

Reviewed Preprint

v2 • November 27, 2024

Revised by authors

Reviewed Preprint

v1 • June 5, 2024

Single Cell Sequencing Provides Clues about the Developmental Genetic Basis of Evolutionary Adaptations in Syngnathid Fishes

Hope M Healey, Hayden B Penn, Clayton M Small, Susan Bassham, Vithika Goyal, Micah A Woods, William A Cresko 

Institute of Ecology and Evolution, University of Oregon, Eugene, United States • School of Computer and Data Science, University of Oregon, Eugene, United States • Knight Campus for Accelerating Scientific Impact, University of Oregon, Eugene, United States

 https://en.wikipedia.org/wiki/Open_access Copyright information

eLife Assessment

This study provides a **valuable** new resource to investigate the molecular basis of the particular features characterizing the pipefish embryo. The authors found both unique and shared gene expression patterns in pipefish organs compared with other teleost fishes. The **solid** data collected in this unconventional model organism will give new insights into understanding the extraordinary adaptations of the Syngnathidae family and will be of interest in the domain of evolution of fish development.

<https://doi.org/10.7554/eLife.97764.2.sa4>

Abstract

Seahorses, pipefishes, and seadragons are fishes from the family Syngnathidae that have evolved extraordinary traits including male pregnancy, elongated snouts, loss of teeth, and dermal bony armor. The developmental genetic and cellular changes that led to the evolution of these traits are largely unknown. Recent syngnathid genome assemblies revealed suggestive gene content differences and provide the opportunity for detailed genetic analyses. We created a single cell RNA sequencing atlas of Gulf pipefish embryos to understand the developmental basis of four traits: derived head shape, toothlessness, dermal armor, and male pregnancy. We completed marker gene analyses, built genetic networks, and examined spatial expression of select genes. We identified osteochondrogenic mesenchymal cells in the elongating face that express regulatory genes *bmp4*, *sfrp1a*, and *prdm16*. We found no evidence for tooth primordia cells, and we observed re-deployment of osteoblast genetic networks in developing dermal armor. Finally, we found that epidermal cells expressed nutrient processing and environmental sensing genes, potentially relevant for the brooding environment. The examined pipefish evolutionary innovations are composed of recognizable cell types, suggesting derived features originate from changes within existing gene networks.

Future work addressing syngnathid gene networks across multiple stages and species is essential for understanding how their novelties evolved.

Introduction

Seahorses, pipefishes, and seadragons are extraordinary fishes in the family Syngnathidae with diverse body plans, coloration, and elaborate structures for paternal brooding. The syngnathid clade comprises over 300 diverse species that vary in conservation status, distribution, ecology, and morphology (Leysen et al., 2011 [↗](#); Manning et al., 2019 [↗](#); Schneider, Woltering, et al., 2023 [↗](#); Stiller et al., 2022 [↗](#)). Syngnathids have numerous highly altered traits, trait losses, and evolutionary novelties. They have elongated snouts bearing small, toothless jaws (Leysen et al., 2011 [↗](#)) specialized for capture of small zooplankton (Van Wassenbergh et al., 2009 [↗](#)). Additionally, syngnathids are distinctly protected by a bony dermal armor rather than scales (Jungerson, 1910 [↗](#)). Other skeletal differences include a lack of ribs and pelvic fins, and an expansion in the number of vertebrae (Schneider, Woltering, et al., 2023). Finally, syngnathids exhibit male pregnancy, which has involved the evolution of specialized brooding tissues and structures (Whittington & Friesen, 2020 [↗](#)). Paternal investment varies among lineages; for example, seadragons tether embryos externally to their tails while seahorses and some pipefishes have enclosed brood pouches proposed to support embryos through nutrient transfer and osmotic regulation (Carcupino, 2002 [↗](#); Melamed et al., 2005 [↗](#); Ripley & Foran, 2006 [↗](#)).

Despite advances in understanding the ecology and evolution of syngnathid novelties, the developmental genetic basis for these traits is largely unknown. The recent production of high quality syngnathid genome assemblies (Qu et al., 2021 [↗](#); Ramesh et al., 2023 [↗](#); Small et al., 2022 [↗](#); Wolf et al., 2024 [↗](#)) provides initial clues for the developmental genetic basis of some evolutionary changes. Studies have found that syngnathids lack several genes with deeply conserved roles in vertebrate development, including pharyngeal arch development (*fgf3*), tooth development (*fgf3*, *fgf4*, and *eve1*, and most Scpp genes), fin development (*tbx4*), and immune function (MHC pathway components) (Lin et al., 2016 [↗](#); Qu et al., 2021 [↗](#); Small et al., 2016 [↗](#); Small et al., 2022 [↗](#); Zhang et al., 2020). Though these gene losses are highly suggestive of leading to unique changes, exploration of the actual developmental consequences of their losses is needed.

To fill this gap in knowledge, we used single cell RNA sequencing (scRNAseq) to investigate how these striking genomic changes have affected the developmental genetic and cellular basis of syngnathids' derived traits. The Gulf pipefish (*Syngnathus scovelli*) is an attractive model for this study (Figure 1A [↗](#)). This species has a high-quality reference genome annotated by NCBI and is amenable to laboratory culture (Anderson & Jones, 2019 [↗](#); Ramesh et al., 2023 [↗](#)). Furthermore, species from the *Syngnathus* genus are used worldwide to address questions about syngnathid evolution in microbial, developmental, functional morphological, histological, transcriptomic, ecotoxicological, and genomic studies (Berglund et al., 1986 [↗](#); Carcupino, 2002 [↗](#); Fuiten & Cresko, 2021 [↗](#); Harada et al., 2022 [↗](#); Harlin-Cognato et al., 2006 [↗](#); Partridge et al., 2007 [↗](#); Ripley & Foran, 2006 [↗](#); Rose et al., 2023 [↗](#); Roth et al., 2012 [↗](#); Small et al., 2016 [↗](#); Small et al., 2013 [↗](#)). In this paper, we focus on a subset of unique syngnathid traits, including their elongated head, toothlessness, dermal armor, and development of embryos inside the brood pouch. These traits represent the diversity of evolutionary changes observed in the syngnathid clade (highly altered, lost, and novel traits); hypotheses from studies of model organisms suggest developmental pathways involved in their evolution (Lin et al., 2016 [↗](#); Roth et al., 2020 [↗](#); Small et al., 2016 [↗](#); Small et al., 2022 [↗](#)).

scRNAseq atlases are a powerful complement to genomic analyses (Shema et al., 2019 [↗](#); Ton et al., 2020 [↗](#)). They can provide crucial insights into types of cells present, genes that distinguish cell types (marker genes), active gene networks, and a means to identify expression of genes of interest within predicted cell types (Farnsworth et al., 2020 [↗](#); Farrell et al., 2018 [↗](#); Williams et al., 2019).

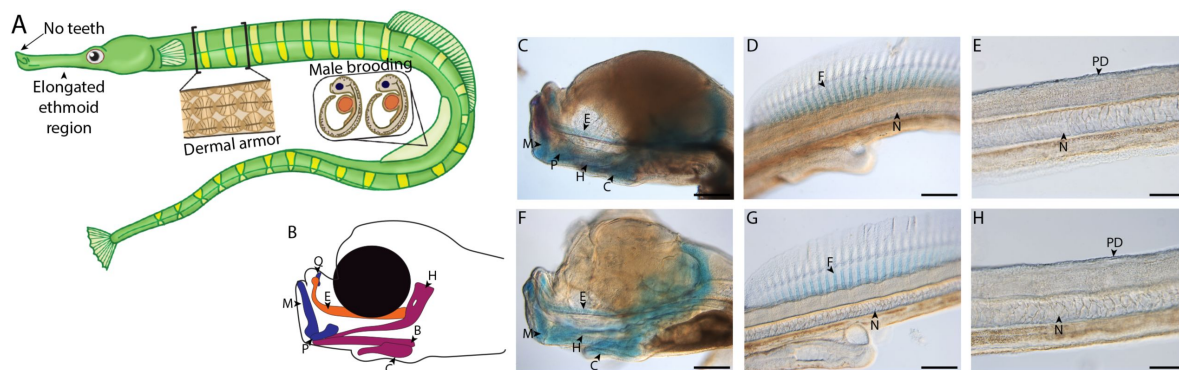


Figure 1

Gulf pipefish exemplify syngnathid derived traits.

Gulf pipefish have elongate snouts, have lost teeth on their oral and pharyngeal jaws, possess dermal armor, and have brood pouches in males (A-B). Cartilage (alcian) and bone (alizarin) stained clutch siblings of embryos from the two scRNAseq samples are shown in panels C-H. Embryos have cartilaginous craniofacial skeletons (panels B, C, F; E marks the (Mes)ethmoid cartilage, C indicates the Ceratohyal, H shows the Hyosymplectic cartilage, M marks the Meckel's cartilage, Q indicates the quadrate, B shows the Basihyal, and P marks the palatoquadrate) with the onset of ossification in the jaw. They have cartilaginous fin radials in the dorsal fin (D and G; F indicates fin radials and N denotes the notochord). The embryos do not have signs of ossification in the trunk where the exoskeleton will form later (panels E and H; PD shows the region where dermal armor primordia will arise). Panels C, D, F, and G, scale bar 200 μ m; E, H scale bar 100 μ m.

