nat. Neurosci.* **2**, 848.
Voiculescu, O., et al. (2001). Hindbrain patterning: Krox20 couples segmentation and specification of regional identity. *Development* **128**, 4967-78.
Voltes, A., et al. (2019). Yap/Taz-TEAD activity links mechanical cues to progenitor cell behavior during zebrafish hindbrain segmentation. *Development* **146**.
Wang, Q., et al. (2017). Cell Sorting and Noise-Induced Cell Plasticity Coordinate to Sharpen Boundaries between Gene Expression Domains. *PLoS Comput. Biol.* **13**, e1005307.

Westerfield, M. (2007). *The Zebrafish Book*. Place, Published, Eugene, Oregon: University of Oregon.
Whitehead, G. G., et al. (2005). fgf20 is essential for initiating zebrafish fin regeneration. *Science* **310**, 1957-60.
Xu, Q., et al. (1995). Expression of truncated Sek-1 receptor tyrosine kinase disrupts the segmental restriction of gene expression in the Xenopus and zebrafish hindbrain. *Development* **121**, 4005-16.
Xu, Q., et al. (1994). Spatially regulated expression of three receptor tyrosine kinase genes during gastrulation in the zebrafish. *Development* **120**, 287-299.
Zannino, D. A., et al. (2014). prdm12b specifies the p1 progenitor domain and reveals a role for V1 interneurons in swim movements. *Dev Biol* **390**, 247-60.
Zhang, H. M., et al. (2012). AnimalTFDB: a comprehensive animal transcription factor database. *Nucleic Acids Res* **40**, D144-9.
Zheng, W., et al. (2004). Fibroblast growth factor 2 is required for maintaining the neural stem cell pool in the mouse brain subventricular zone. *Dev. Neurosci.* **26**, 181-96.

Figures

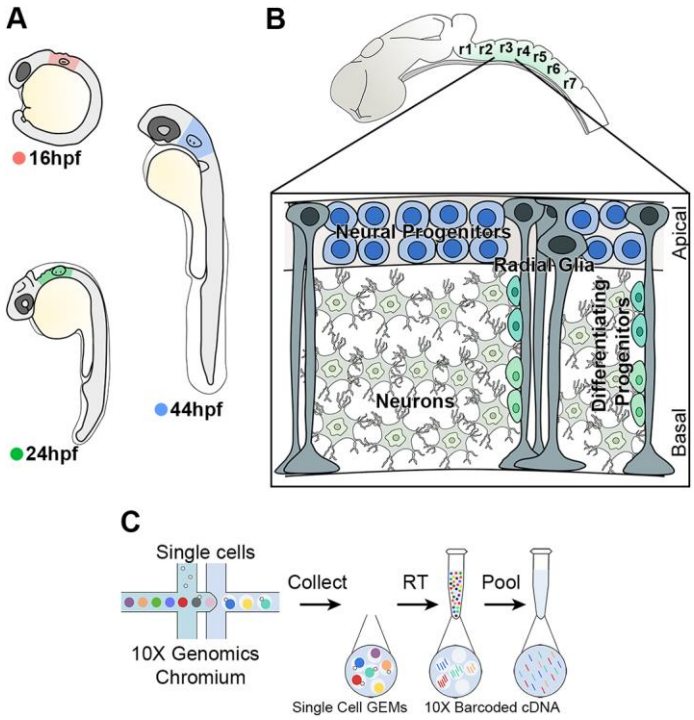


Figure 1. High throughput scRNA-seq strategy from the developing hindbrain.
(A) The hindbrain of 16 hpf (pink), 24 hpf (green) and 44 hpf (blue) embryos was collected for scRNA-seq. (B) Drawing of zebrafish hindbrain with a zoomed-in view of the stereotypical hindbrain cell composition at 44 hpf. Progenitors and radial glia cell bodies occupy the ventricular region, while differentiating progenitors and neurons are in the mantle zone. (C) Schematic of the 10X Genomics Chromium workflow.

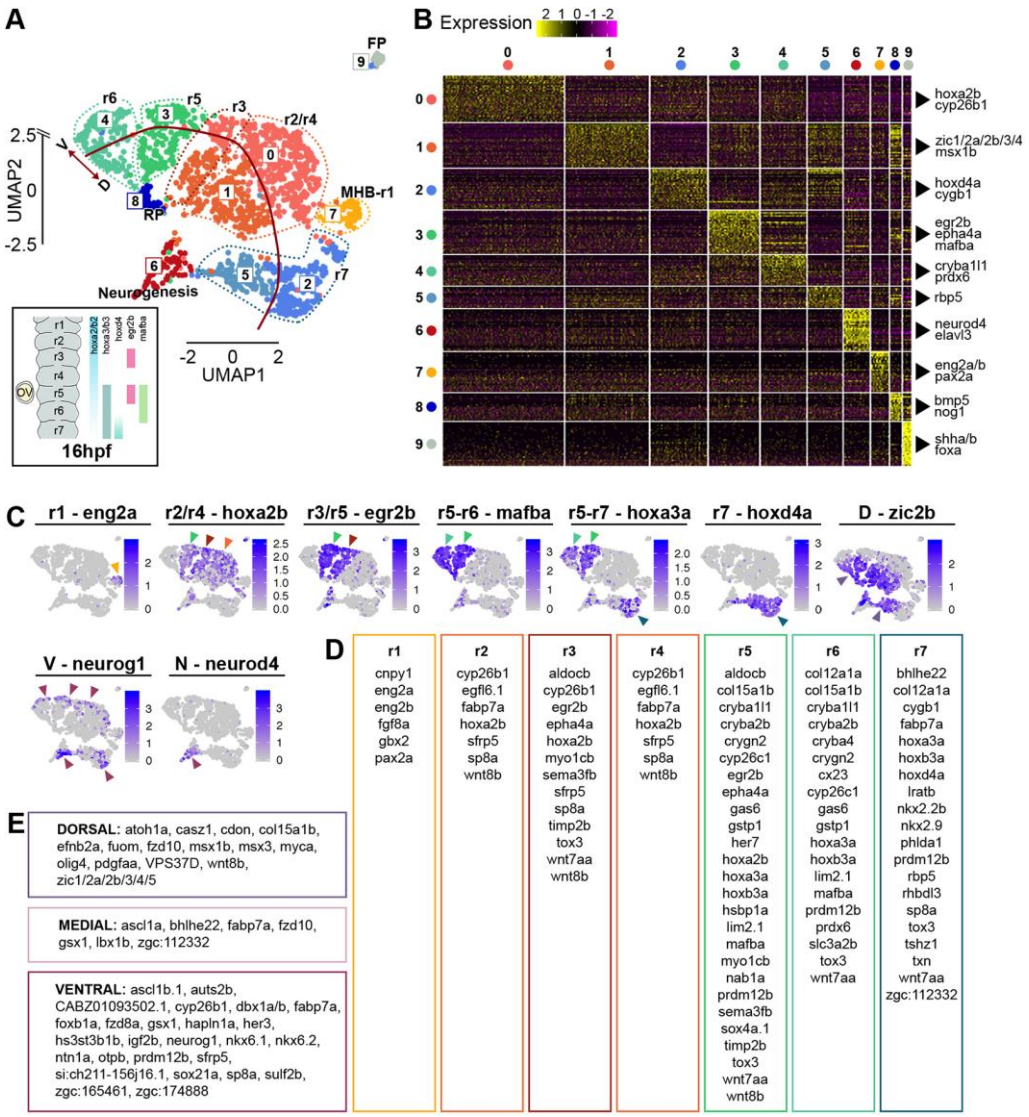


Figure 2. Cell population composition and signatures of the 16 hpf hindbrain

(A) Unsupervised UMAP plot subdivides hindbrain cells into 10 clusters (C0-C89). Dotted lines segregate different rhombomeres (r), midbrain-hindbrain boundary (MHB); floor plate (FP), roof plate (RP) and cells undergoing neurogenesis are also highlighted. The red line separates dorsal versus ventral cells. UMAP2 (y-axis) is discontinuous. Below the UMAP, a schematic view of the zebrafish hindbrain at 16 hpf and selected segmental genes. (B) Heatmap of the top 30 genes significantly enriched in each cluster; representative gene names are shown close to each cluster. The full gene list is in Fig. S3. (C) UMAP plots showing the log normalised counts of representative genes. Colour intensity is proportional to the expression level. Arrow heads point to relevant domain of expression, colour refers to cluster of origin. (D) Summary of rhombomere-specific genes extracted from the top 30 significantly enriched. (E) Summary of genes restricted along the D-V axis.

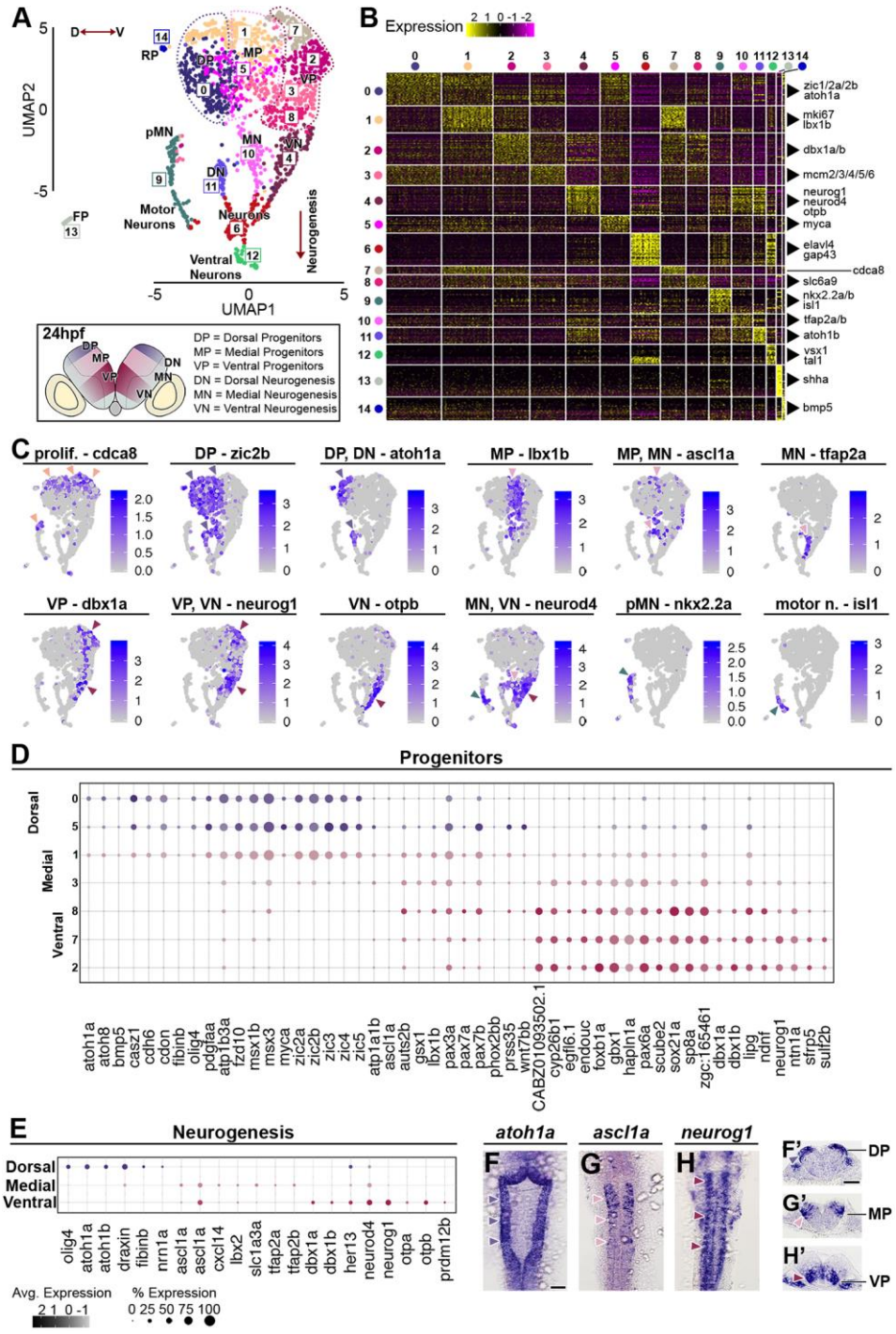


Figure 3. Cell population composition and signatures of the 24 hpf hindbrain

(A) Unsupervised UMAP plot subdivides hindbrain cells into 15 clusters. Dotted lines segregate dorsal (dark violet), medial (pink) and ventral (maroon) progenitors, red arrowed lines indicate the D-V axis and the direction of neurogenesis. Below the UMAP, a schematic drawing of a representative transverse section of a 24 hpf zebrafish hindbrain at the level of the otic vesicle (DP = Dorsal Progenitors, MP = Medial Progenitors, VP = Ventral Progenitors, pMN = progenitors Motor Neurons, DN = Dorsal Neurogenesis, MN =

Medial Neurogenesis, VN = Ventral Neurogenesis, FP = Floor Plate, RP = Roof Plate). (B) Heatmap of the top 30 genes significantly enriched in each cluster; representative gene names are shown close to each cluster. The full gene list is in Fig. S5. (C) UMAP plots showing the log normalised counts of selective representative genes. Colour intensity is proportional to the expression level of a given gene. Arrow heads point to relevant domain of expression, colour refers to cluster of origin. (D) Dot Plot of genes with dorso-ventral restricted expression in progenitors. (E) Dot Plot of factors with restricted expression in differentiating progenitors. Dot size corresponds to the percentage of cells expressing the feature in each cluster, while the colour represents the average expression level. Whole mount in situ hybridization showing the expression pattern of *atoh1a* (F, F'), *ascl1a* (G, G') and *neurog1* (H, H'). (F-H) Dorsal view and (F'-H') 40 µm hindbrain transverse section at the level of r4-r5/r5-r6. Scale bar: 50 µm.