Specifically, scRNAseq captures RNA expression profiles from individual cells, allowing cell types to be inferred *post-hoc*. scRNAseq has successfully been applied to syngnathid adult kidneys (Parker et al., 2022 [DOI](#)), but there are no published syngnathid developmental atlases.

Here we report the first developmental scRNAseq atlas for syngnathids from late embryogenesis staged Gulf pipefish. We delineate the overall structure of this atlas, which describes 38 cell clusters composed of 35,785 cells, and use these data to make inferences about the morphological evolution of several syngnathid innovations. In addition to inferring present cell types and their underlying genetic networks, we detail spatial expression patterns using *in situ* hybridization experiments of select marker and other candidate genes in pipefish embryos and juveniles. We found genes of conserved signaling pathways expressed during craniofacial development but did not detect evidence of tooth primordia. The embryonic dermis and epidermis, respectively, expressed genes for the dermal armor development (bone development pathways) and genes potentially involved in interaction with the male brood pouch (e.g., nutrient acquisition genes). Overall, this atlas provides a deeper understanding of the development of Gulf pipefish and identifies gene candidates for understanding the development of syngnathid evolutionary innovations. In addition to these discoveries, this atlas provides a significant resource for researchers studying syngnathid evolution and development.

Results

Valuable scRNAseq atlas for studying syngnathid development

We produced the first developmental scRNAseq atlas for a syngnathid from two samples comprising 20 similarly staged embryos from pregnant, wild-caught Gulf pipefish (*Syngnathus scovelli*) males. The samples represent a late organogenesis developmental stage (Figure 1B-H [DOI](#)). These embryos had a primarily cartilaginous skeleton with minimal mineralization, including jaw cartilages that were at the onset of mineralization and ethmoid elongation. The embryos also possessed cartilaginous dorsal fin pterygiophores but had no signs of dermal armor mineralization. This stage is referred to as “frontal jaws” in the literature on syngnathids (Sommer et al., 2012 [DOI](#)).

The atlas included 35,785 cells (19,892 and 15,893 cells from each sample; S1:S2), which formed 38 cell clusters (Figure 2A [DOI](#), Supplementary File 1, 2). We classified cells into 4 different broad tissue types – epithelial, connective, neural, and muscle – using Seurat-identified marker genes and published model organism resources (S3:S57). We next used Seurat identified marker genes to pinpoint single marker genes that were most unique to each cluster (Figure S58, Supplementary File 3, 4, 5). We completed *in situ* hybridization using Gulf and bay pipefish embryos for cell clusters for which examining gene expression would help hone and validate cluster annotations (Figure S59:S71).

In total, our atlas contained 13,027 connective tissue cells (excluding cells from blood, immune, and the digestive system) from 14 clusters, 10,112 nervous system cells from ten clusters, 4,363 muscle cells from five clusters, 4,133 blood cells from three clusters, 650 immune cells from two clusters, 432 pigment cells from one cluster, 370 epidermal cells from one cluster, and 137 gut cells from one cluster. Within the connective tissue cell types, we also identified cartilage (302 cells), developing bone (442 cells), fins (253 cells), and notochord (693 cells). The number of recovered cells per identity may not necessarily represent organismal cellular proportions because of potential variability in dissociation success for different cell types (Denisenko et al., 2020 [DOI](#); Venema et al., 2022).

Discovery of cell cluster function and state using KEGG analysis

To affirm identities and discover potential properties of each cluster, we completed a KEGG pathway analysis for each cluster using Seurat's marker genes (Figure S72, [Figure 2B](#)). For eight of the clusters (1, 4, 6, 9, 11, 15, 19, and 24), we did not find any significantly enriched pathways, possibly due to similar gene expression profiles across cell types that reduced the number of identified markers. However, we found one or more significantly enriched pathways for the other 29 cell clusters. We observed enriched pathway terms that supported cluster annotations. For example, 'phototransduction' in the retina cluster, 'melanogenesis' in the pigment cluster, 'cardiac muscle contraction' in muscle clusters, and 'neuroactive ligand receptor interaction' in neuronal clusters.

The inferred KEGG pathways demonstrated some commonalities across the different tissue types, including in signaling pathways and cell states. Notably, our identified KEGG terms delineated progenitor and differentiated cell clusters. Based on their KEGG terms, we classified clusters 8, 10, and 16 as possible neural, muscle, and connective tissue progenitor cells, respectively. We also detected expression of *pax3a* and *pax3b*, muscle primordia markers, in cluster 10, supporting this annotation. These clusters had enriched KEGG terms associated with cell division ('cell cycle', 'DNA replication', 'nucleotide excision repair', and 'homologous recombination'), and lacked enrichment for KEGG pathways present with differentiated cell types of their lineage. Specifically, cluster 8 lacked the neural KEGG term 'neuroactive ligand receptor interaction', cluster 10 lacked muscle KEGG terms 'adrenergic signaling in cardiomyocytes', 'calcium signaling pathways', and 'cardiac muscle contraction', and cluster 16 lacked connective tissue term 'ECM receptor interaction'. To complement these findings, we completed a cell differentiation analysis using CytoTRACE for neural, muscle, and connective clusters (Figure S73). Clusters 8, 10, and 16 had the lowest scores in each respective comparison, which indicated undifferentiated cell states. Thus, it is likely that clusters 8, 10, and 16 represented undifferentiated cells within the major lineages of neural, muscle, and connective cells.

Commonalities of cell clusters, unique networks, and elusive cell types identified in network analysis

We built gene networks/modules from 3,000 variable genes using weighted gene network correlation analysis (WGCNA; [Langfelder & Horvath, 2008](#)). This produced 43 gene modules in total (Supplementary File 8, 9), assessed for each cluster-module pair for their strength of association ([Figure 3A](#), Supplementary File 10, 11) and every module's dependence on each cluster for their network connectivity (Figure S74, Supplementary File 12, 13). Using the genes from each network, we completed a KEGG pathway analysis to identify whether gene modules indicated specific cellular pathways or states ([Figure 3B](#)). We initially explored whether these network-cluster associations could reveal commonalities between cell clusters or identify whether particular clusters contained multiple cell identities.

First, we asked whether gene modules that associate with three or more cell clusters signify commonalities between clusters that have similar cell types. We identified seven gene modules (6, 7, 21, 14, 17, 41, and 42) that each associate with three or more cell clusters. These gene modules do connect clusters of similar identities or cell states. For example, Modules 6 and 7 associate with connective tissue cells. Module 6, the larger of the two modules, contains numerous other KEGG pathways found in most connective cell clusters, such as 'cytokine-cytokine receptor interaction' ([Figure 3C](#)). Interestingly, module 21, associated with clusters 8, 10, 16, and 25, contains genes from KEGG pathways related to the 'cell cycle' and 'cellular senescence', supporting the results found in our cluster based KEGG pathway analysis ([Figure 3D](#)). Module 21 also contained 'Notch signaling' genes, possibly due to similar correlations with cell cycle genes; however, these were only expressed in the neural cell clusters (8 and 25).

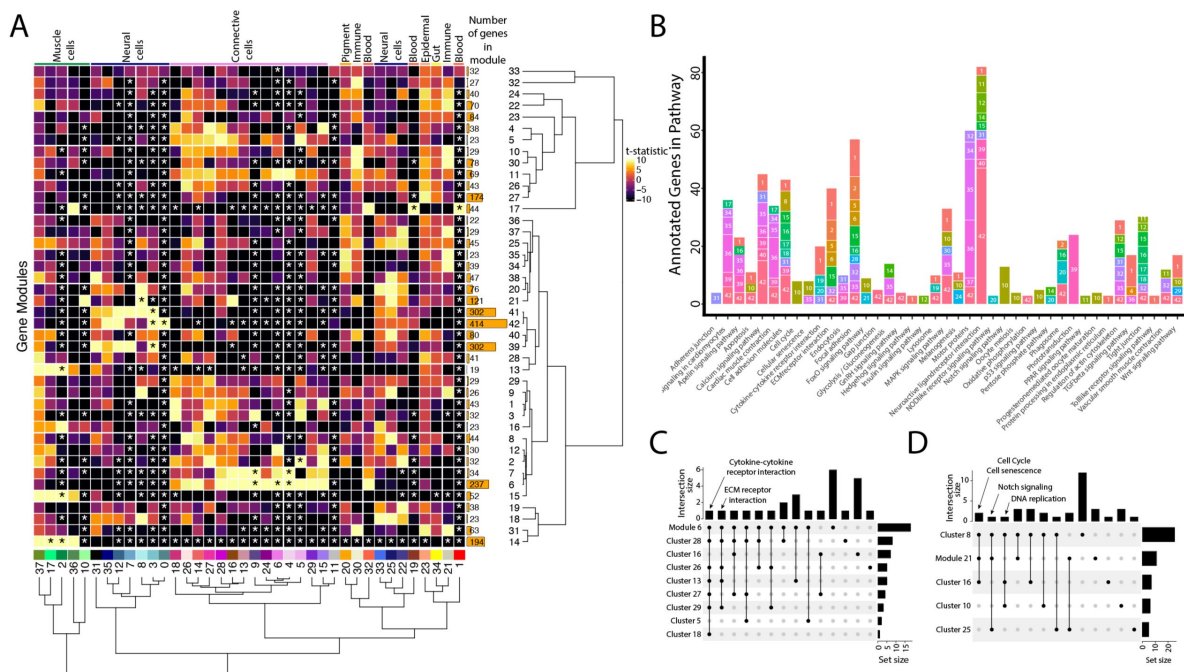


Figure 3

Weighted Gene Network Analysis (WGCNA) identifies gene modules that define and unite cell clusters.

A) The strength of association between the gene modules and cell clusters is shown in panel A with dendrogram clustering illustrating the distance between modules and cell clusters. Gene modules are represented by rows and cell clusters by columns. The modules and clusters are clustered using the Pearson distance method. The number of genes in each gene module are shown in the right hand bar plots. Cell clusters are colored based on their identity. The asterisks indicate the module-cluster relationships that have a p-value less than .05 from a two-sided permutation test after correction for multiple tests (FDR). The heatmap is colored by t-statistics in a range of -10 to 10, with highly positive values in yellow and highly negative values in black. B) The identified gene modules possess genes from KEGG pathways. The bars are labeled with the gene module and the size of each bar corresponds to the number of genes from the KEGG pathway in the module. Since WGCNA modules do not have p-values, only KEGG pathways with more than two genes included in the gene module are shown on the plot. C) Identified gene modules contain similar KEGG pathways as the cell clusters that correlated with them. These relationships are shown in Upset plots where each row is a cell cluster or gene module, each column represents KEGG pathways shared by the modules and clusters (shared condition is shown filled in black dots connected by lines), the intersection size is the number of pathways in common between the set of modules and clusters, and the set size is the number of pathways that are enriched in each cluster and module. Panel C1 highlights that 'cytokine-cytokine receptor interaction' and 'ECM receptor interaction' are present in module 6 as well as 6 and 4 connective cell clusters respectively. Panel C2 shows that 'cell cycle' and 'senescence' are present in module 21 as well as clusters 8 and 16, 'Notch signaling' genes are present in clusters 8 and 25 as well as module 21, and 'DNA replication' is present in clusters 8, 10, and 16.

Where one cell cluster is associated with multiple gene networks, we wondered if multiple cell identities existed within the cell cluster. We explored this possibility by examining the pigment cell cluster (cluster #20) and whether the five different correlated gene modules (#34, 35, 36, and 37, and 38) are expressed in distinct subgroups of cells within the cluster (S75). We found cases of more than one genetic network (modules 35 and 38) expressed in the same cells. Modules 35 and 38 included conserved pigment genes, *pmela*, *mlana*, and *dct* in module 35 and *try1b* and *pmel* in module 38, that mark melanocytes (Figure S6; Du et al., 2003 [DOI](#); Johnson et al., 2011 [DOI](#); Lamason et al., 2005 [DOI](#); Thisse & Thisse, 2004 [DOI](#)). However, we also found non-overlapping expression of networks, notably modules 36 and 37. We inferred that module 36 is associated with xanthophores and xanthoblasts due to the presence of *plin6* and *scarb1*, genes involved with lipid binding and activity in xanthophores (Ahi et al., 2020 [DOI](#)). On the other hand, module 37 likely represents iridophores and iridoblasts because it contains *pnp4a*, which is involved in purine-nucleoside phosphorylase activity in iridoblasts (Kimura et al., 2017 [DOI](#)).

Conserved signaling pathways are active during syngnathid craniofacial development

The specialized pipefish feeding apparatus is composed of an elongate, tubular snout, toothless mandible and pharyngeal jaws, large tendons, and associated muscles. Therefore, numerous cell types contribute to their distinct faces: cartilage, bone, tendon, muscle, and connective tissues as well as their progenitors. We sought to identify markers of these cell types and signaling pathways active in them.

We found marker genes uniquely expressed in the face, genes that mark cell types important to craniofacial development, and markers with potentially relevant functions for craniofacial development using *in situ* hybridizations of cell cluster marker genes. For instance, we observed the marker for osteochondro-mesenchymal cells (cluster #6), *elnb*, specifically expressed at the intersection between the ethmoid plate and palatoquadrate as well as on the Meckel's cartilage (Figure 4B [DOI](#)). Although *elnb* is observed in the zebrafish cranial skeleton, it is primarily studied for its proposed role in teleost heart evolution (Miao et al., 2007 [DOI](#); Moriyama et al., 2016 [DOI](#)).

Other genes identified here as cell markers were not uniquely craniofacial but provide insights into the cell types that comprise the face. For example, *tnmd* expression marked tendons and ligaments (cluster #9) throughout the face and body (Figure 4C [DOI](#); Figure S65). Our finding is consistent with *tnmd*'s role in tenocyte development in model systems, namely zebrafish and mouse (Chen & Galloway, 2014 [DOI](#); Docheva et al., 2005 [DOI](#)). Identifying tenocyte cells is particularly relevant in syngnathid fishes where tendons are enlarged and store elastic energy necessary for their specialized feeding (Van Wassenbergh et al., 2007 [DOI](#)).

The last category of markers contained genes with regulatory roles important for craniofacial development such as *prdm16*, the marker for osteochondro-mesenchymal cells (cluster #5). The gene *prdm16* mediates the methylation of histones and regulates gene expression, key for promoting craniofacial chondrocyte differentiation (Ding et al., 2013 [DOI](#); Kaneda-Nakashima et al., 2022 [DOI](#); Lomeli Carpio Shull et al., 2020 [DOI](#)). We found *prdm16* expressed in mesenchymal cells directly above the ethmoid plate and in fins (Figure 4D [DOI](#); Figure S61). Comparison of pipefish with a related, short snouted fish (stickleback) identified similarity in fin expression, but differences in craniofacial staining (S76). Stickleback expressed *prdm16* in the hindbrain, gill arches, and lower jaw, consistent with published zebrafish data (Ding et al., 2013 [DOI](#)), but staining was additionally detected in the ethmoid region in pipefish.

We next examined signaling pathways active in craniofacial development. Our KEGG pathway analysis revealed that MAPK, Wnt, and TGF-beta signaling pathways were significantly enriched in one or more craniofacial contributing cell clusters (module #6; Figure 2B [DOI](#), Figure 3C [DOI](#), Figure 4A [DOI](#)). In addition to these KEGG findings, expression of ligands and receptors that are members

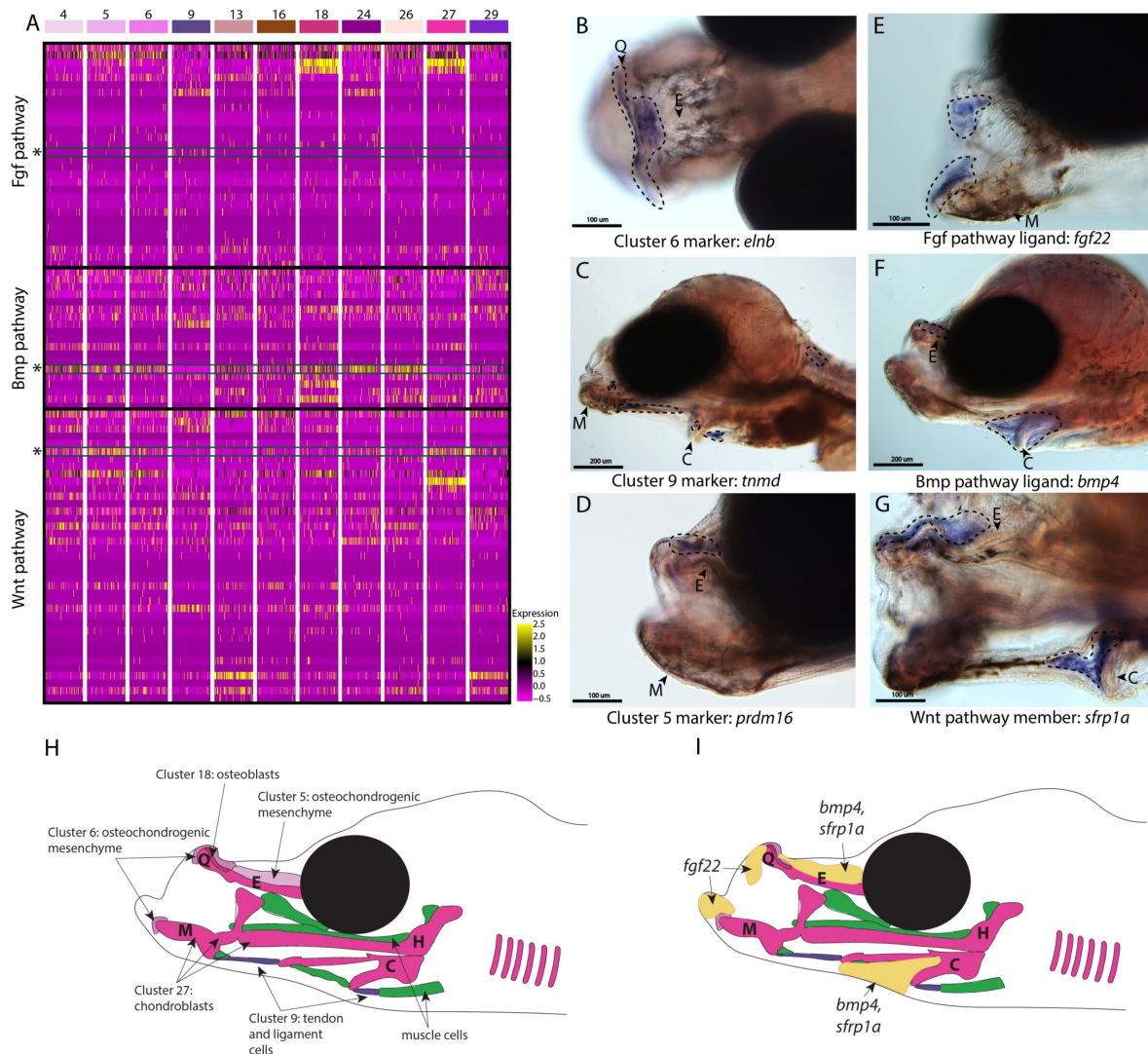


Figure 4

Conserved cell types and gene pathways build unique faces of syngnathids.