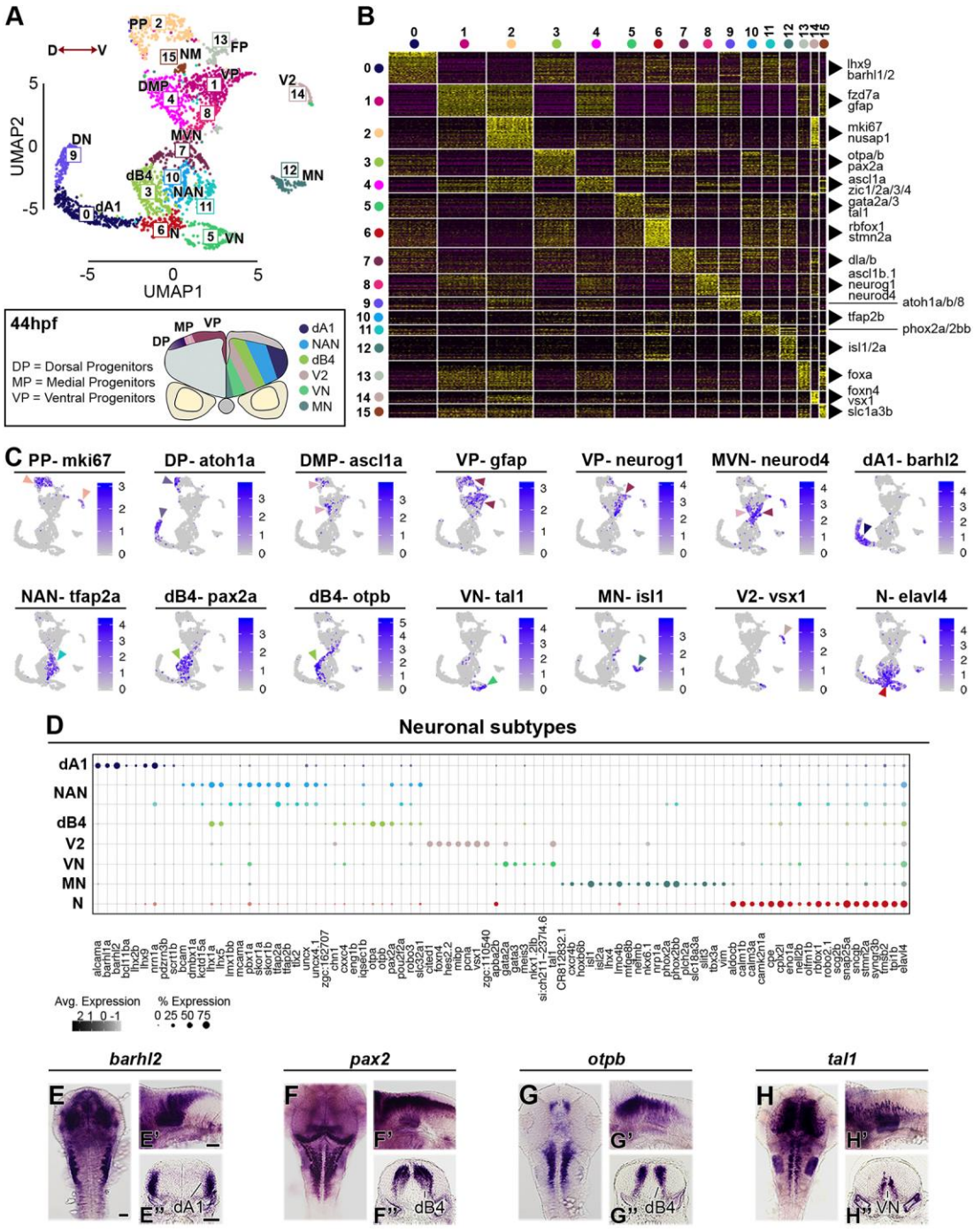


Figure 4. Neuronal complexity of the 44 hpf hindbrain

(A) Unsupervised UMAP plot subdivides cells into 16 clusters. Red arrowed lines indicate the D-V axis. Below the UMAP, a schematic drawing of a representative transverse section of a 44 hpf zebrafish hindbrain at the level of the otic vesicle (PP = Proliferative Progenitors, DMP = Dorso-Medial Progenitors, VP = Ventral Progenitors, MVN = Medio-Ventral Neurogenesis, DN = Dorsal Neurogenesis, dB4 = GABAergic interneurons, NAN = NorAdrenergic Neurons, dA1 = dorsal neurons, N = neurons, VN = Ventral Neurons, V2 = Interneurons, MN = Motor Neurons, FP = Floor Plate, NM = Neuromast). (B) Heatmap of the

top 30 genes significantly enriched in each cluster, representative gene names are shown close to each cluster. For the full gene list refer to Fig. S6. (C) UMAP plots showing the log normalised counts of selective representative genes. Colour intensity is proportional to the expression level of a given gene. Arrow heads point to relevant domain of expression, colour refers to cluster of origin. (D) Dot Plot showing neuronal subtype molecular signature. Dot size corresponds to the percentage of cells expressing the feature in each cluster, while the colour represents the average expression level. Whole mount in situ hybridization showing the expression pattern of *barhl2* (E, E'), *pax2* (F, F'), *otpb* (G, G') and *tal1* (H-H'). (E-H) Dorsal view, (E'-H') lateral view and (E''-H'') 40 µm hindbrain transverse section at the level of r4-r5/r5-r6. Scale bar: 50 µm.

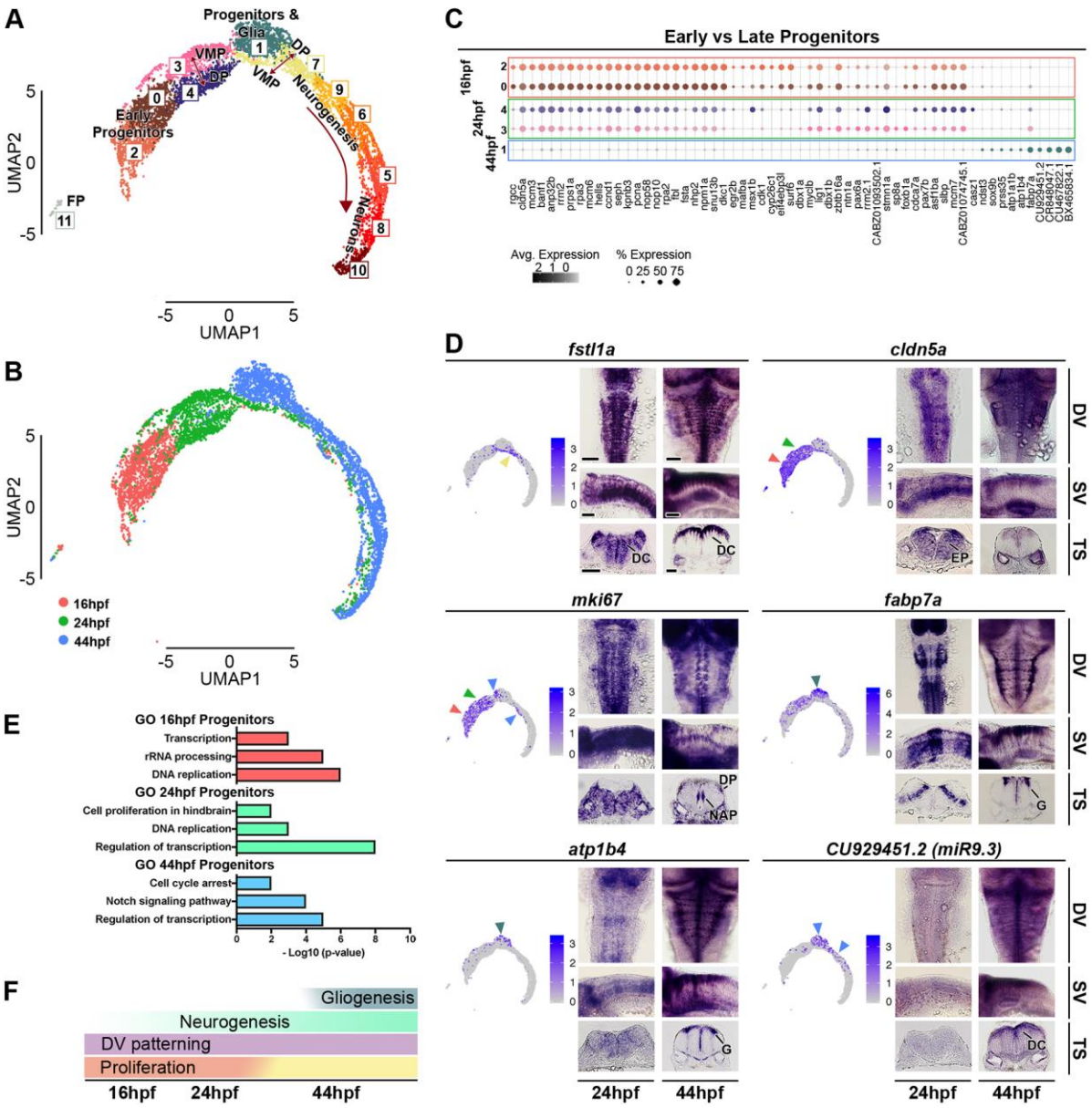


Figure 5. Analysis of aggregated 16 hpf, 24 hpf and 44 hpf data

(A) Unsupervised UMAP plot of cells from 16 hpf, 24 hpf and 44 hpf subdivides them into 12 clusters (DP = Dorsal Progenitors, VMP = Ventral-Medial Progenitors, FP = Floor Plate). (B) Dot Plot showing molecular signature of dorsal and ventral progenitors at the three stages. Dot size corresponds to the percentage of cells expressing the feature in each cluster, while the colour represents the average expression level. The full gene list of top 30 significantly enriched factors is in Fig. S8. (C) UMAP plots with cells coloured based on their stage of origin: 16 hpf (pink), 24 hpf (green) and 44 hpf (blue). (D) UMAP plots showing the log normalised counts of representative genes. Colour intensity is proportional to the expression level of a given gene. Whole mount in situ hybridization showing the expression pattern of *cldn5a*, *fstl1a*, *mki67*, *fabp7a*, *atp1b4* and *CU929451.2 (miR9.3)* at

24 hpf and 44 hpf. Arrow heads point to relevant domain of expression, colour refers to cluster of origin. Dorsal view (DV), side view (SV) and 40 µm hindbrain transverse section (TS) at the level of r4-r5/r5-r6 are shown for each gene. Scale bar: 50 µm. EN = Early Neurogenesis, EP = Early Progenitors, DC = Differentiating Cells, DP = Dorsal Progenitors, NAP = Non-Apical Proliferation, G = Glia. (E) Selected Gene Ontology (GO) terms at 16 hpf (pink), 24 hpf (green) and 44 hpf (blue) are shown. X-axis is -log10(p-value). (F) Summary of global hindbrain changes along the temporal axis.

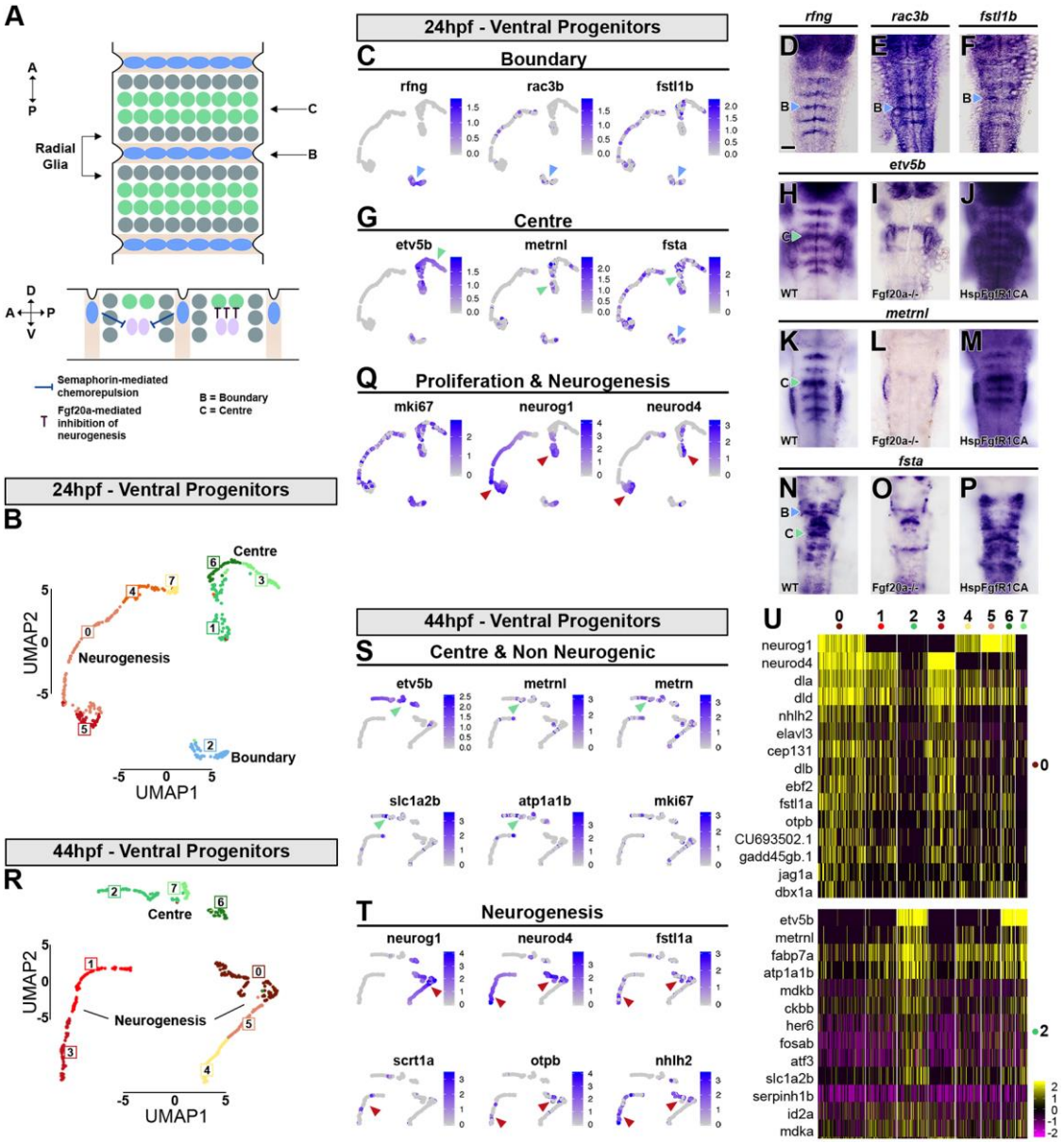


Figure 6. Transcriptional signature of boundary cells and segment centre progenitors

(A) Schematic drawing representing anterior-posterior organization within hindbrain segments. Boundary cells are in cyan, neurogenic progenitors in grey and segment centre cells in green. Below is a side view showing the role of boundary cells in maintaining fgf20a neurons (pink) at the centre of each segment, mediated by semaphorins. Fgf20 signaling maintains undifferentiated progenitors. (B) Supervised clustering of 24 hpf ventral progenitors. 8 clusters are identified: C7, C0 are progenitors; C5 is the neurogenic domains; C5 are boundary cells; and C1, C6 and C3 are segment centre progenitors. UMAP plot showing the expression distribution of boundary (C), segment centre (G) and proliferation and neurogenic genes (Q). Arrow heads point to relevant domain of expression, colour refers to cluster of origin. Whole mount in situ hybridization of

boundary (D-F) and segment centre genes (H, K, N). Segment centre-specific gene expression is dependent on Fgf20 signalling, as fgf20a-/- embryos have loss of *etv5b* (I), *metrnl* (L) and *fsta* (O) expression, whereas constitutive activation of FgfR1 induces their ectopic expression (J, M, P). (R) Supervised clustering of 44 hpf ventral progenitors. 8 clusters are identified: C4, C5 are progenitors; C0, C1, C3 are neurogenic domains; C2, C7, C6 are segment centre progenitors. UMAP plot showing the expression distribution of segment centre and non-neurogenic genes (S) and neurogenic genes (T). Arrow heads point to relevant domain of expression, colour refers to cluster of origin. (U) Heatmap of the top 15 genes enriched in each cluster.