Three main conserved signaling pathways are enriched in connective cell types, MAPK signaling (including Fgf signaling), TGF-beta signaling (including Bmp signaling), and Wnt signaling. Receptors and ligands expression patterns are shown in panel A heatmap from all cell types with cells present in the head. This heatmap features 100 cells downsampled from each cluster and illustrates that many genes from these families are expressed in these cells. Yellow lines indicate high expression of a gene, while hot pink lines indicate no expression. The pathways are boxed in black. Rows representing *fgf22*, *bmp4*, and *sfrp1a* expression are marked with an asterisk and green box for each respective section of signaling (Fgf, Bmp, and Wnt). Panels B, C, and D are *in situ* hybridizations of three marker genes, *elnb*, *tnmd*, and *prdm16*. The genes *prdm16* and *elnb* mark osteochondrogenic mesenchyme and *tnmd* marks tendons and ligaments. Panels E, F, and G, show expression patterns of three pathway representatives *fgf22*, *bmp4*, and *sfrp1a*. All three genes are expressed in the face: *fgf22* at the tip of the mandible and *bmp4* and *sfrp1a* above the ethmoid and near the ceratohyal. Staining is circled with dashed lines. The Meckel's cartilage (M), (mes)Ethmoid cartilage (E), Quadrate (Q), and Ceratohyal (C) are labeled. Panel C is a dorsal view. Panels B, C, E, F, G, H, I are in lateral view. *In situ* experiments of *fgf22* were completed using 10dpf Gulf pipefish. *In situ* experiments of *bmp4*, *sfrp1a*, *tnmd*, *elnb*, and *prdm16* were completed using wild caught bay pipefish at the onset craniofacial elongation. Panels H and I are summary illustrations of our findings, panel H shows where cells from various clusters were present in the developing head and panel I illustrates where *bmp4*, *sfrp1a*, and *fgf22* were expressed. Panel H is colored according to cell type.

of these pathways (Fgf, a MAPK pathway, Wnt, and Bmp, a TGF-beta pathway) was observed in cell clusters of both differentiated and undifferentiated states (**Figure 4A**). We chose three genes, one from each major pathway, for *in situ* hybridizations: *fgf22* (a Fgf ligand present in the actively dividing cells' module), *bmp4* (a Bmp ligand present in the largest connective tissue module), and *sfrp1a* (a Wnt pathway enabler present in a cartilage gene module). Although mouse and zebrafish studies identified *fgf22* expression only in the nervous system (Miyake & Itoh, 2013; Umemori et al., 2004), we found *fgf22* is expressed at the tip of the palatoquadrate and Meckel's cartilage, the gill arches, fins, and brain in pipefish (**figure 4E**; S77). These same expression patterns were observed in stickleback, suggesting a surprising co-option of fin and craniofacial expression in percomorph fish.

We observed *bmp4* and *sfrp1a* expressed above the ethmoid plate and along the ceratohyal in pipefish (**figure 4F-G**). The gene *bmp4* has a conserved role in craniofacial development, particularly important at later stages for driving chondrocyte differentiation (Wang et al., 2024; Zhou et al., 2013). However, the specificity of *bmp4* expression to mesenchyme around the ceratohyal and ethmoid was not observed in stickleback fish, which had broad craniofacial expression (including the jaws, tooth germs, and gill arches; S78). Interestingly, *sfrp1a* expression has not been observed in the palate of mice or ethmoid region of zebrafish, but *sfrp1a* craniofacial expression was identified in stickleback (lower jaw and gill arches; S79) and has been observed in other fishes (Ahi et al., 2014; Schneider, Gunter, et al., 2023; Schilling & Kimmel, 1997; Swartz et al., 2011; Wang et al., 2024).

Gulf pipefish retain tooth development genes but likely lack onset of tooth development

Previous papers have identified possible candidate genes for the loss of teeth in syngnathid fishes including those from genes that initiate tooth bud formation (*fgf4*, *eve1*), regulate tooth morphogenesis (*fgf3*, *fgf4*), and synthesize tooth minerals (*scpp4*, *scpp7*, *scpp9*, *odam*, and *scpp5*; Lin et al., 2016; Qu et al., 2021; Small et al., 2016; Small et al., 2022; Zhang et al., 2020). However, it is unknown whether syngnathid tooth development initiates then halts or whether it never begins. We searched for signs of early tooth primordia within our atlas to ask whether tooth development might initiate in syngnathids. Additionally, we examined whether genes present in mature teeth are still expressed in syngnathids and what types of cells express them.

Our thorough examination of cell clusters for identity annotation did not find a tooth primordium cluster. We therefore searched for tooth primordia by examining expression of specific odontogenesis marker genes (*aldh1a2*, *bmp4*, *dlx2a*, *dlx3b*, *lef1*, *lhx6a*, *lhx8*, *msx1a*, *msx2*, and *wnt10a*; **Figure 5A**). We observed several primordium genes expressed in our atlas. However, there was no cluster with every marker gene expressed in over 10% of cells. Several markers (*dlx3b*, *lef1*, *msx1a*, *msx2*, and *wnt10a*) were expressed in cluster #28, a fin cluster distinguished by *hoxa13a* and *hoxa13b* expression. Since we previously noted that cluster #16 seems to be a primordial connective tissue cluster, we wondered if it could contain tooth primordial cells. In this cluster, we found the following percentage of expression of each gene in the cells: *bmp4* in 46%, *aldh1a2* in 29%, *lef1* in 24%, *dlx2a* in 18%, *dlx3b* in 16.4%, *eda* in 12.7%, *msx1a* in 10.6%, *pitx2* in 10%, *lhx6a* in 8%, *wnt10a* in 4.6%, *msx2* in 2.65%, *lhx8* in 1.8%, and *shha* in 1%.

We suggest that, given the low expression of most tooth marker genes, cluster #16 is unlikely to contain tooth primordial cells. To test this hypothesis, we examined spatial gene expression using *in situ* hybridization of *bmp4*, *pitx2*, *lhx6a*, *dlx2a*, and *dlx3b* in pipefish to ask whether the two definitive cell types are present, namely the dental epithelium (marked by *pitx2* and *bmp4*) and dental mesenchyme (distinguished by *dlx2a*, *dlx3b*, and *lhx6a*; **Figure 5B-F**; Gibert et al., 2019; A. Tucker & Sharpe, 2004). Tooth-specific expression of *dlx2a* is observed solely in the dental mesenchyme in zebrafish and mice, however, it is also expressed in the dental epithelium in medaka (Stock et al., 2006). For this study, we labelled *dlx2a* as a mesenchyme marker, though it

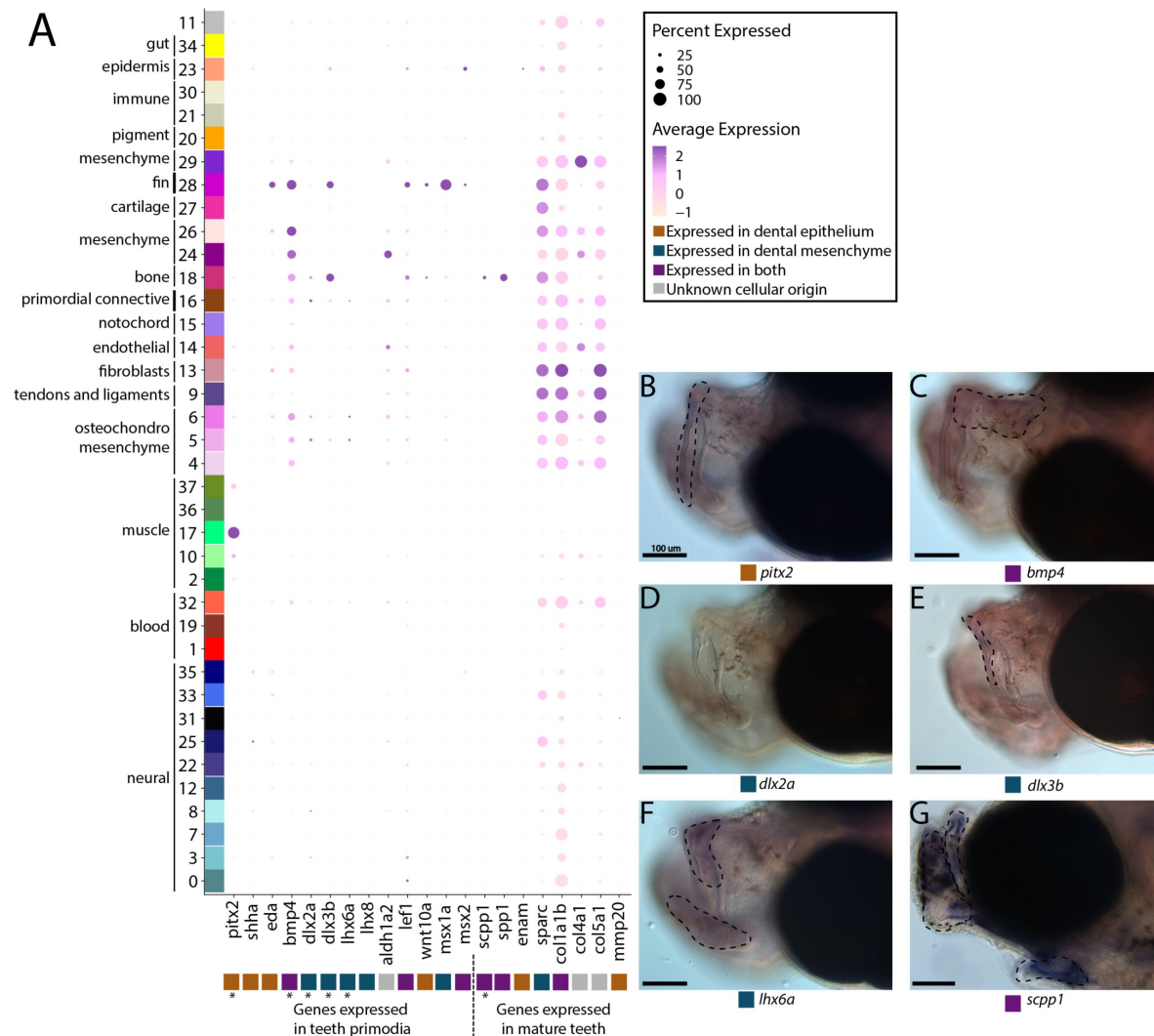


Figure 5

Pipefish do not possess identifiable tooth primordium cells, but continue to express tooth development genes in other contexts.

Panel A presents a dotplot of genes classified by the tissue layer in which they are reported to be expressed during tooth development in other vertebrate (Tucker and Sharpe 2004 [\[4\]](#); Gilbert et al. 2019; Kawasaki 2009 [\[4\]](#)). The x-axis contains the assayed genes, with asterisks under the genes that were also examined with *in situ* hybridizations. The y-axis contains all the cell clusters. The size of the dot is representative of the percentage of cells from the cluster that express the gene. The color of the dots is average expression of the gene in the cluster (darker purples represent higher expression). Panel B includes *in situ* hybridizations of selected tooth primordia markers (*bmp4*, *pitx2*, *lhx6a*, *dlx2a*, and *dlx3b*) and mature tooth marker (*scpp1*). *In situ* experiments of *bmp4*, *pitx2*, *lhx6a*, *dlx2a*, and *dlx3b* were completed with wild caught bay pipefish that had begun craniofacial elongation. *In situ* experiments of *scpp1* were completed using 9dpf Gulf pipefish. The scale bars for all images represents 100 μ m.

could be expressed in both dental tissues in a syngnathid outgroup. We found expression of all genes except *dlx2a* in the developing jaws. However, *bmp4*, *pitx2*, *lhx6a*, and *dlx3b* were expressed throughout the face rather than in the punctate pattern observed in tooth primordia development.

We next investigated *scpp* genes (*enam*, *scpp1*, and *spp1*), which are expressed in late tooth development and tooth maintenance in other vertebrates (**Figure 5A**). These genes also have some expression outside of teeth such as in dental bone (*scpp1*, *spp1*; Kawasaki, 2009) and fins (*enam*; Jain et al., 2007). We identified *spp1* and *scpp1* expression in 54.5% and 24% of bone cells, respectively, and sparse *spp1* expression (<3%) in other connective tissue cell types. We found minimal expression of *enam* (in 11.4%) in epidermal cells and in less than 8% of muscle cells. In whole mount *in situ* hybridization, we found that *scpp1* is expressed in all developing pipefish bones, both endochondral and dermal (**Figure 5G**). Since *scpp* gene losses observed in syngnathids have been hypothesized to be responsible for their tooth loss (Lin et al., 2016; Qu et al., 2021; Zhang et al., 2020), we explored *scpp* gene family content in the close, tooth-bearing relative to syngnathids, the blue spotted cornetfish, and we found several gene losses (*scpp4*, *scpp7*, and *scpp9*) and a likely pseudogene (*scpp5*) (Figure S80; Hughes et al., 2018; Stiller et al., 2022).

Exploration of additional tooth maturation genes (*col1a1b*, *col4a1*, *col5a1*, *sparc*, and *mmp20b*) similarly found that these genes were expressed in non-tooth derivatives, including connective tissue, smooth muscle, and neural cells.

Tooth and skeletal genes are expressed during dermal armor development

Syngnathid dermal armor is mineralized dermal bone underneath the skin (**Figure 6D, G, H**). It is unknown when dermal armor primordia initiate and how they are patterned. Spatial expression analysis of *pitx2* and *dlx3b* in search of tooth primordia instead revealed expression of these genes in possible dermal armor primordia (**Figure 6A,B, E,F**). We found *pitx2* staining localized dorsally to striated muscle underneath developing dermal armor. The gene *dlx3b* is expressed in a repeating pattern along the body in the epidermal and dermal tissues. Both staining patterns were in discrete regions of the muscle and epidermal layers rather than being continuously expressed across the tissues. We did not observe the expression of other tooth primordium genes (*bmp4*, *lhx6a*, and *dlx2a*) in this region.

Because the embryos from our atlas had not begun dermal armor mineralization, the atlas cannot be directly used for discovery of genes active in dermal armor. However, the atlas contains osteoblasts from craniofacial bones which we used to create osteoblast-specific gene networks. We therefore asked whether these osteoblast genes were present in mineralizing dermal armor at later stages. *In situ* hybridization expression analysis revealed that *scpp1*, an osteoblast and tooth mineralization gene, and *ifitm5*, an osteoblast gene, were expressed in the dermal armor at the onset of mineralization (**Figure 6C**; Figure S68).

Epithelial expression of immune and nutrient-processing genes may facilitate embryo-paternal interactions in the brood pouch

Within the brood pouch, embryos could interact with male placenta-like tissues, the male brood pouch epithelium, and/or the pouch microbiome. Once the thin chorion is shed, the embryos' epidermis is directly exposed to the pouch environment. We therefore asked if the embryonic epidermal cells expressed nutrient acquisition and/or immune genes that would indicate active transfer of nutrients and immune response.

Within our larger KEGG analysis, we asked whether nutrient absorption and immune KEGG terms were among the enriched pathways for the epidermal cells. We identified 106 enriched genes in the 'endocytosis' pathway (p-value=.036; **Figure 7A**). Four metabolism pathways ('galactose', 'glutathione', 'sphingolipid', and 'starch and sucrose') are also enriched. No immune-related KEGG

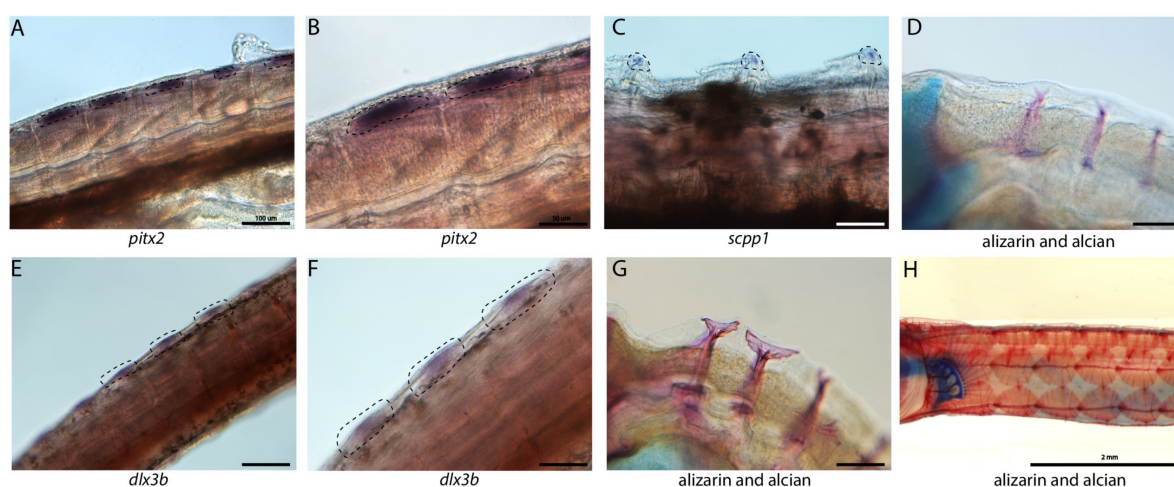


Figure 6

Tooth and bone development genes expressed during exoskeleton development. We discovered *pitx2* (A and B) and *dlx3b* (F and E) expression during possible emergence of exoskeletal primordium in wild caught bay pipefish. The embryos used for these *in situ* experiments were the same stage as those from [Figure 5](#), at the beginning of craniofacial elongation. Figures A and F have 100 μm scale bars, figures B and E have 50 μm scale bars. We further found *scpp1* is expressed at the mineralization front of the exoskeleton in 12dpf Gulf pipefish (C) and has a 50 μm scale bar. Alizarin and alcian stained pipefish are shown in panels D (12dpf Gulf pipefish), G (1dpf Gulf pipefish), and H (adult Gulf pipefish) to illustrate how the exoskeleton forms. Panels D and G have 100 μm scale bars.