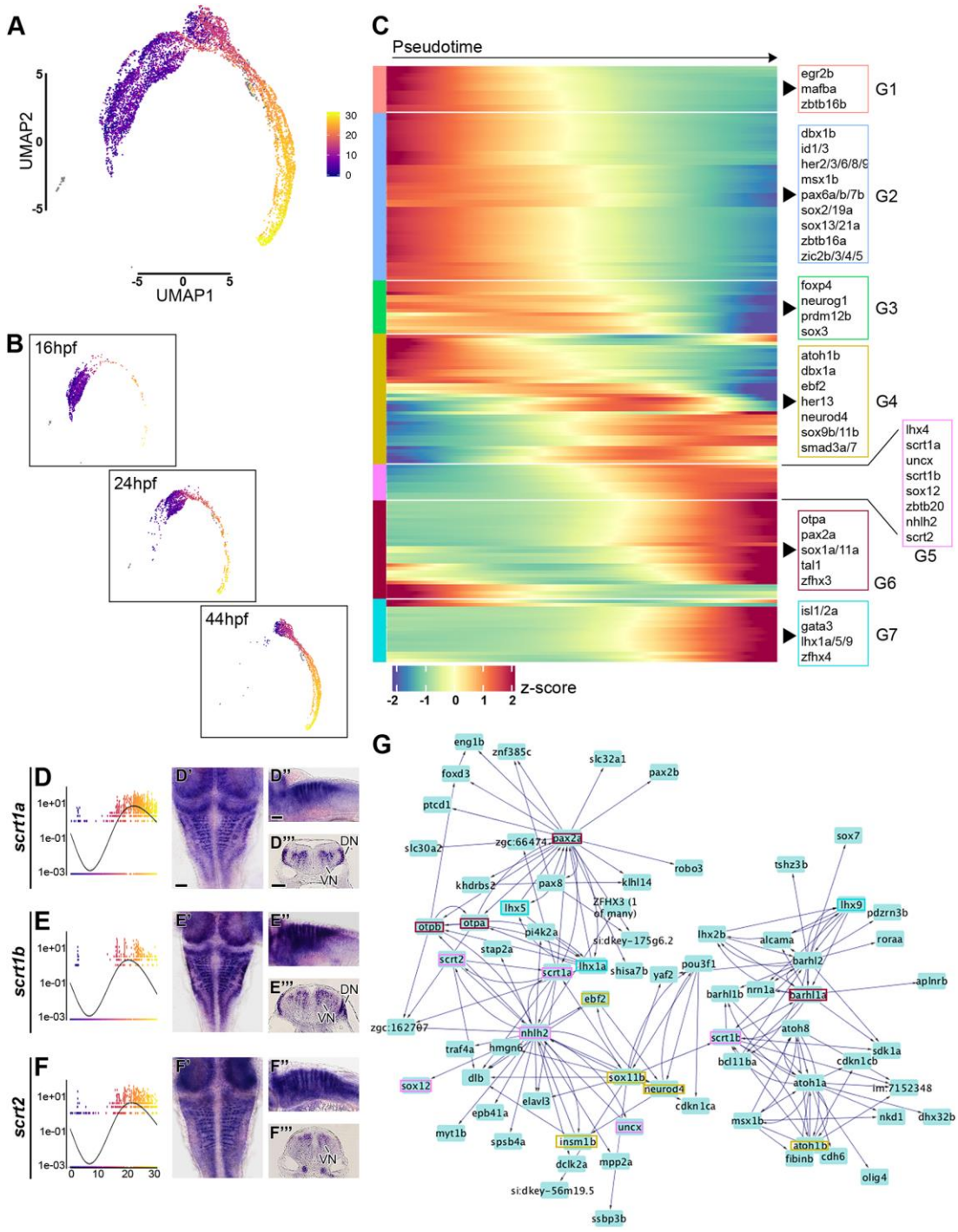


Figure 7. Analysis of transcription factor expression during hindbrain neurogenesis

(A) Monocle3 pseudo-temporal ordering of 16 hpf, 24 hpf and 44 hpf hindbrain cells superimposed onto the aggregate UMAP. Cells are coloured based on their progression along pseudotemporal space (from pseudotime 0 in violet to end of differentiation in yellow). (B) Individual pseudotemporal plots representing cells distribution at each developmental stage. (C) Heatmap showing selected TFs clustered by pseudotemporal expression pattern (q values<0.01). Pseudotime ordering goes from left (progenitor state)

to right (differentiated neurons). Selected transcription factors are shown for each group (G1-G7). The full gene list is in Supplementary Figure 13. (D-F) Expression of *scrt1a*, *scrt1b* and *scrt2* during pseudotime. Whole mount in situ hybridization at 44 hpf for *scratch* genes is shown in dorsal view (D'-F'), side view (D''-F'') and hindbrain sections (D'''-F'''). Scale bar: 50 µm. VN = Ventral Neurogenesis, DN = Dorsal Neurogenesis. (G) Using GENIE3, a directed network of interactions was predicted among the genes in the 44 hpf scRNA-seq data set. The *scratch* genes network was viewed and extracted in Cytoscape; boxes highlight TFs present in the above heatmap, and colours match the group of origin in (C).

Tables

Table 1 – Primer sequences for antisense probe generation

Gene Name	Primer sequences 5’ to 3’
Atoh1a	Fw CCAACGTCGTGCAGAAA Rw gaaatTAATACGACTCACTATAGgAACCCATTACAAAGCCCAGATA
Ascl1a	Fw CAAAGAGCCAAGGGACTAAGAG Rw gaaatTAATACGACTCACTATAGgCCCAGCATTGTAAAGGCAAAG
Barhl2	Fw GCCACCTCCTCCTTTCTAATC Rw gaaatTAATACGACTCACTATAGgGCTGTCCACGGTTCCTAATAA
Otpb	Fw CTCACGGGCTCATACAACTATT Rw gaaatTAATACGACTCACTATAGgGACGCAGGTGTCAACAATTTAG
Tal1	Fw GCGGAACAGTATGGGATGTAT Rw gaaatTAATACGACTCACTATAGgCTGGAATGGTGTAGTCCTCTTG
Cldn5a	Fw AGCAGACAACCTGACCAAAG Rw gaaatTAATACGACTCACTATAGgTGGCACAAGCACGAAGAT
Fstl1a	Fw CCGCCGTACCATTGAGAAA Rw gaaatTAATACGACTCACTATAGgAGCAGTGTGGTCATCCTTTAC
Mki67	Fw AGCCAGAAGATGCCAAACTTA Rw gaaatTAATACGACTCACTATAGgGGACTACCTCACCAGCACTAAAC
Fabp7a	Fw GCAATGTTACCAAACCCACAAT Rw gaaatTAATACGACTCACTATAGgACAAAGGCAGGCCTCAATAA
Atp1b4	Fw GCCATGTTTGCTGGTTGTATG Rw gaaatTAATACGACTCACTATAGgGTGTCGTGTTGGACGTTAAGA
CU929451.2	Fw TGCCTCAGCAGTGTCTAAAG Rw gaaatTAATACGACTCACTATAGgTCAGACACATTTGGTAGCTTCA
Rac3b	Fw CAATGTGATGGTGGATGGTAAAC Rw gaaatTAATACGACTCACTATAGgACCCAACCTGTGAGAGTAGTA
Fstl1b	Fw CAGTCCAGTCGTGTGTTATGT Rw gaaatTAATACGACTCACTATAGgTGTGCTGGTCTTCATCTTCTC
Fsta	Fw CTGTGGTCCTGGAAAGAGATG Rw gaaatTAATACGACTCACTATAGgGACTCATCTTTGCATCCCATAAAC
Plp1a	Fw ATGCTCTGCCTTCAGCTTATC Rw gaaatTAATACGACTCACTATAGgCATGGAAACCAACCCTCTCTAC
Her4.4	Fw CCGCCGTACCATTGAGAAA Rw gaaatTAATACGACTCACTATAGgAGCAGTGTGGTCATCCTTTAC
Rtca	Fw GCTGAAATGGCACCTCAAATAG Rw gaaatTAATACGACTCACTATAGgCCTGTTCGCATTCTGGATGTA
Dusp1	Fw CTGAGGTGATCTTGCCAGTATT Rw gaaatTAATACGACTCACTATAGgGACAATCCCTGAGCAACCTATAA
Zbtb18	Fw ATCCACCTCAGCACACATTT Rw gaaatTAATACGACTCACTATAGgCCCACTCTTACCTTCACCTTTC
Ebf2	Fw GTCATGGGTCTCAGCTCTTATC Rw gaaatTAATACGACTCACTATAGgTGGCAACCTCCTCACAATC
Atp1a1b	Fw GACCATCCCATCACTGCTAAA Rw gaaatTAATACGACTCACTATAGgCCTCGTACGCCAGAGAAATAG
Ptf1a	Fw CACAGGCTTAGACTCTTTCTCC Rw gaaatTAATACGACTCACTATAGgCCCGTAGTCTGGGTCATTTG
Prdm8	Fw TCGCTCCTTGTGGACTAATG Rw gaaatTAATACGACTCACTATAGgCTGGCTTCTGTTGGTTGATTG
Nusap1	Fw AACTGTCCTCACCACCAATAAA Rw gaaatTAATACGACTCACTATAGgGACAAACGAGACGAAAGCTAAAC
Ccnd1	Fw CGAGCTCCAGCTTTCTTACTT Rw gaaatTAATACGACTCACTATAGgGCCAGATCCCACTTCAGTTTAT
Cdc8a	Fw CACCGCTGAAGTCTACAATGA Rw gaaatTAATACGACTCACTATAGgGACGGGTACAGCACAAGAATA

Table 2 – qPCR primers

Gene Name	Primer sequences 5' to 3'	Species	Marker Region
ß-actin	Fw CGAGCTGTCTTCCCATCCA Rw TCACCAACGTAGCTGTCTTT	Zebrafish	Housekeeping gene
Otx2	Fw CAAGCAACCACCTTACACGG Rw TCGTCTCTGCTTTCGAGGAG	Zebrafish	Anterior head
Egr2b (Krox20)	Fw GGACATTACGAGCAGATAAACG Rw CTGCTGGAGTAGGCTAAGTCG	Zebrafish	Hindbrain
Mafba	Fw AGCGTTTGATGGATACAGGG Rw TGGTGTTGATGGTGATGGTG	Zebrafish	Hindbrain
Hoxb2a	Fw CAGAGATTCAAGGTGGACTCG Rw AGTAGCTGCGTGTTGGTATAC	Zebrafish	Hindbrain
Etv5b	Fw CTCTTTCAAGACCTCAGCCAG Rw GCTCATCTCCCTCTTTATTTTCG	Zebrafish	Hindbrain, FGF readout
Hoxb6a	Fw GGGAAAAGCATCTACCCTGA Rw CGACCAGCGTTACCGAAG	Zebrafish	Spinal Cord
xFgfR1	Fw CTGCTCTATCAGTTGCCCG Rw CCCAGTTGATGCTCTGAACA	Xenopus	Heat Shock Tg(hsp70:ca-fgfr1)

SUPPLEMENTARY FIGURES

Fig.S1

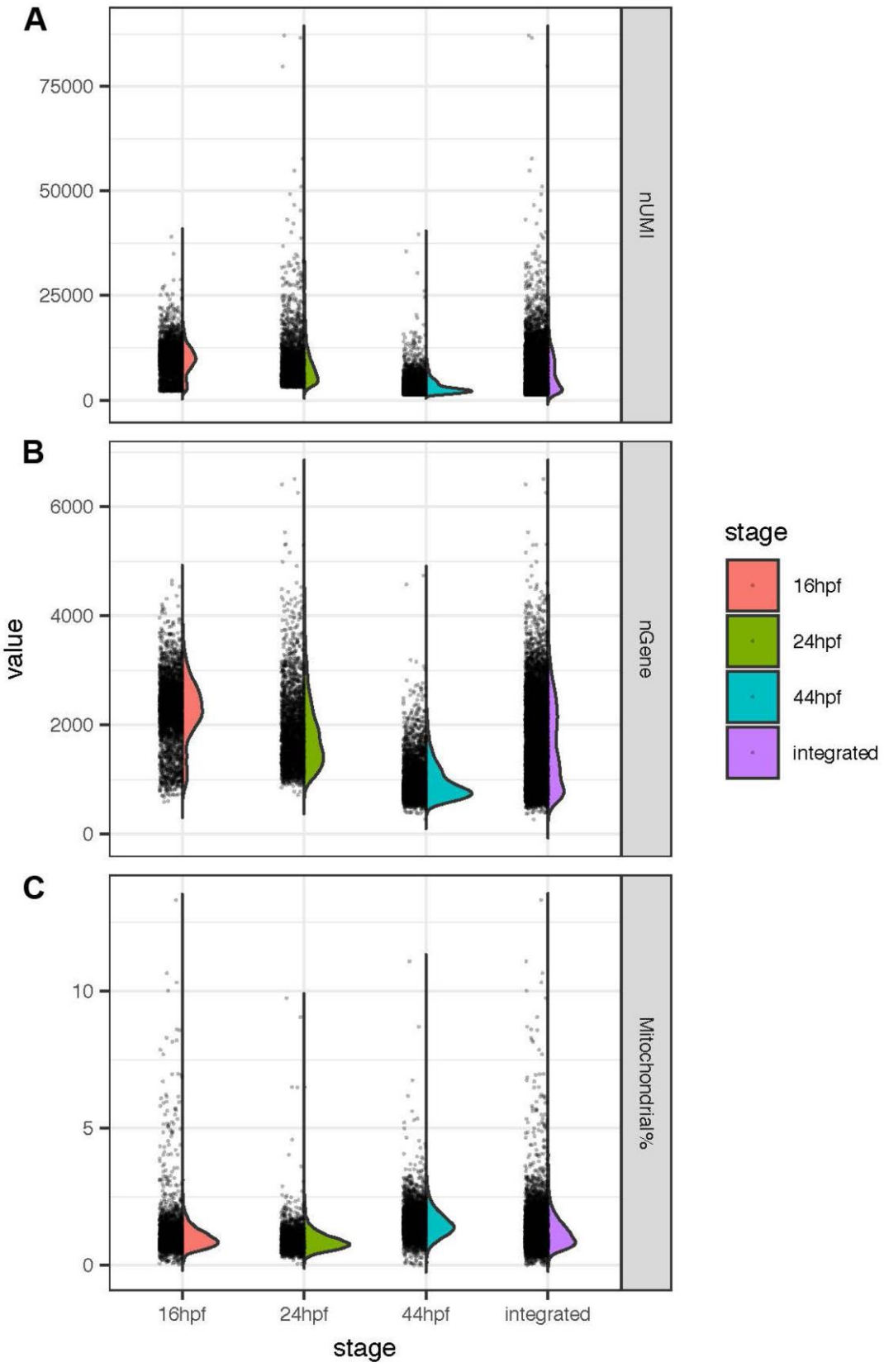


Figure S1. Quality Control matrix.