terms were enriched in the epidermis. For comparison, we investigated whether these KEGG terms are also enriched in epidermal cells of non-brooding fishes. We completed a KEGG pathway analysis on a comparably staged zebrafish single cell RNA sequencing atlas (3 days post fertilization; [Lange et al., 2023](#) [↗](#)). The 3dpf zebrafish epidermal cells did not have a significant enrichment of the ‘endocytosis’ pathway (23 up-regulated genes, $p\text{-value}=0.99$) or any metabolism pathway. However, there are 11 endocytosis genes up-regulated in both zebrafish and pipefish epidermal cells, suggesting conserved expression of these genes in pipefish.

We next examined pipefish epidermal gene networks for presence of nutrient absorption or immune genes. We found that the largest epidermal gene network (#16) contained a striking enrichment of C type lectin genes, carbohydrate binding proteins that possess antimicrobial properties (**Figure 7B** [↗](#)). This network contained 14 total lectin genes expressed in the epidermal cells: five galactose-specific lectin nattolectin, 2 alpha-N-acetylgalactosamine-specific lectins, one L-rhamnose-binding lectins, four ladderlectin, one C-type lectin 37Dd-like, and one C-type lectin domain family 4 member G-like. Through examining lectin gene expression in the entire dataset, we found that these genes were specific to epidermal cells. Interestingly, previous literature has identified an upregulation of C-type lectins in brood pouch tissues throughout different stages of syngnathid pregnancy ([Roth et al., 2020](#) [↗](#); [Small et al., 2013](#) [↗](#); [Whittington et al., 2015](#)).

We did not find any expression of C-type lectin genes in zebrafish epidermal cells (Figure S81), unlike in pipefish. However, it is possible there are unannotated C-type lectin genes in zebrafish that remained cryptic in the atlas. Interestingly, the ‘C-type lectin receptor signaling pathway’ was significantly enriched in zebrafish epidermal cells (13 genes, $p\text{-value}=0.04$) but not in pipefish epidermal cells (34 genes, $p\text{-value}=0.4$). Although these results seem paradoxical, the KEGG term ‘C-type lectin signaling pathway’ does not include any of the C-type lectin genes themselves. Overall, these data suggest that the expression of C-type lectin genes in the pipefish embryonic epidermis is potentially unique and warrants further investigation.

Discussion

Our study examines the development of syngnathids, with a particular focus on novel and adaptive characters, using single cell RNA sequencing of Gulf pipefish embryos coupled with *in situ* experiments of gene expression. Our single cell atlas represents early craniofacial skeleton development in Gulf pipefish, at a stage when the cartilages of the head skeleton were formed but the face has not elongated. We used the atlas to explore craniofacial and dermal armor development and to investigate potential interactions between the embryos and the brood pouch environment. Our dataset is both an opportunity to explore the developmental genetic underpinnings of syngnathid innovations, and a resource for teleost researchers for future studies in this fascinating lineage.

Our atlas represents a novel resource for Evo-Devo research

Developmental single cell atlases have elucidated cell identities and genetic pathways active in model organisms such as zebrafish, mice, and chick ([Farnsworth et al., 2020](#) [↗](#); [Farrell et al., 2018](#) [↗](#); [Feregrino et al., 2019](#) [↗](#); [Morrison et al., 2017](#) [↗](#); [Soldatov et al., 2019](#) [↗](#); [Wagner et al., 2018](#) [↗](#); [Williams et al., 2019](#)). In less traditional models, the majority of scRNAseq atlases are produced from adult tissues, allowing investigations into cell types, population differences, and genetic networks ([Chari et al., 2021](#) [↗](#); [Fuess & Bolnick, 2023](#) [↗](#); [Hain et al., 2022](#) [↗](#); [Hong et al., 2023](#) [↗](#); [Koiwai et al., 2023](#) [↗](#); [Parker et al., 2022](#) [↗](#); [Potts et al., 2022](#) [↗](#); [Royan et al., 2021](#) [↗](#); [Songco-Casey et al., 2022](#) [↗](#); [Vonk et al., 2023](#) [↗](#); [Woych et al., 2022](#) [↗](#)). Creation of developmental scRNAseq atlases in non-model organisms is just beginning to accelerate, but many emerging models still lack such a resource, limiting our understanding of their development ([Healey et al.,](#)

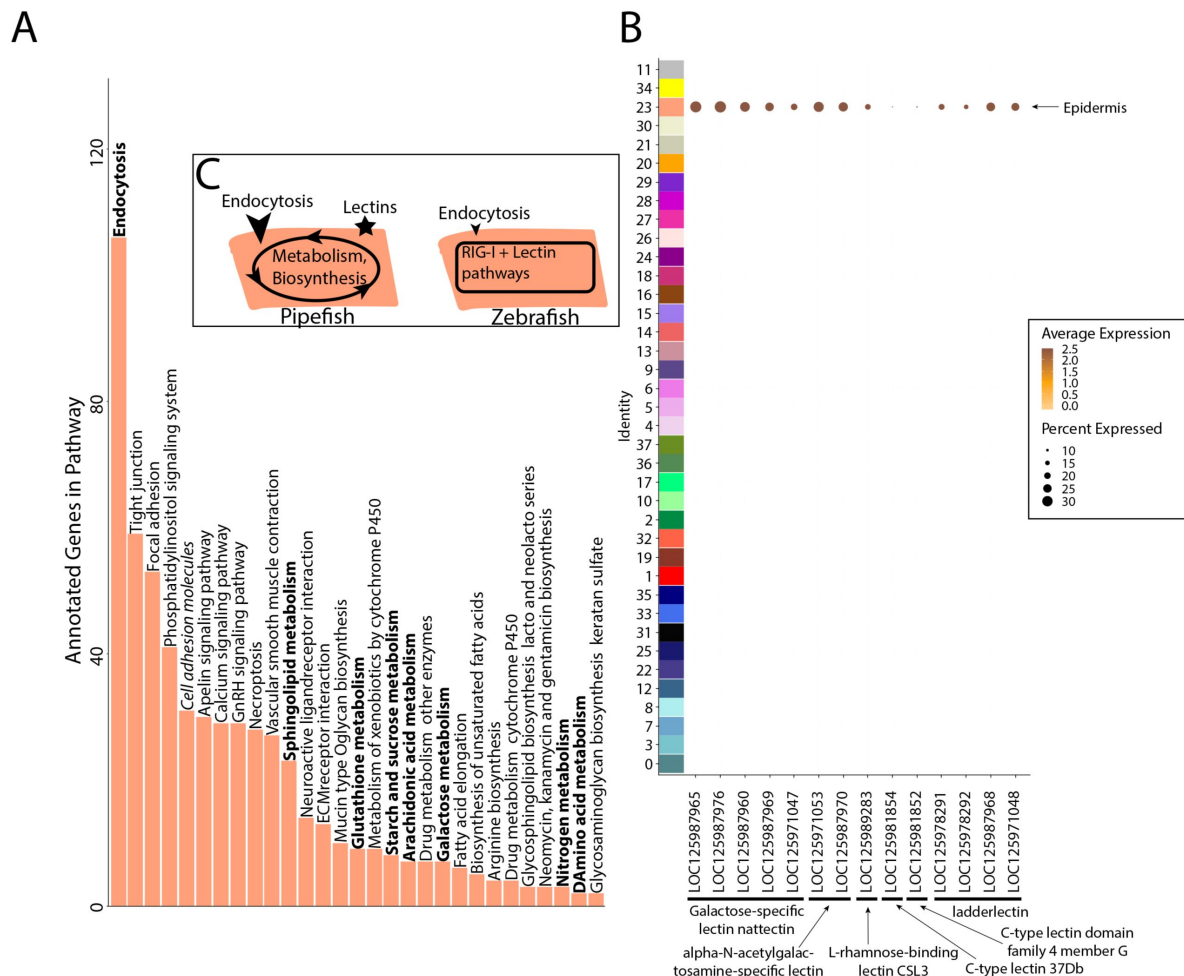


Figure 7

Gene expression signatures suggest embryonic interactions within brood pouch environment.

Epidermal cells (panel A), with pathways that suggest increased endocytosis and metabolism pathways are in bold text. Pathways upregulated in 3dpf zebrafish epidermal cells are in italics. Pipefish epidermal cells also express 15 lectin genes not found in other cell types (panel B). We suggest an epidermal cell model (panel C), in which we predict pipefish have an enrichment of nutrient processing genes and lectins in comparison with zebrafish cells.

2022 [↗](#); Salamanca-Díaz et al., 2022 [↗](#); Steger et al., 2022 [↗](#); Ton et al., 2023 [↗](#)). Our single cell atlas is one of the first created to understand the development of derived traits in a non-model organism.

For syngnathids specifically, this atlas represents an important step towards understanding the genetic nature of unique syngnathid traits. Numerous developmental genetic changes can lead to evolutionary innovations, including the evolution of novel genes, gene duplications, gene losses, gene family expansions or contractions, evolution of regulatory elements, co-option of gene regulatory networks, re-wiring of gene networks, assembly of novel gene networks, and/or the emergence of novel cells (Arendt et al., 2016 [↗](#); Căestro et al., 2007 [↗](#); Teichmann & Babu, 2004 [↗](#); Wagner, 2011 [↗](#); Wagner & Lynch, 2010 [↗](#)). Considering these possibilities, we examined select syngnathid traits and speculated on developmental genetic mechanisms influencing their evolution. Continuing to investigate these proposed mechanisms through expanded scRNAseq atlases and other studies will be critical for understanding syngnathid evolution.

Conserved pathways may contribute to derived syngnathid heads

Syngnathids have highly derived heads including an elongated ethmoid region, uniquely shaped hyoid, and altered muscles and tendons to support specialized “pivot feeding”. The developmental underpinnings of these derived traits have remained underexplored. In our atlas, we identified cell types present in the developing pipefish head and genetic pathways active in those cell types. We identified numerous cells that were present in the developing face: cartilage, bone, tendons, ligaments, osteochondrogenic mesenchyme, fibroblasts, and unclassified connective tissue cells. Overall, we did not find any unrecognizable cell types, suggesting that genetic modifications within conserved cell types may drive craniofacial modifications.

We next investigated signaling pathways expressed in these cells to determine whether and to what extent developmental genetic reorganization might have occurred. Specifically, we examined the expression of one gene each from three different highly conserved signaling pathways: Wnt (*sfrp1a*), TGF-beta (*bmp4*), and MAPK (*fgf22*). Using *in situ* hybridizations, we found *sfrp1a* and *bmp4* expressed dorsal of the elongating ethmoid plate and surrounding the ceratohyal, suggesting that Wnt and Bmp signaling may be active in the lengthening structures. These two genes are proposed to influence the development of elongated and broadened craniofacial morphologies in other species (Ahi et al., 2014 [↗](#); Schneider, Gunter, et al., 2023; Tucker et al., 2000 [↗](#)). Pipefish *prdm16* is similarly expressed dorsal to the elongating ethmoid plate. Since *prdm16* regulates Wnt and TGF-beta signaling and these genes regulate chondrocyte differentiation (Bjork et al., 2010 [↗](#); Kaneda-Nakashima et al., 2022 [↗](#); Shull et al., 2022 [↗](#); Wang et al., 2024 [↗](#)), their co-expression might suggest a prolonged period of chondrocyte differentiation along the pipefish ethmoid region. Future studies investigating this testable hypothesis would clarify whether prolonged chondrocyte differentiation broadly underlies craniofacial diversity.

We found *fgf22* expressed in the mandible, gill arches, and fins using *in situ* hybridizations, but not in the elongating regions of the head. Interestingly, *fgf22* expression has not been reported in craniofacial development of any other species (Miyake & Itoh, 2013 [↗](#)), but we found similar craniofacial and fin expression in stickleback fish. Fgf signaling, however, is a conserved and essential pathway for craniofacial development (Crump et al., 2004 [↗](#); Leerberg et al., 2019 [↗](#); McCarthy et al., 2016 [↗](#); Szabo-Rogers et al., 2008 [↗](#); Walshe & Mason, 2003 [↗](#); Woronowicz & Schneider, 2019 [↗](#)), raising the possibility that *fgf22* has been co-opted into a role played by different *fgf* genes in other species. Future work should consider a relationship between the novel craniofacial expression of *fgf22* and the loss of Fgf ligands *fgf3* and *fgf4* in syngnathids. If *fgf22* is active in existing gene networks, particularly those where *fgf3* or *fgf4* is active in other species, then its novel expression may indicate evolved genetic redundancy prior to the gene losses.

Our analysis suggests ways in which unique syngnathid craniofacial structure could have evolved through genetic network evolution. Unusual expression location (e.g., *fgf22*) and specificity (e.g., *sfrp1a*, *bmp4*, and *prdm16*) in pipefish compared to zebrafish and stickleback suggests that changes in signaling gene deployment and/or content within craniofacial gene networks, particularly genes from Wnt, Fgf, or TGF-beta families, could underly the exceptionally elongated syngnathid face.

Early, not late, tooth development is likely at the root of evolutionary tooth loss

Tooth loss has occurred independently in numerous lineages and has often been studied to understand the developmental basis of character loss. For instance, research in birds and turtles found that tooth programs initiate but are subsequently truncated, explaining toothlessness in mature animals (Chen et al., 2000 [DOI](#); Tokita et al., 2013 [DOI](#)). Additional studies in birds have found losses in tooth maturation genes (specifically *scpp* genes; Sire et al., 2008 [DOI](#)). Since numerous primordium and maturation genes are lost in syngnathids (Lin et al., 2016 [DOI](#); Qu et al., 2021 [DOI](#); Small et al., 2016 [DOI](#); Small et al., 2022 [DOI](#); Zhang et al., 2020), we asked if syngnathids begin tooth development at all.

We found that early tooth development genes were still expressed in pipefish, which is unsurprising given their pleiotropic roles, but found no convincing evidence either in the atlas or in spatial gene expression analysis of tissues for tooth primordia. The possibility remains, however, that a transient cell population could fall at a different developmental time than assayed here. Barring this caveat, syngnathid *fgf3* and *fgf4* losses could have resulted in insufficient Fgf signaling from the oral epithelium to the dental mesenchyme, causing failure of tooth initiation (Small et al., 2022 [DOI](#); Stock et al., 2006 [DOI](#)).

If our finding of the loss of earliest stages of tooth development is consistent across developmental stages and syngnathids, why then have syngnathids retained some members of the *scpp* tooth maturation gene cluster? Since studies in birds propose *scpp* gene losses can occur from relaxed selection (Sire et al., 2008 [DOI](#)), we speculated syngnathids lost *scpp* genes with expression limited to teeth and retained genes with ancestrally pleiotropic expression patterns. In Gulf pipefish, we found the retained genes *scpp1*, *spp1*, and *enam*, are expressed in structures outside of tooth development, suggesting developmental pleiotropic constraint. Specifically, we found *spp1* and *scpp1* expressed in osteoblasts which is consistent with zebrafish (Bergen et al., 2022 [DOI](#); Kawasaki, 2009 [DOI](#); Liu et al., 2016 [DOI](#)) and *enam* expressed in the epidermis which has not been reported in zebrafish (Goldsmith et al., 2003 [DOI](#); Jain et al., 2007 [DOI](#); Liu et al., 2016 [DOI](#)). Through examining the conservation of the *scpp* genes in close syngnathid relatives, we found that most *scpp* gene losses (*scpp4*, *scpp7*, and *scpp9* and a functional *scpp5*) are shared with a tooth-bearing outgroup to the family, and likely occurred prior to the loss of teeth in syngnathids. Overall, our analysis favors the hypothesis that pleiotropic *scpp* genes were retained in syngnathids while other, more tooth-specific *scpp* genes were lost due to relaxed selection.

Redeployment of the bone gene network to build dermal armor

The syngnathid dermal armor is a type of evolutionary novelty, which can arise through either differentiation of serially repeated elements or *de novo* origination, derived from either the redeployment of existing gene networks, rewiring of existing gene networks, or the assemblage of new gene networks (Wagner & Lynch, 2010 [DOI](#)). Currently, there is no understanding of the developmental genetic underpinnings of the syngnathid dermal armor.

We identified *dlx3b* and *pitx2* expression in tissues where the dermal plates mineralize later in development. Using *in situ* hybridizations, we showed that epithelial and dermal layers expressed *dlx3b* and underlying muscle cells expressed *pitx2*. In some species, dermal bone, plate, or denticle development occurs from the re-deployment of tooth gene regulatory networks (Mori & Nakamura,

2022 [\[link\]](#)). However, this does not appear to be the case in syngnathids because the dermal armor lacks the characteristic epithelial–mesenchyme interactions distinguished by *pitx2* expression in the epithelia.

Instead, dermal armor might originate from the co-option of existing bone development gene regulatory networks. Expression of the gene *dlx3b* has been observed in epithelia and mesenchyme during dermal and perichondral bone development in zebrafish (Verreijdt et al., 2006 [\[link\]](#)). In addition to *dlx3b*, we observed bone development genes *scpp1* and *ifitm5* expressed in the ossifying dermal armor. Future studies could test our hypothesis that the dermal armor evolved through re-deployed osteoblast networks by examining osteoblast gene network expression over time.