(A) Distribution of number of Unique Molecular Identifiers (UMIs), (B) number of genes per cell (nGene), and (C) percentage of mitochondrial reads are shown for 16 hpf (pink), 24 hpf (green), 44 hpf (blue) and the aggregated data set (violet).

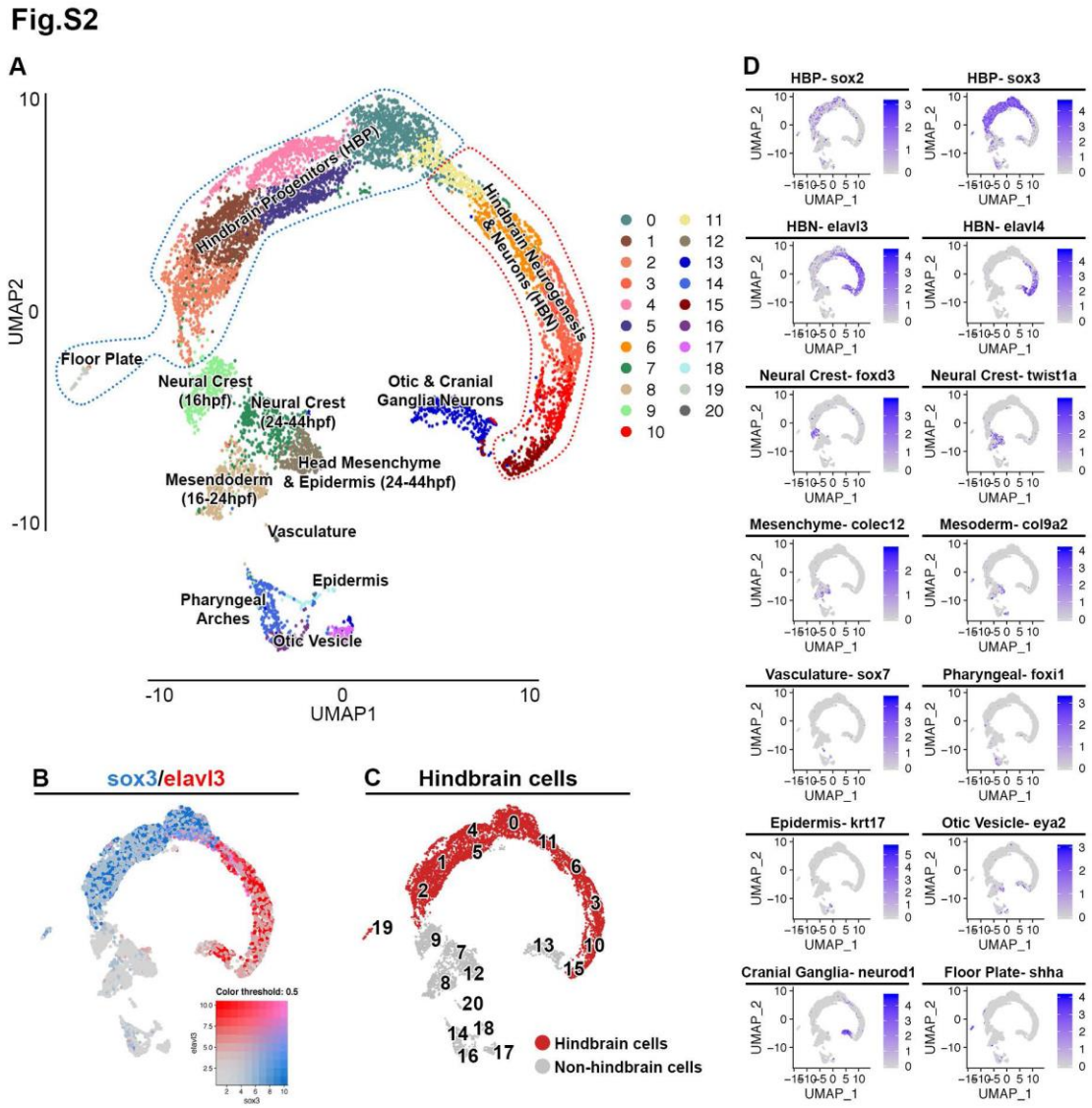


Figure S2. Mapping of the hindbrain and surrounding tissues.

(A) UMAP representation of the aggregated data (16 hpf, 24 hpf and 44 hpf), where the clustering of cells depicts their transcriptional similarity. (B) UMAP plot showing the expression distribution of *sox3* (blue) and *elavl3* (red). (C) High levels of *sox3* and/or *elavl3* expression demarcate the hindbrain territory. Cluster identity (A) was defined based on expression of known marker genes (*sox2*, *sox3* = hindbrain progenitors HBP; *elavl3*, *elavl4* = hindbrain neurons HBN; *foxd3*, *twis1* = neural crest; *colec12* = head mesenchyme; *col9a2* = mesoderm; *sox7* = vasculature; *foxi1* = pharyngeal arches; *krt17* = epidermis; *eya2* = otic vesicle; *neurod1* = cranial ganglia; *shha* = floor plate) (D). Colour intensity is proportional to the expression level of a given gene.

Fig.S3

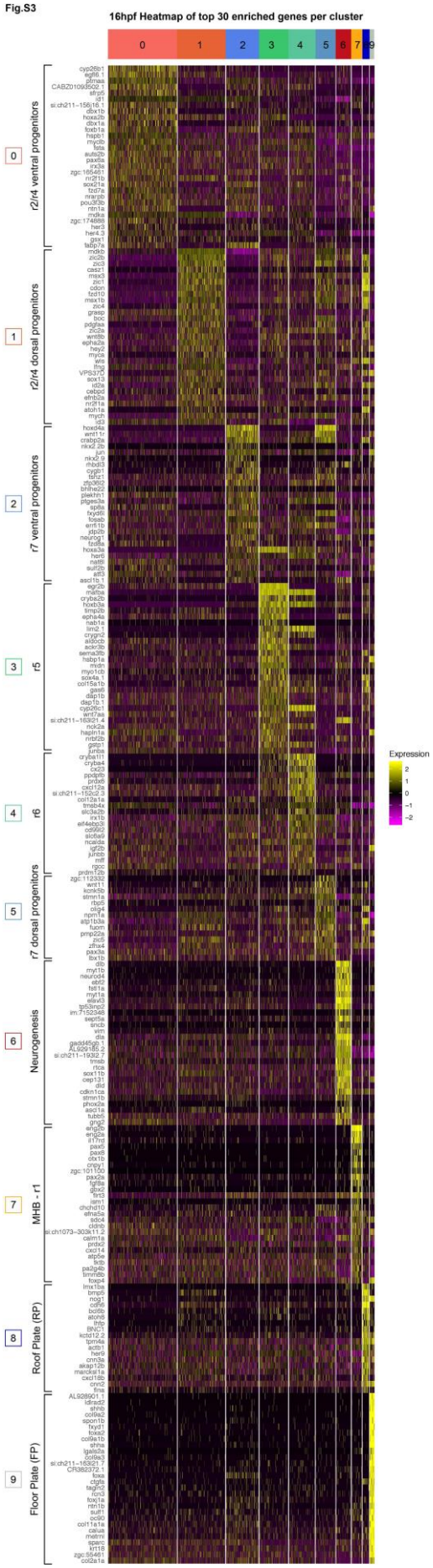


Figure S3. Heatmap of the top 30 significant markers per cluster at 16 hpf.

Full heatmap of the top 30 significant markers per cluster, if available, at 16 hpf.

Fig.S4

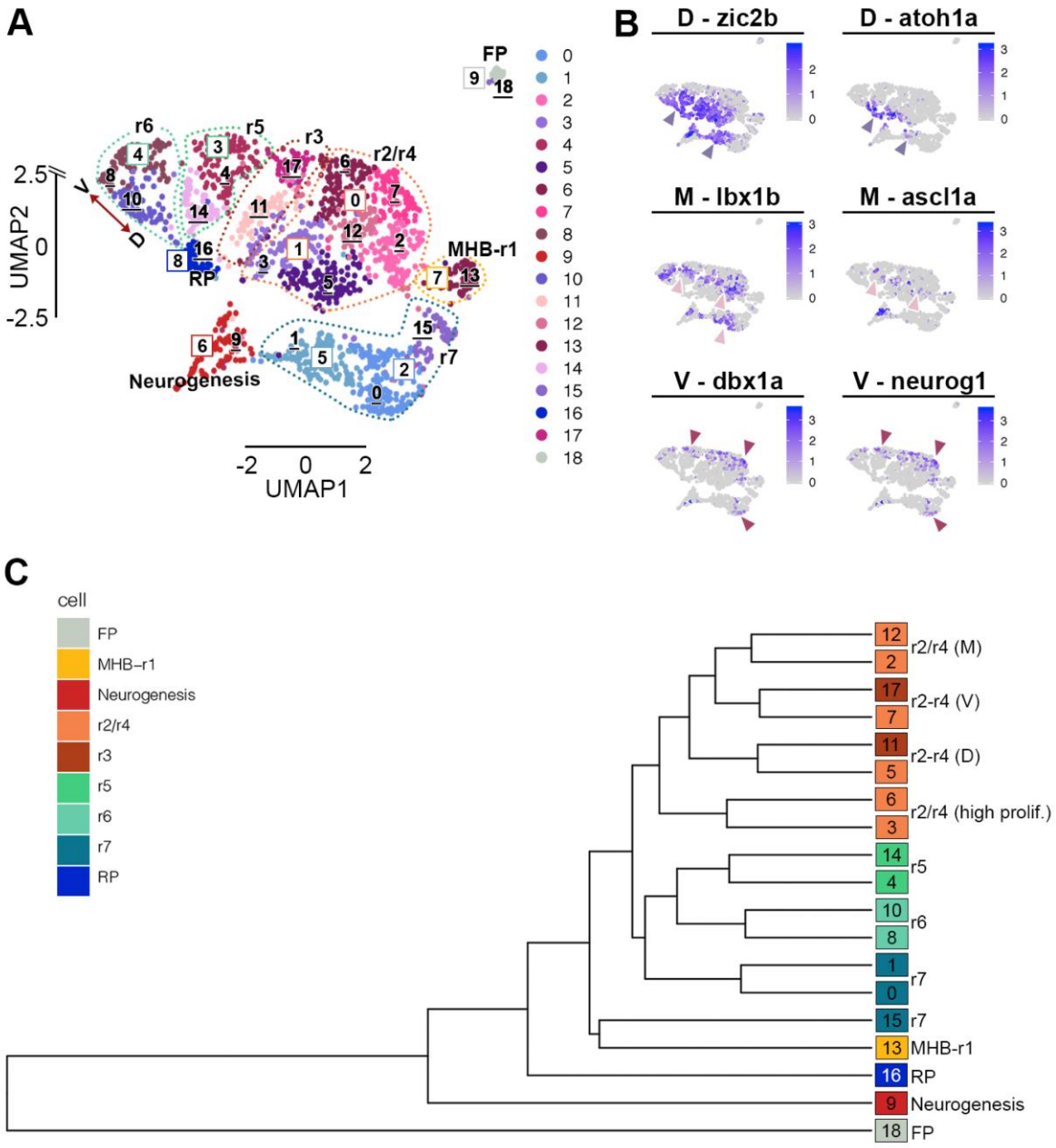


Figure S4. Subclustering of 16 hpf hindbrain.

(A) Higher resolution analysis of the 16 hpf hindbrain cells identifies 19 clusters. Dorsal, medial and ventral progenitors are separated in distinct clusters along the anterior-posterior axis. Clusters identified in Fig. 2A are overlaid to visualize rhombomere identity. r3 is now separated from r2/r4, and multiple clusters appear in the r2/r4 domain. UMAP2 (y-axis) is discontinuous. (B) Dorsal (*zic2b, atoh1a*), medial (*lbx1b, ascl1a*) and ventral (*dbx1a, neurog1*) gene expression domains are reported. Colour intensity is proportional to the expression level of a given gene. (C) Analysis with PlotClusterTree in Seurat to reveal

the transcriptomic similarities between clusters. Cluster identity in (A) and (C) are colour coded as indicated.

Fig.S5

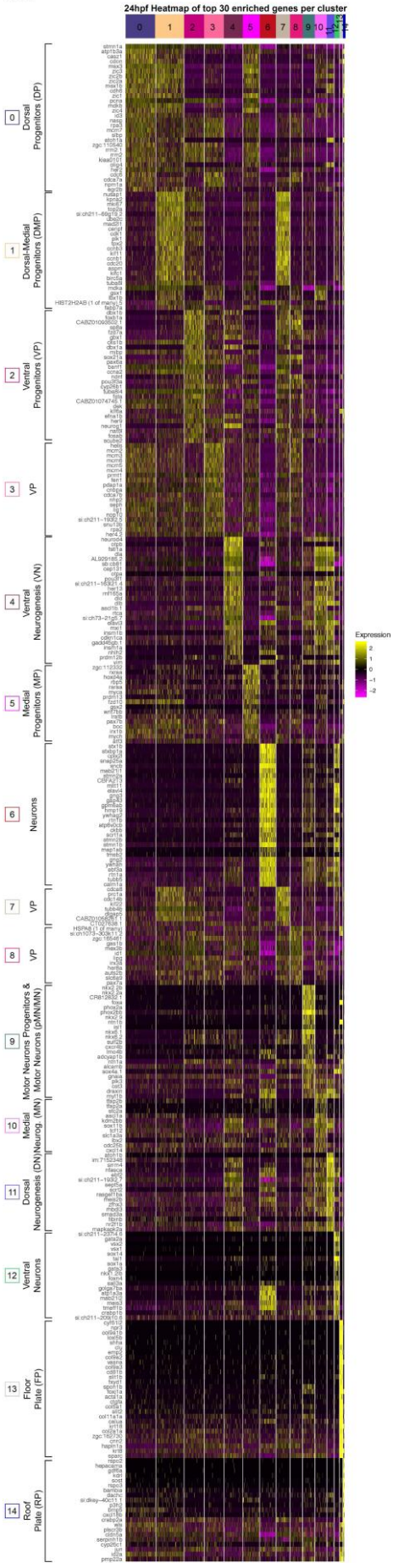


Figure S5. Heatmap of the top 30 significant markers per cluster at 24 hpf.

Full heatmap of the top 30 significant markers per cluster, if available, at 24 hpf.

Fig.S6

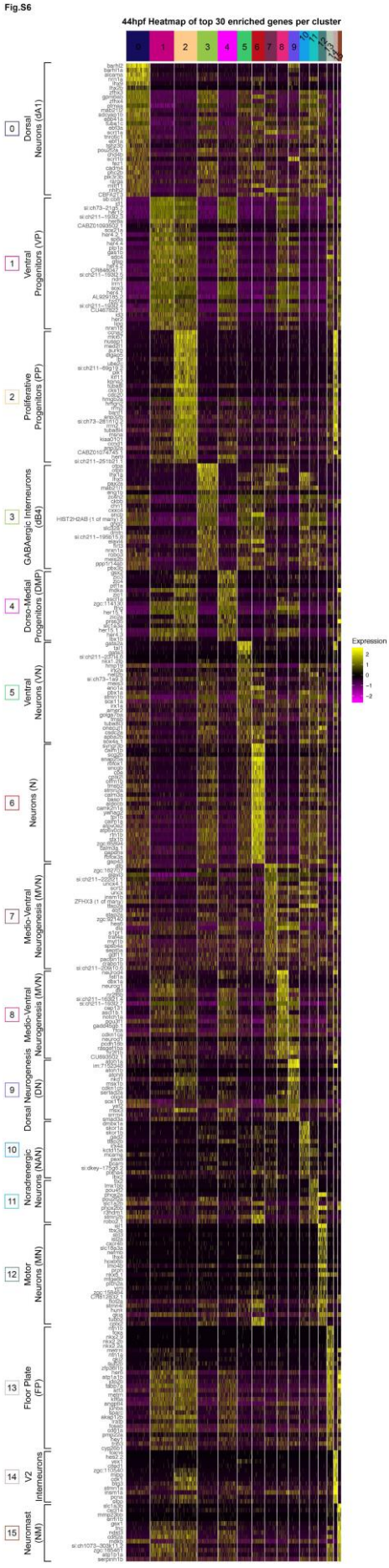


Figure S6. Heatmap of the top 30 significant markers per cluster at 44 hpf.

Full heatmap of the top 30 significant markers per cluster, if available, at 44 hpf.

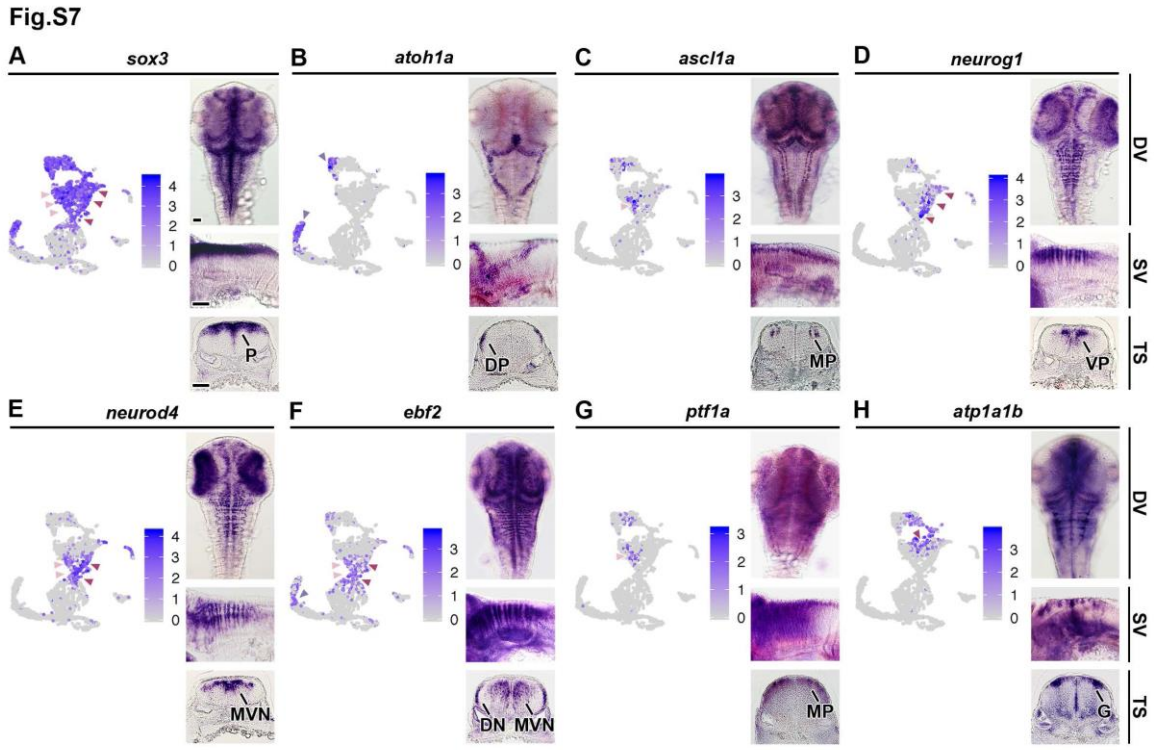


Figure S7. Selected expression patterns of progenitors and differentiation factors at 44 hpf.

Progenitor marker *sox3* (A) is widely expressed in the ventricular zone. Proneural genes *atoh1a* (B), ascl1a (C) and *neurog1* (D) are differentially expressed along the D-V axis at 44 hpf, similarly to their distribution at 24 hpf. *neurod4* (E) is found in medio-ventral differentiating progenitors, *ebf2* (F) has an expression domain resembling *neurod4*, and also expressed in some differentiated neurons, while *ptf1a* (G) is found medially in differentiating cells. *atp1a1b* is a newly identified marker of glial cells (H). For each gene the UMAP plot shows gene expression from the 44 hpf scRNA-seq data; colour intensity is proportional to the expression level of a given gene. In situ hybridization images are shown for dorsal view (DV), side view (SV) and transverse section (TS) at the level of r4-r5/r5-r6. scRNA-seq and in situ hybridization expression patterns strongly correlate. P = Progenitors, DP = Dorsal Progenitors, MP = Medial Progenitors, VP = Ventral Progenitors, MVN = Medio-Ventral Neurogenesis, DN = Dorsal Neurogenesis, G = Glia.

Fig.S8

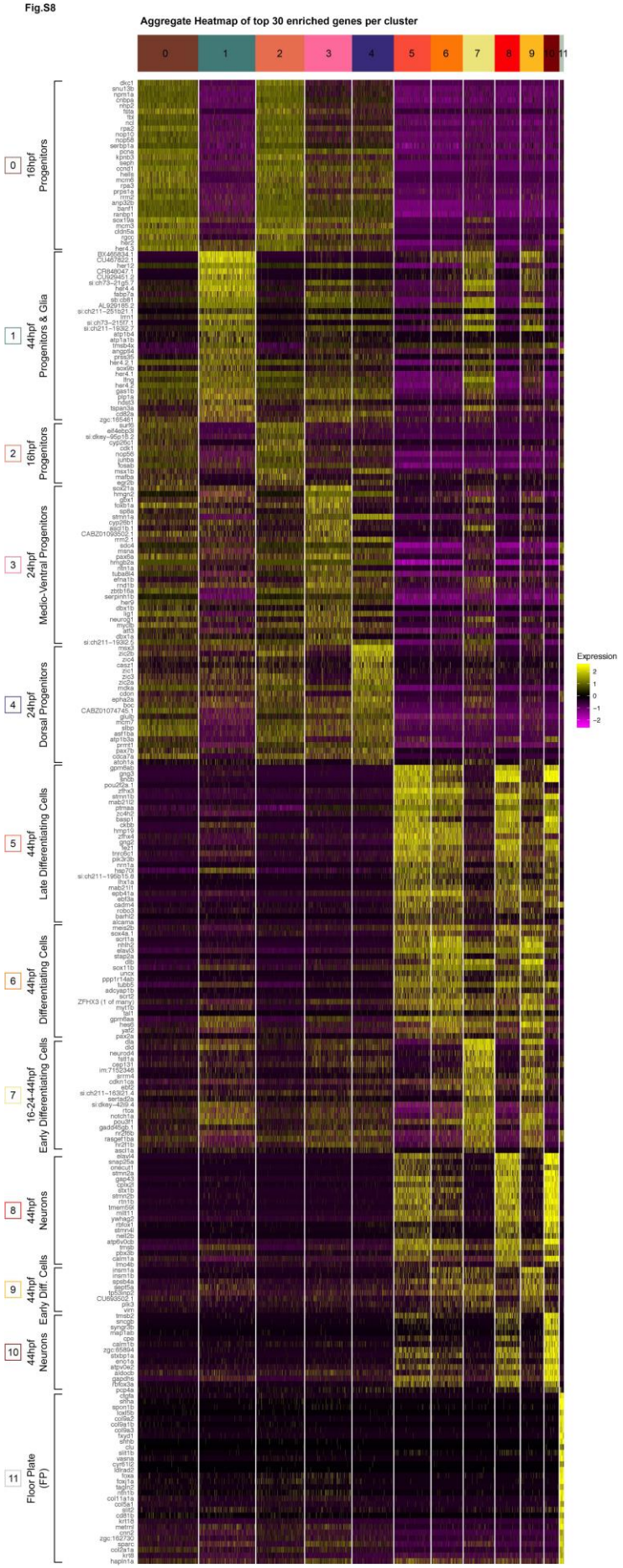


Figure S8. Heatmap of the top 30 significant markers per cluster for the aggregate data set.

Full heatmap of the top 30 significant markers per cluster, if available, for the aggregate data set.

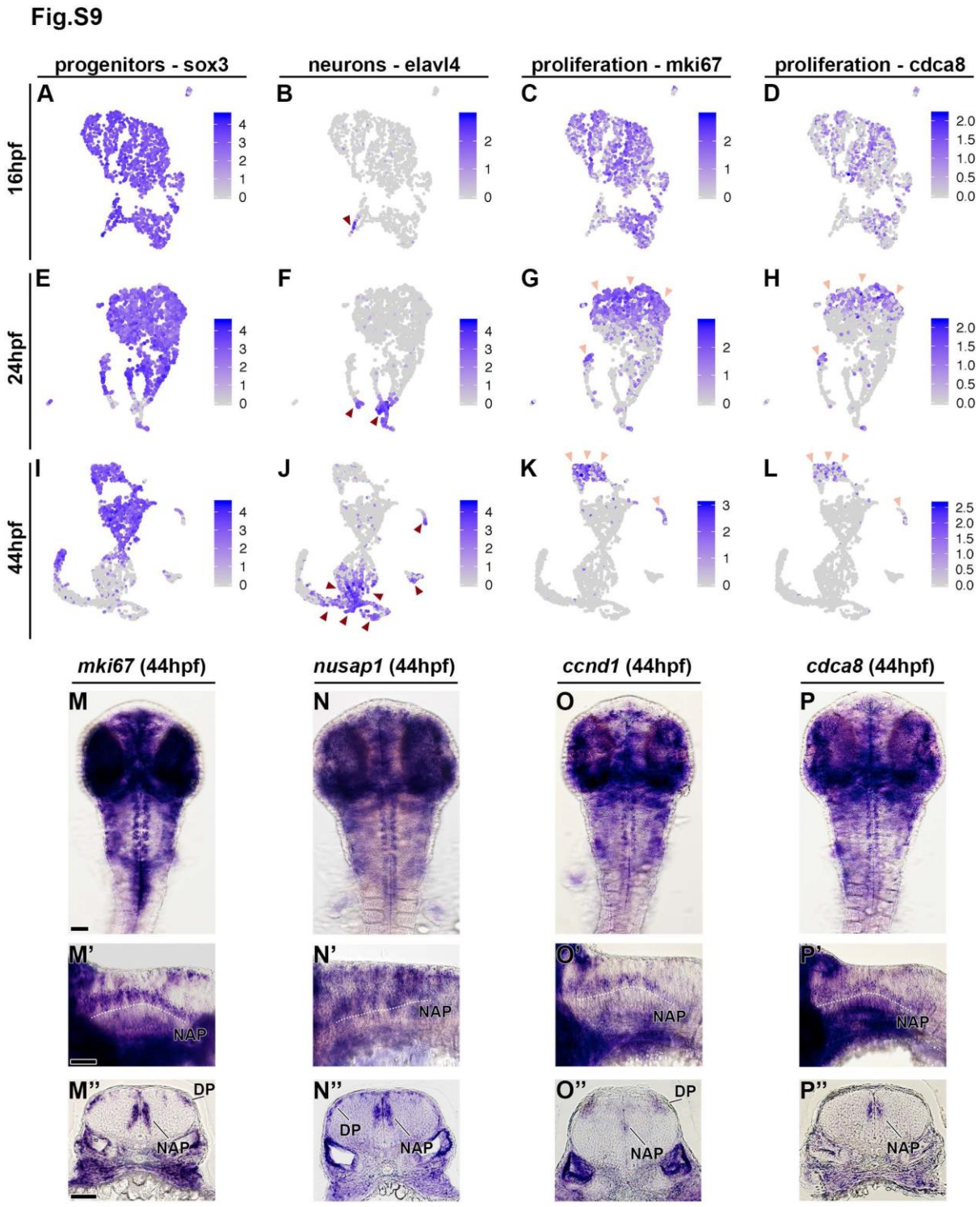


Figure S9. Progenitor, neurogenesis and proliferation gene expression at different stages.

For each gene, the UMAP plot shows gene expression from 16 hpf (A-D), 24 hpf (E-H) and 44 hpf (I-L) scRNA-seq data. Progenitor cells are marked by *sox3* expression (A, E, I), neurogenesis by *elavl4* (B, F, J) and proliferation by *mki67* (C, G, K) and *cdca8* (D, H, L). Whole mount in situ hybridization at 44 hpf of *mki67* (M), *nusap1* (N), *ccnd1* (O) and *cdca8*

(P). Dorsal view (M-P), side view (M'-P') and 40 µm hindbrain transverse section at the level of r4-r5/r5-r6 (M''-P''). Scale bar: 50 µm.

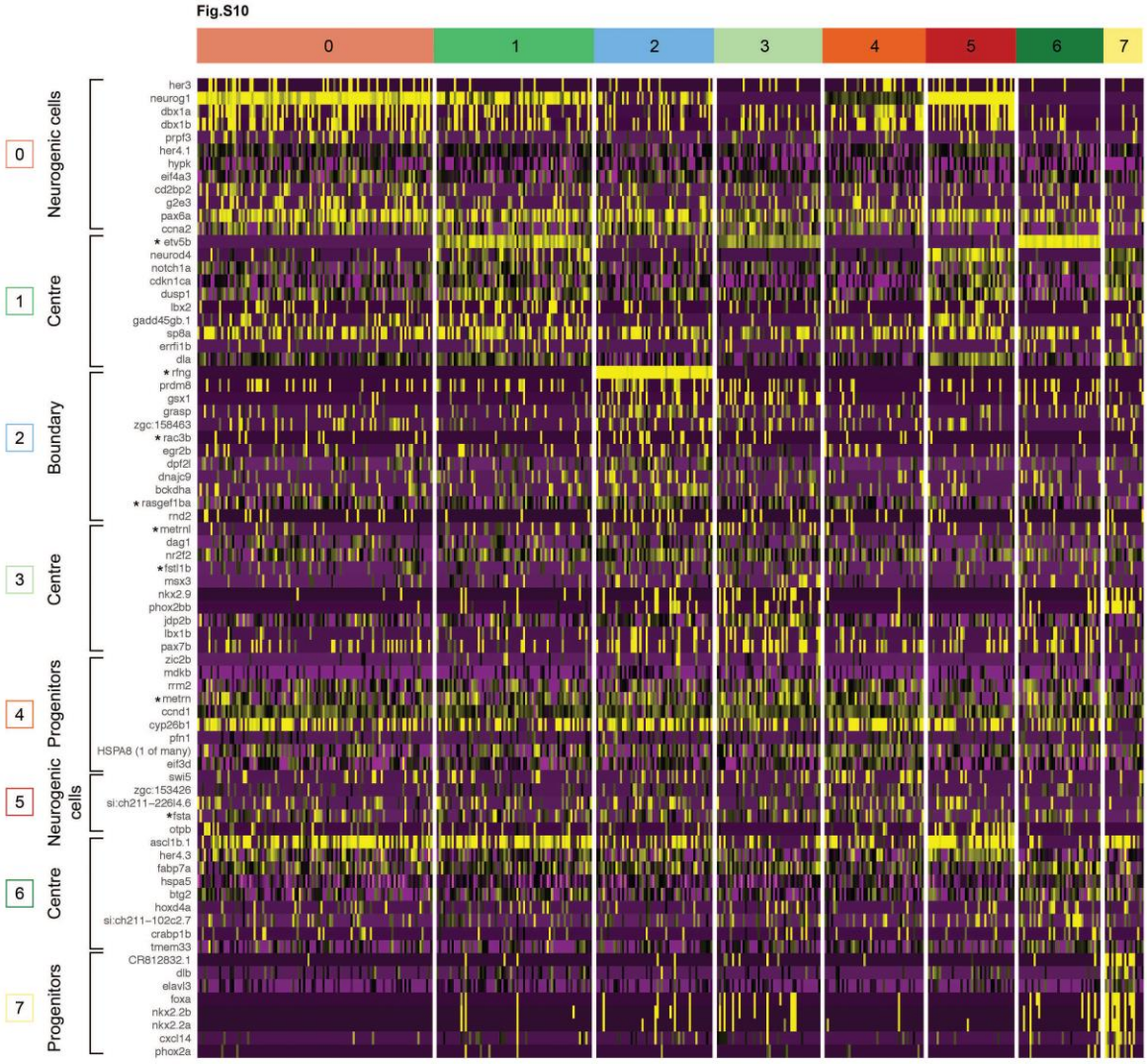


Figure S10. Selected top markers for 24 hpf boundary and centre supervised analysis. Heatmap of selected markers (pval<0.1, logfc>0.1, detected at a minimum fraction of 20% of tested cells) of the supervised clustering analysis done on 24 hpf ventral progenitors (VP). Validated and known makers of boundary and rhombomere centre cells are highlighted with an asterisk.

Fig.S11

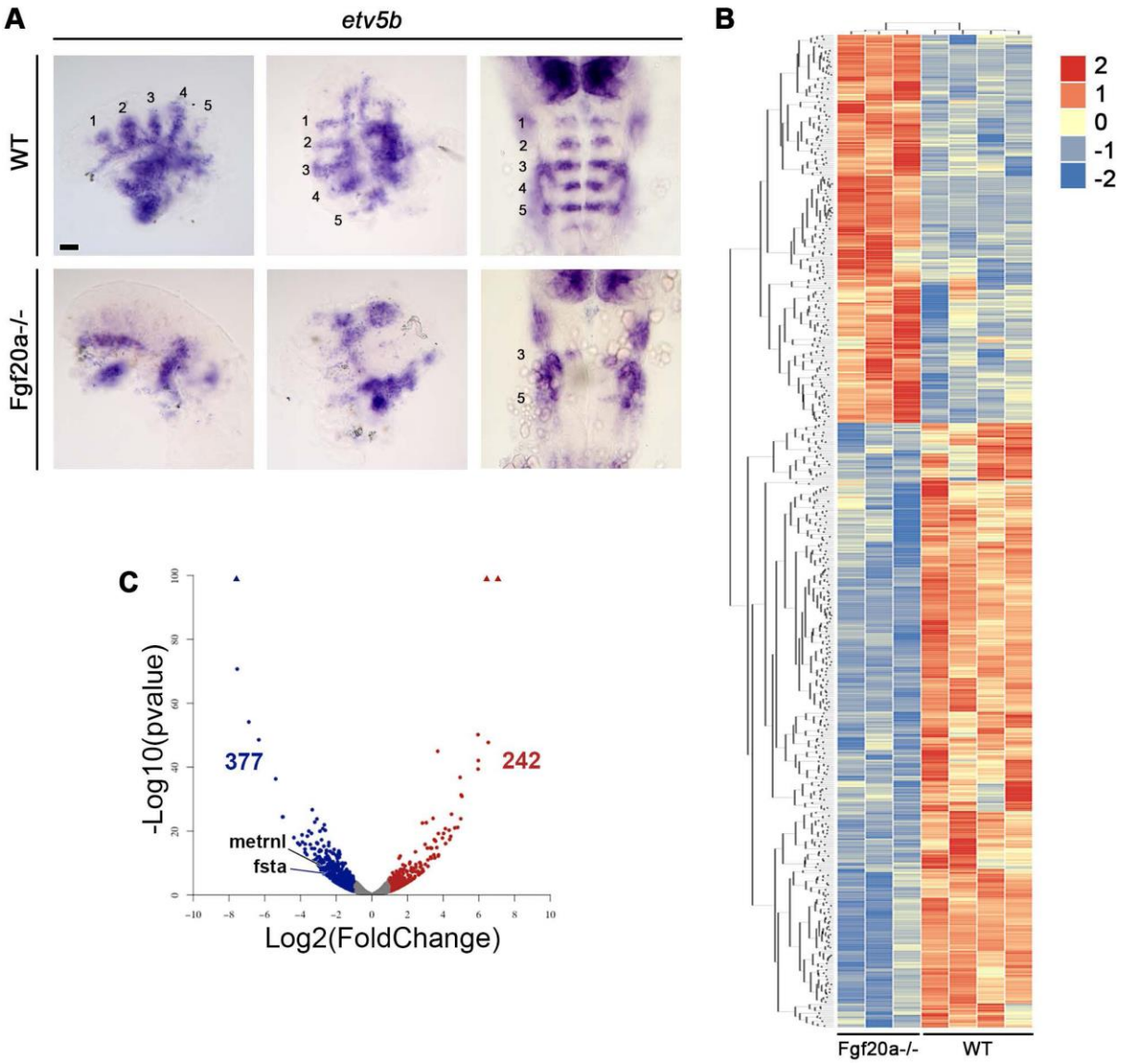


Figure S11. Fgf20a-/- bulk RNA-seq identifies metrnl and fsta as new Fgf20 targets in the hindbrain.

(A) Examples of *etv5b* expression in dissected hindbrain for wild-type (WT) and *fgf20a-/-* 24 hpf embryos. Five stripes of segment centre expression occur in WT embryos, together with otic vesicle and cranial ganglia expression domains. In *fgf20a-/-* embryos only weak r3 and r5 stripes are present, while otic vesicle and cranial ganglia expression domains are unaffected. Representative whole embryos are also shown. Strong expression in domains outside the hindbrain probably masks changes in the hindbrain (e.g. *etv5b*). Scale bar: 50 µm. (B) Heatmap showing RNA-seq expression levels of significantly differentially expressed genes between 4 WT and 3 *fgf20a-/-* dissected tissues. Hierarchical clustering groups the WT tissues and the mutants in separate clusters, suggesting genome wide

similarities in dissected samples of the same genotype. Colour scale depicts low to high expression in blue to red shades, respectively. (C) Volcano plot shows 377 significantly downregulated genes in blue and 242 upregulated in red. *metrnl* and *fsta* are among the downregulated factors. Grey dots are non-significant genes, x-axis Log2(Fold Change) and y-axis -Log10(pvalue).

Fig.S12

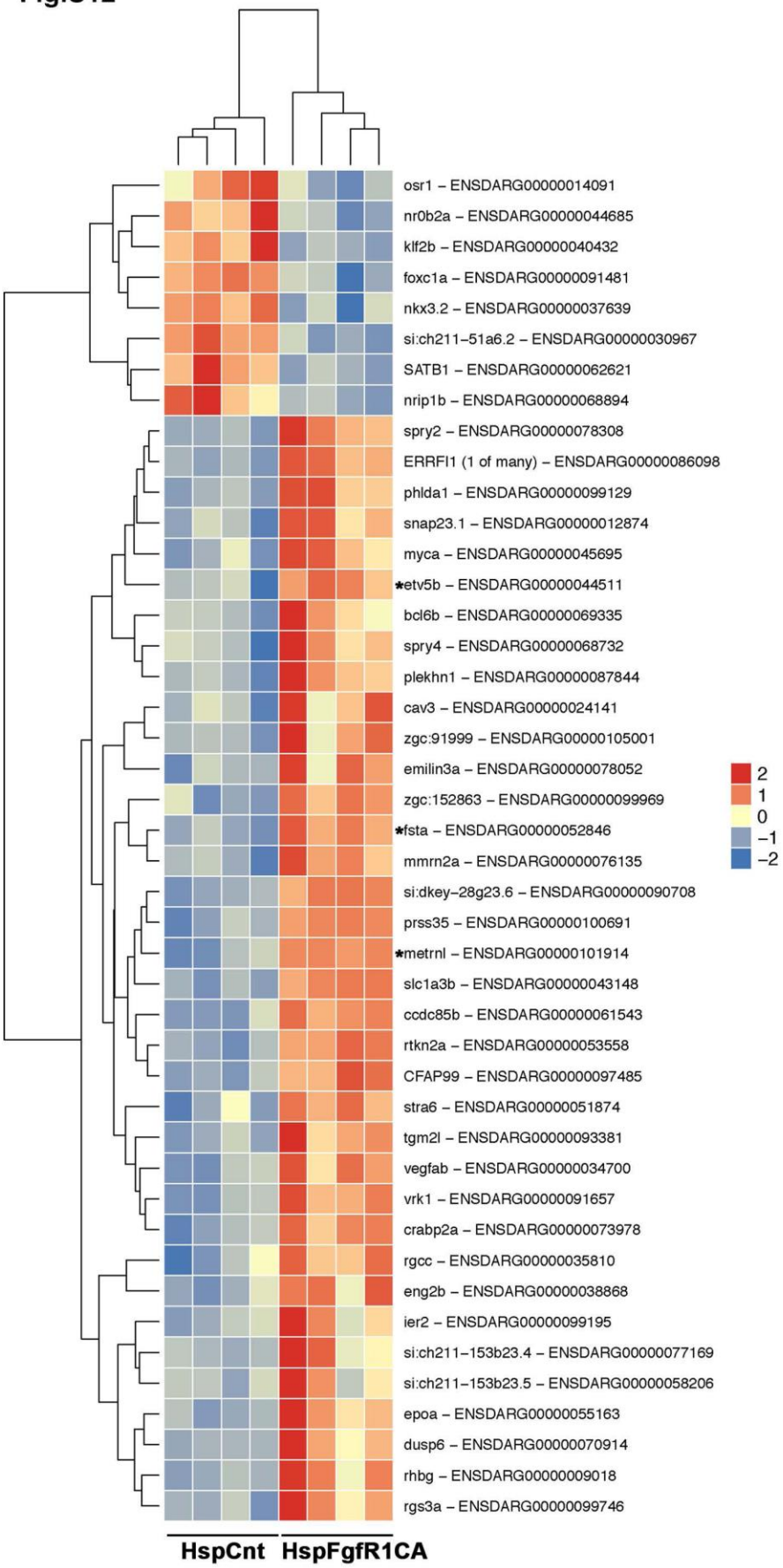


Figure S12. Constitutive activation of FgfR1 ectopically induces *etv5b*, *metrnl* and *fsta* expression.

Heatmap shows RNA-seq expression levels of significantly differentially expressed genes between 4 heat shocked controls (HspCnt) and 4 heat shocked constitutive active FgfR1 (HspFgfR1CA) dissected tissues. Hierarchical clustering groups the 4 HspCnt tissues and the 4 HspFgfR1CA in separate clusters, suggesting genome wide similarities in dissected samples of the same genotype. Colours scale depicts low to high expression in blue to red shades, respectively. 8 genes are significantly downregulated in HspFgfR1CA, while 36 are upregulated. Among the upregulated genes, known Fgf signaling targets are found (e.g. *spry2*, *spry4* and *etv5b*) and in addition *metrnl* and *fsta* are found that are expressed in hindbrain segment centres.

Fig.S13

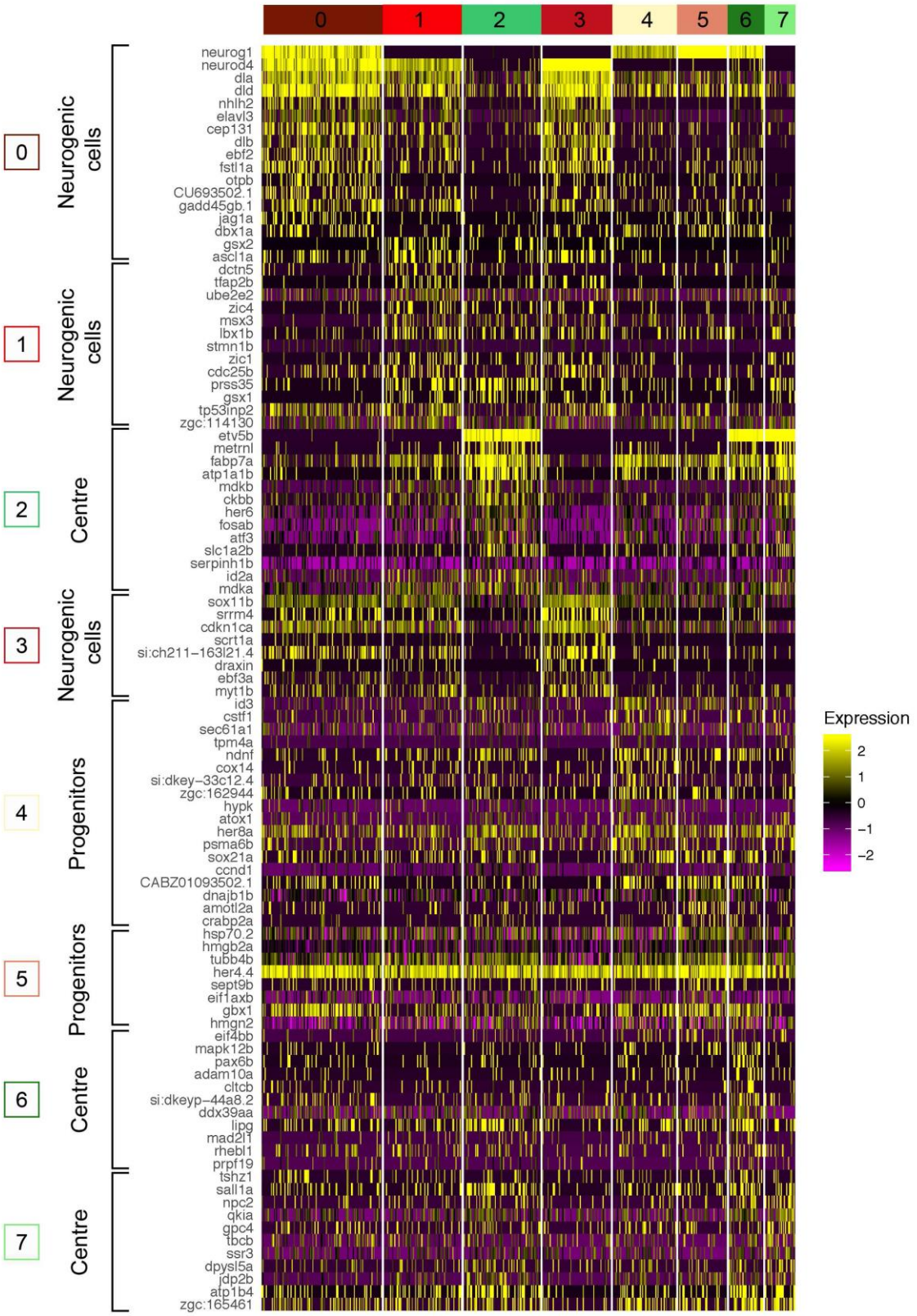


Figure S13. Heatmap of the top 15 significant markers per cluster for 44 hpf centre progenitors supervised analysis.

Full heatmap of the top 15 significant markers per cluster, if available, for 44 hpf centre progenitors supervised analysis.

Fig.S14

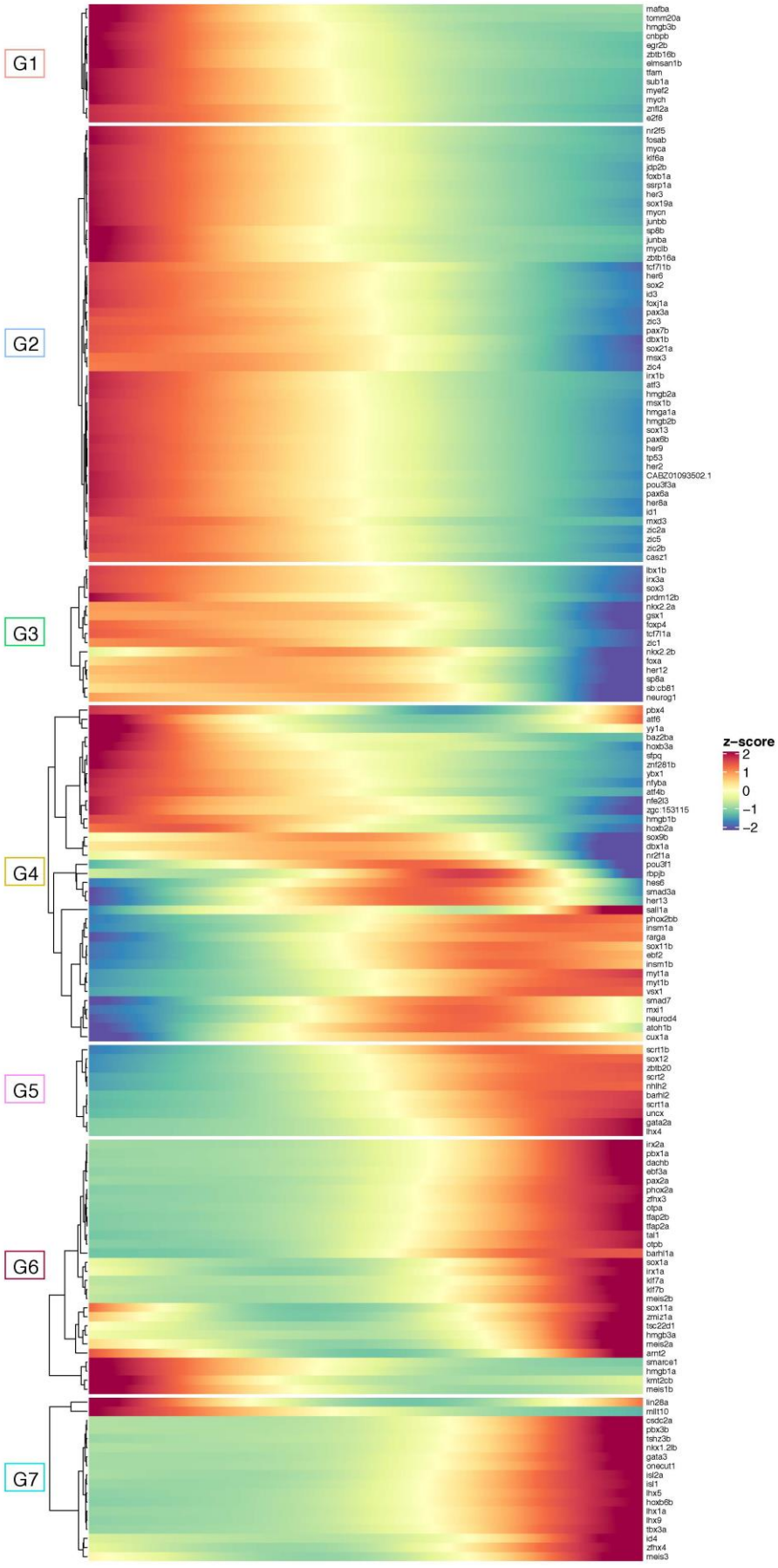
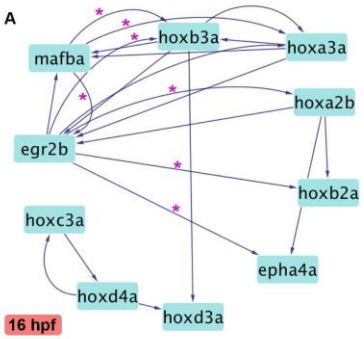


Figure S14. Heatmap of selected transcription factors changing with pseudotime.
Full heatmap of selected transcription factors changing along the pseudo-temporal axis.

Fig.S15

A



B

Regulatory Gene	Target Gene	Weight	Refs
mafba	egr2b	0.0617	Chomette et al., 2006; Labalette et al., 2011
egr2b	hoxa2b	0.0632	Nonchev et al., 1996a; Nonchev et al., 1996b
egr2b	hoxb2a	0.0469	Sham et al., 1993; Nonchev et al., 1996a
mafba	hoxa3a	0.0983	Manzanares et al., 1999; Manzanares et al., 1997
hoxb3a	hoxa3a	0.0875	Manzanares et al., 2001
mafba	hoxb3a	0.1841	Manzanares et al., 1999; Manzanares et al., 1997; Manzanares et al., 2002
egr2b	hoxb3a	0.0380	Manzanares et al., 2002
egr2b	epha4a	0.1782	Theil et al., 1998

Figure S15. Known interactions retrieved with Genie3 at 16 hpf.

(A) Interaction among *egr2b* (*krox20*), *hox* genes (*hoxa2b, hoxb2a, hoxa3a, hoxb3a, hoxc3a, hoxd3a, hoxd4a*) and *mafba* from the predicted gene regulatory network. * marks known interaction that have been validated in vivo. (B) Table summarizing the retrieved known interactions with relative weight predicted by Genie3 interaction and references of the correspondent in vivo validations.

SUPPLEMENTARY TABLES

Table S1. Hindbrain cells.

Spreadsheet containing the names of cells considered to be hindbrain on the basis of marker gene expression.

Click here to Download Table S1

Table S2. Expression pattern summary.

Spreadsheet 1 - Table S2.1. Expression pattern summary of selected genes differentially expressed at 16 hpf.

Spreadsheet 2 - Table S2.2. Expression pattern summary of selected genes differentially expressed at 24 hpf.

Spreadsheet 3 - Table S2.3. Expression pattern summary of selected genes differentially expressed at 44 hpf.

Spreadsheet 4 - Table S2.4. Expression pattern summary of selected genes differentially expressed in the Aggregate data set.

Spreadsheet 5 - Table S2.5. Expression pattern summary of differentially expressed genes between boundary and centre progenitors at 24 hpf.

Spreadsheet 6 - Table S2.6. Expression pattern summary of enriched genes in centre progenitors at 44 hpf.

References are listed below.

Click here to Download Table S2

Table S3. Differential expression of significant genes between wild-type and Fgf20a-/-.

Bulk RNA-seq analysis of 4 wild-type (WT) and 3 Fgf20a-/- dissected hindbrain tissues.

Click here to Download Table S3

Table 4. Differential expression of significant genes between heat shock controls and HspFgfR1CA.

Bulk RNA-seq analysis of 4 heat shock control (HspCnt) and 4 heat shock constitutive active FgfR1 (HspCAFgfR1) dissected hindbrain tissues.

Click here to Download Table S4

Table S5. GENIE3 predicted interactions (IM>0.025)

Genie3 table of interactions presenting regulatoryGene, targetGene and weight of the interaction (IM>=0.025).

Spreadsheet 1 - Table S5.1. Predicted interaction at 16hpf

Spreadsheet 2 - Table S5.2. Predicted interaction at 24hpf

Spreadsheet 3 - Table S5.3. Predicted interaction at 44hpf

Click here to Download Table S5

SUPPLEMENTARY REFERENCES

Allende, M. L., et al. (1994). The expression pattern of two zebrafish achaete-scute homolog (ash) genes is altered in the embryonic brain of the cyclops mutant. *Dev. Biol.* **166**, 509-30.

Amoyel, M., et al. (2005). Wnt1 regulates neurogenesis and mediates lateral inhibition of boundary cell specification in the zebrafish hindbrain. *Development* **132**, 775-85.

Ando, H., et al. (2005). Lhx2 mediates the activity of Six3 in zebrafish forebrain growth. *Dev. Biol.* **287**, 456-68.

Appel, B., et al. (1995). Motoneuron fate specification revealed by patterned LIM homeobox gene expression in embryonic zebrafish. *Development* **121**, 4117-25.

Bae, Y. K., et al. (2005). Patterning of proneuronal and inter-proneuronal domains by hairy- and enhancer of split-related genes in zebrafish neuroectoderm. *Development* **132**, 1375-85.

Beretta, C. A., et al. (2011). All four zebrafish Wnt7 genes are expressed during early brain development. *Gene Expr Patterns* **11**, 277-84.

Cadwallader, A. B., et al. (2006). Combinatorial expression patterns of heparan sulfate sulfotransferases in zebrafish: I. The 3-O-sulfotransferase family. *Dev. Dyn.* **235**, 3423-31.

Caron, A., et al. (2012). Wnt/beta-catenin signaling directly regulates Foxj1 expression and ciliogenesis in zebrafish Kupffer's vesicle. *Development* **139**, 514-24.

Cheng, C. W., et al. (2007). Zebrafish homologue irx1a is required for the differentiation of serotonergic neurons. *Dev. Dyn.* **236**, 2661-7.

Cheng, Y. C., et al. (2004). Notch activation regulates the segregation and differentiation of rhombomere boundary cells in the zebrafish hindbrain. *Dev. Cell* **6**, 539-50.

Choe, S. K., et al. (2011). A screen for hoxb1-regulated genes identifies ppp1r14al as a regulator of the rhombomere 4 Fgf-signaling center. *Dev. Biol.* **358**, 356-67.

Chomette, D., et al. (2006). Krox20 hindbrain cis-regulatory landscape: interplay between multiple long-range initiation and autoregulatory elements. *Development* **133**, 1253-62.

Colombo, A., et al. (2006). Zebrafish BarH-like genes define discrete neural domains in the early embryo. *Gene Expr Patterns* **6**, 347-52.

Dee, C. T., et al. (2008). Sox3 regulates both neural fate and differentiation in the zebrafish ectoderm. *Dev. Biol.* **320**, 289-301.

Del Giacco, L., et al. (2006). Differential regulation of the zebrafish orthopedia 1 gene during fate determination of diencephalic neurons. *BMC Dev. Biol.* **6**, 50.

Duncan, R. N., et al. (2015). Identification of Wnt Genes Expressed in Neural Progenitor Zones during Zebrafish Brain Development. *PLoS One* **10**, e0145810.

Elsen, G. E., et al. (2008). Zic1 and Zic4 regulate zebrafish roof plate specification and hindbrain ventricle morphogenesis. *Dev. Biol.* **314**, 376-92.

Elsen, G. E., et al. (2009). The autism susceptibility gene met regulates zebrafish cerebellar development and facial motor neuron migration. *Dev. Biol.* **335**, 78-92.

Emoto, Y., et al. (2005). Retinoic acid-metabolizing enzyme Cyp26a1 is essential for determining territories of hindbrain and spinal cord in zebrafish. *Dev. Biol.* **278**, 415-27.

Esain, V., et al. (2010). FGF-receptor signalling controls neural cell diversity in the zebrafish hindbrain by regulating olig2 and sox9. *Development* **137**, 33-42.

Filipek-Gorniok, B., et al. (2015). The NDST gene family in zebrafish: role of NDST1B in pharyngeal arch formation. *PLoS One* **10**, e0119040.

Filippi, A., et al. (2005). The basic helix-loop-helix olig3 establishes the neural plate boundary of the trunk and is necessary for development of the dorsal spinal cord. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 4377-82.
Fjose, A., et al. (1994). Expression of the zebrafish gene hlx-1 in the prechordal plate and during CNS development. *Development* **120**, 71-81.
Ghosh, P., et al. (2018). Analysis of novel caudal hindbrain genes reveals different regulatory logic for gene expression in rhombomere 4 versus 5/6 in embryonic zebrafish. *Neural Dev* **13**, 13.
Gorsi, B., et al. (2010). Dynamic expression patterns of 6-O endosulfatases during zebrafish development suggest a subfunctionalisation event for sulf2. *Dev. Dyn.* **239**, 3312-23.
Grinblat, Y., et al. (2001). zic Gene expression marks anteroposterior pattern in the presumptive neurectoderm of the zebrafish gastrula. *Dev. Dyn.* **222**, 688-93.
Gu, X., et al. (2005). Molecular cloning and expression of a novel CYP26 gene (cyp26d1) during zebrafish early development. *Gene Expr Patterns* **5**, 733-9.
Guner, B., et al. (2007). Cloning of zebrafish nkx6.2 and a comprehensive analysis of the conserved transcriptional response to Hedgehog/Gli signaling in the zebrafish neural tube. *Gene Expr Patterns* **7**, 596-605.
Gupta, M., et al. (2013). Identification and expression analysis of zebrafish glypicans during embryonic development. *PLoS One* **8**, e80824.
Haddon, C., et al. (1998). Multiple delta genes and lateral inhibition in zebrafish primary neurogenesis. *Development* **125**, 359-70.
Hans, S., et al. (2004). her3, a zebrafish member of the hairy-E(spl) family, is repressed by Notch signalling. *Development* **131**, 2957-69.
Hirate, Y., et al. (2006). Canopy1, a novel regulator of FGF signaling around the midbrain-hindbrain boundary in zebrafish. *Curr. Biol.* **16**, 421-7.
Hong, E., et al. (2013). Cholinergic left-right asymmetry in the habenulo-interpeduncular pathway. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 21171-6.
Hsu, L. S., et al. (2010). Zebrafish calcium/calmodulin-dependent protein kinase II (cam-kii) inhibitors: expression patterns and their roles in zebrafish brain development. *Dev. Dyn.* **239**, 3098-105.
Kani, S., et al. (2010). Proneural gene-linked neurogenesis in zebrafish cerebellum. *Dev. Biol.* **343**, 1-17.
Kelly, G. M., et al. (1995). Zebrafish wnt8 and wnt8b share a common activity but are involved in distinct developmental pathways. *Development* **121**, 1787-99.
Kinkhabwala, A., et al. (2011). A structural and functional ground plan for neurons in the hindbrain of zebrafish. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 1164-9.
Knight, R. D., et al. (2005). AP2-dependent signals from the ectoderm regulate craniofacial development in the zebrafish embryo. *Development* **132**, 3127-38.
Kondrychyn, I., et al. (2017). Transcriptional Complexity and Distinct Expression Patterns of auts2 Paralogs in Danio rerio. *G3 (Bethesda)* **7**, 2577-2593.
Korzh, V., et al. (1998). Expression of zebrafish bHLH genes ngn1 and nrd defines distinct stages of neural differentiation. *Dev. Dyn.* **213**, 92-104.
Kotkamp, K., et al. (2014). Pou5f1/Oct4 promotes cell survival via direct activation of mych expression during zebrafish gastrulation. *PLoS One* **9**, e92356.
Koudijs, M. J., et al. (2005). The zebrafish mutants dre, uki, and lep encode negative regulators of the hedgehog signaling pathway. *PLoS Genet* **1**, e19.
Kudoh, T., et al. (2001). A gene expression screen in zebrafish embryogenesis. *Genome Res.* **11**, 1979-87.
Labalette, C., et al. (2011). Hindbrain patterning requires fine-tuning of early krox20 transcription by Sprouty 4. *Development* **138**, 317-26.
Lee, H. C., et al. (2017). Embryonic expression patterns of Eukaryotic EndoU ribonuclease family gene endouC in zebrafish. *Gene Expr Patterns* **25-26**, 66-70.
Lee, S. A., et al. (2003). The zebrafish forkhead transcription factor Foxi1 specifies epibranchial placode-derived sensory neurons. *Development* **130**, 2669-79.

Li, J., et al. (2014). Temporal and spatial expression of the four Igf ligands and two Igf type 1 receptors in zebrafish during early embryonic development. *Gene Expr Patterns* **15**, 104-11.
Li, Y., et al. (2018). Temporal and spatial expression of fgfbp genes in zebrafish. *Gene* **659**, 128-136.
Liu, Q., et al. (2006). cadherin-6 message expression in the nervous system of developing zebrafish. *Dev. Dyn.* **235**, 272-8.
Love, C. E., et al. (2012). Expression and retinoic acid regulation of the zebrafish nr2f orphan nuclear receptor genes. *Dev. Dyn.* **241**, 1603-15.
Lun, K., et al. (1998). A series of no isthmus (noi) alleles of the zebrafish pax2.1 gene reveals multiple signaling events in development of the midbrain-hindbrain boundary. *Development* **125**, 3049-62.
Luo, N., et al. (2016). Syndecan-4 modulates the proliferation of neural cells and the formation of CaP axons during zebrafish embryonic neurogenesis. *Sci. Rep.* **6**, 25300.
Manzanares, M., et al. (2001). Independent regulation of initiation and maintenance phases of Hoxa3 expression in the vertebrate hindbrain involve auto- and cross-regulatory mechanisms. *Development* **128**, 3595-607.
Manzanares, M., et al. (1999). Conserved and distinct roles of kreisler in regulation of the paralogous Hoxa3 and Hoxb3 genes. *Development* **126**, 759-69.
Manzanares, M., et al. (1997). Segmental regulation of Hoxb-3 by kreisler. *Nature* **387**, 191-5.
Manzanares, M., et al. (2002). Krox20 and kreisler co-operate in the transcriptional control of segmental expression of Hoxb3 in the developing hindbrain. *EMBO J.* **21**, 365-76.
Mckeown, K. A., et al. (2012). Disruption of Eaat2b, a glutamate transporter, results in abnormal motor behaviors in developing zebrafish. *Dev. Biol.* **362**, 162-71.
Melvin, V. S., et al. (2013). A morpholino-based screen to identify novel genes involved in craniofacial morphogenesis. *Dev. Dyn.* **242**, 817-31.
Minchin, J. E., et al. (2008). Sequential actions of Pax3 and Pax7 drive xanthophore development in zebrafish neural crest. *Dev. Biol.* **317**, 508-22.
Miyake, A., et al. (2012). Neucrin, a novel secreted antagonist of canonical Wnt signaling, plays roles in developing neural tissues in zebrafish. *Mech. Dev.* **128**, 577-90.
Moens, C. B., et al. (1998). Equivalence in the genetic control of hindbrain segmentation in fish and mouse. *Development* **125**, 381-91.
Moens, C. B., et al. (1996). valentino: a zebrafish gene required for normal hindbrain segmentation. *Development* **122**, 3981-90.
Nakaya, N., et al. (2007). Expression patterns of alternative transcripts of the zebrafish olfactomedin 1 genes. *Gene Expr Patterns* **7**, 723-9.
Nechiporuk, A., et al. (2007). Specification of epibranchial placodes in zebrafish. *Development* **134**, 611-23.
Nepal, C., et al. (2016). Transcriptional, post-transcriptional and chromatin-associated regulation of pri-miRNAs, pre-miRNAs and moRNAs. *Nucleic Acids Res.* **44**, 3070-81.
Nikaido, M., et al. (2013). A systematic survey of expression and function of zebrafish frizzled genes. *PLoS One* **8**, e54833.
Nikolaou, N., et al. (2009). Lunatic fringe promotes the lateral inhibition of neurogenesis. *Development* **136**, 2523-33.
Nonchev, S., et al. (1996a). The conserved role of Krox-20 in directing Hox gene expression during vertebrate hindbrain segmentation. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 9339-45.
Nonchev, S., et al. (1996b). Segmental expression of Hoxa-2 in the hindbrain is directly regulated by Krox-20. *Development* **122**, 543-54.
Ochi, H., et al. (2009). Lbx2 regulates formation of myofibrils. *BMC Dev. Biol.* **9**, 13.
Oxtoby, E., et al. (1993). Cloning of the zebrafish krox-20 gene (krx-20) and its expression during hindbrain development. *Nucleic Acids Res.* **21**, 1087-95.

Pauls, S., et al. (2012). Lens development depends on a pair of highly conserved Sox21 regulatory elements. *Dev. Biol.* **365**, 310-8.
Piotrowski, T., et al. (2000). The endoderm plays an important role in patterning the segmented pharyngeal region in zebrafish (Danio rerio). *Dev. Biol.* **225**, 339-56.
Prince, V. E., et al. (1998). Zebrafish hox genes: expression in the hindbrain region of wild-type and mutants of the segmentation gene, valentino. *Development* **125**, 393-406.
Radomska, K. J., et al. (2016). Characterization and Expression of the Zebrafish qki Paralogs. *PLoS One* **11**, e0146155.
Rauch, G. J., Lyons, D.A., Middendorf, I., Friedlander, B., Arana, N., Reyes, T., and Talbot, W.S. (2003). Submission and Curation of Gene Expression Data. ZFIN Direct Data Submission (http://zfin.org).
Riley, B. B., et al. (2004). Rhombomere boundaries are Wnt signaling centers that regulate metameric patterning in the zebrafish hindbrain. *Dev. Dyn.* **231**, 278-91.
Rohrschneider, M. R., et al. (2007). Zebrafish Hoxb1a regulates multiple downstream genes including prickle1b. *Dev. Biol.* **309**, 358-72.
Selland, L. G., et al. (2018). Coordinate regulation of retinoic acid synthesis by pbx genes and fibroblast growth factor signaling by hoxb1b is required for hindbrain patterning and development. *Mech. Dev.* **150**, 28-41.
Sham, M. H., et al. (1993). The zinc finger gene Krox20 regulates HoxB2 (Hox2.8) during hindbrain segmentation. *Cell* **72**, 183-96.
Sun, Z., et al. (2008). Discovery and characterization of three novel synuclein genes in zebrafish. *Dev. Dyn.* **237**, 2490-5.
Tanaka, H., et al. (2007). Novel mutations affecting axon guidance in zebrafish and a role for plexin signalling in the guidance of trigeminal and facial nerve axons. *Development* **134**, 3259-69.
Tendeng, C., et al. (2006). Cloning and embryonic expression of five distinct sfrp genes in the zebrafish Danio rerio. *Gene Expr Patterns* **6**, 761-71.
Theil, T., et al. (1998). Segmental expression of the EphA4 (Sek-1) receptor tyrosine kinase in the hindbrain is under direct transcriptional control of Krox-20. *Development* **125**, 443-52.
Thisse, B., Pflumio, S., Fürthauer, M., Loppin, B., Heyer, V., Degrave, A., Woehl, R., Lux, A., Steffan, T., Charbonnier, X.Q. And Thisse, C. (2001). Expression of the zebrafish genome during embryogenesis (NIH R01 RR15402). ZFIN Direct Data Submission (http://zfin.org).
Thisse, B., Thisse, C. (2004). Fast Release Clones: A High Throughput Expression Analysis. ZFIN Direct Data Submission (http://zfin.org).
Thisse, B., Wright, G.J., Thisse, C. (2008). Embryonic and Larval Expression Patterns from a Large Scale Screening for Novel Low Affinity Extracellular Protein Interactions. ZFIN Direct Data Submission (http://zfin.org).
Thisse, C., And Thisse, B. (2005). High Throughput Expression Analysis of ZF-Models Consortium Clones. ZFIN Direct Data Submission (http://zfin.org).
Thompson, M. A., et al. (1998). The cloche and spadetail genes differentially affect hematopoiesis and vasculogenesis. *Dev. Biol.* **197**, 248-69.
Vanderlaan, G., et al. (2005). Gli function is essential for motor neuron induction in zebrafish. *Dev. Biol.* **282**, 550-70.
Wang, H., et al. (2007). Isolation and expression of zebrafish zinc-finger transcription factor gene tsh1. *Gene Expr Patterns* **7**, 318-22.
Webb, K. J., et al. (2011). The Enhancer of split transcription factor Her8a is a novel dimerisation partner for Her3 that controls anterior hindbrain neurogenesis in zebrafish. *BMC Dev. Biol.* **11**, 27.
Woods, I. G., et al. (2005). The you gene encodes an EGF-CUB protein essential for Hedgehog signaling in zebrafish. *PLoS Biol.* **3**, e66.

Xu, Q., et al. (1995). Expression of truncated Sek-1 receptor tyrosine kinase disrupts the segmental restriction of gene expression in the Xenopus and zebrafish hindbrain. *Development* **121**, 4005-16.
Yao, J., et al. (2010). Atoh8, a bHLH transcription factor, is required for the development of retina and skeletal muscle in zebrafish. *PLoS One* **5**, e10945.
Yeo, S. Y., et al. (2001). Overexpression of a slit homologue impairs convergent extension of the mesoderm and causes cyclopia in embryonic zebrafish. *Dev. Biol.* **230**, 1-17.
Yu, H. H., et al. (2005). Semaphorin signaling guides cranial neural crest cell migration in zebrafish. *Dev. Biol.* **280**, 373-85.
Zannino, D. A., et al. (2014). prdm12b specifies the p1 progenitor domain and reveals a role for V1 interneurons in swim movements. *Dev. Biol.* **390**, 247-60.
Zhao, Q., et al. (2005). Expression of cyp26b1 during zebrafish early development. *Gene Expr Patterns* **5**, 363-9.
Zhu, S., et al. (2012). Activated ALK collaborates with MYCN in neuroblastoma pathogenesis. *Cancer Cell* **21**, 362-73.