Signatures of embryonic interactions within the novel pouch environment

Syngnathid embryos are reared within the brood pouch, a novel structure and environment composed of male-derived tissues (epithelium and placental-like tissues that include specialized cell types) that harbors a pouch microbiome (Stölting & Wilson, 2007 [\[link\]](#)). During pregnancy, the male brood pouch undergoes numerous changes including increased vascularization and altered expression of immune genes (Harada et al., 2022 [\[link\]](#); Ripley et al., 2010 [\[link\]](#); Roth et al., 2020 [\[link\]](#); Small et al., 2013 [\[link\]](#); Whittington et al., 2015). Researchers predict that these changes relate to nutrient and waste transfer and prevention of embryonic rejection and bacterial infection (Dudley et al., 2021 [\[link\]](#); Whittington & Friesen, 2020 [\[link\]](#)). However, there are few studies that examine whether and how embryos interact with the brood pouch environment (Kvarnemo et al., 2011 [\[link\]](#); Ripley & Foran, 2006 [\[link\]](#)). To consider whether embryos have specializations for life in this brood pouch environment, we asked about cell type-specific expression of nutrient acquisition and/or immune genes.

Pipefish embryos uptake paternally derived carbohydrates, proteins, and lipids (Kvarnemo et al., 2011 [\[link\]](#); Ripley & Foran, 2006 [\[link\]](#)). Our data suggest that this uptake could occur through the embryonic epidermis. Specifically, we noticed an enrichment of endocytosis and metabolism genes in epidermal cells. Epidermal absorption of maternally derived nutrients has been suggested in viviparous fishes (Tengfei et al., 2021 [\[link\]](#); Wourms, 1981 [\[link\]](#)). Interestingly, microvilli, a type of cellular projection, have been observed on the anal fin of developing seahorses (Wetzel & Wourms, 2004 [\[link\]](#)). In light of our findings in pipefish, possibly these seahorse microvilli could be functionally equivalent to those on the small intestine, maximizing nutrient absorption from the environment.

During pregnancy, the male brood pouch increases expression of C-type lectin genes (Roth et al., 2020 [\[link\]](#); Small et al., 2013 [\[link\]](#); Whittington et al., 2015). These genes are transmembrane or secreted receptors that sense self or non-self and are primarily studied for their role in innate and adaptive immunity (Brown et al., 2018 [\[link\]](#)). We identified 14 C-type lectin genes expressed in the embryonic epidermis. Our work suggests that lectin genes are produced by both the father and the embryos, but their function is still unclear. Syngnathid research has primarily suggested that lectin genes are produced to prevent bacterial infection (Melamed et al., 2005 [\[link\]](#)), though they could be important for male-embryo recognition.

Overall, our findings suggest that pipefish embryos have evolved to be specialized for development within the brood pouch by expressing genes related to nutrient acquisition and immunity. Future studies could provide insights into when nutrient acquisition and lectin genes are expressed in development, their functional role, and how their expression varies across syngnathid lineages that have exposed versus enclosed embryos, for example, to examine how embryonic development has been impacted by the brood pouch.

Conclusions

Our study represents the first scRNAseq developmental atlas in syngnathids, and one of the first non-model developmental scRNAseq atlases, providing a major step forward for evo-devo research. We used our atlas to begin addressing questions on the evolution and development of syngnathid innovations including their unique craniofacial structure, loss of teeth, dermal armor, and development within the male brood pouch. By combining scRNAseq analysis with spatial expression data from *in situ* hybridization, we made important discoveries in cell type identity and distribution as well as spatial expression of marker and signaling genes. We found that syngnathids express conserved signaling pathways during craniofacial development, suggesting that alterations within these pathways may be important for the evolution of their craniofacial skeletons. We did not find evidence for tooth primordia within syngnathids and propose that genetic changes early in tooth development could have led to their loss of teeth. We propose that the re-deployment of bone gene networks, but probably not tooth gene networks, could play a role in the dermal armor development. Finally, we observed an enrichment of endocytosis genes and many C-type lectin genes in epidermal cells, which suggests ways these cells might interact with the brood pouch environment. Our atlas advances our understanding of syngnathid development and evolution and provides resources for developmental genetic analysis in nascent evo-devo model species.

Methods

Single Cell RNA Sequencing Libraries Preparation

We created scRNAseq atlases from embryos of wild caught Gulf pipefish (*Syngnathus scovelli*, acquired from collaborator Emily Rose using collection permit SAL-21-0182-E), and all work was performed under an approved IACUC protocol. We harvested 20 embryos per pouch from two wild caught male pipefish. The embryos were at a stage before the tubular face was fully elongated, and while the head skeleton is cartilaginous with minimal signs of mineralization of superficial intramembranous bones. This corresponds to a stage termed “frontal jaws” in a recent description of pipefish development (Sommer et al., 2012 [DOI](#)).

We dissociated the embryos using 460ul of .25% trypsin in water and 40ul 100 mg/mL collagenase I (Sigma C0130-200mg) for 16 minutes. We filtered cells using a 40uM cell strainer (Thomas Scientific #1181X52). We quantified cell concentrations using the TC20 Automated Cell Counter (Biorad) then diluted the samples to 800 cells/ul in .04% BSA in PBS. The University of Oregon Genomics and Cell Characterization Core (GC3F; <https://gc3f.uoregon.edu> [DOI](#)) prepared single cell libraries for each sample using 10X Genomics Single Cell 3' Genome Expression mRNAseq kit with NextGEM v3.1 chemistry. We sequenced these libraries on an S4 lane on the NovaSeq 6000 at the GC3F. To improve the 3' UTR genome annotations, we also prepared scISOrSeq libraries from the first embryonic sample and from dissociated pouch cells from pregnant and nonpregnant males. These libraries were produced in accordance with (Healey et al., 2022 [DOI](#)). Embryonic, pregnant pouch and non-pregnant pouch libraries were sequenced separately on PacBio Sequel II - SMRT Cells 8M.

To turn the scISOrSeq reads into gene models, we followed the pipeline from Healey et al. 2022 [DOI](#). We ran the script (scISOr_Seq_processing.py from https://github.com/hopehealey/scISOrSeq_processing [DOI](#)) to remove barcodes, identify cell barcodes, and demultiplex with the single cell flag and appropriate barcodes (5' CCCATGTACTCTGCGTTGATACCACTGCT and 3' CTACACGACGCTCTCCGATCT). We aligned the reads to the 2022 Gulf pipefish genome (GenBank: GCA_024217435.2) using minimap v2.9 (Li, 2018 [DOI](#)). We filtered the reads using cDNA cupcake to remove duplicate transcripts (Tseng, 2021 [DOI](#)). We used SQANTI3 to identify gene models and filter them (Tardaguila et al., 2018 [DOI](#)). We merged the SQANTI3 annotations with the Gulf pipefish

genome (NCBI GenBank: GCF_024217435.2) using TAMA merge (Kuo et al., 2017 [DOI](#)). Since the Gulf pipefish genome does not contain mitochondrial genes, we appended the annotation and fasta files with the Gulf pipefish mitochondrial genome (NCBI RefSeq: NC_065499.1).

Single Cell Atlas Construction

We ran Cell Ranger (10X Genomics 3.0.2) using our scRNAseq reads, the Gulf pipefish genome assembly with mitochondrial genome, and the modified gene annotations. Cell Ranger estimated 20,733 cells for sample one, 23,682 genes expressed, and 21,039 mean reads per cell. For sample two, Cell Ranger predicted 17,626 cells, 23,740 genes expressed, and 29,804 mean reads per cell. We analyzed Cell Ranger's output using Seurat (v4.1.0) on R (v4.0.2; Butler et al., 2018 [DOI](#); Hafemeister & Satija, 2019 [DOI](#)).

To remove extraneous RNA counts from the dataset, we used SoupX (v1.5.2; Young & Behjati, 2020 [DOI](#)). We identified doublet scores for our dataset using scrublet (v0.2.3). The doublet removal step reduced the first sample by 114 cells (from 20,733 cells to 20,619 cells) and second sample by 167 cells (from 17,626 cells to 17,459 cells). We finally removed cells with less than 500 features, greater than 9000 features, greater than 1E5 RNA counts, with a scrublet score greater than the detected threshold (.76 for sample 1 and .21 for sample 2), or greater than 10% mitochondrial reads. The second filtering step removed 727 cells from sample one (20,619 cells to 19,892 cells) and 1,566 cells from sample two (from 17,459 cells to 15,893 cells).

We normalized the datasets with SCTransform (v0.3.3). We used Seurat's integration tools, SelectIntegrationFeatures using 3,000 feature genes, FindIntegrationAnchors using SCT normalization, and IntegrateData using SCT normalization, to integrate the two datasets. After integration, our combined atlas had 35,785 cells (Supplementary File 1,2). We then used the integrated dataset to complete PCA analysis. We tested using a variety of principle components for further analysis and chose 30 PCs for our analysis based on the clear delineation of major cell types. We next clustered the cells using 30 PCs and plotted the data on a UMAP with Seurat.

Single Cell Atlas Cluster Identification

To identify cluster identities, we used the RNA assay of the scRNAseq data to find cluster markers with Seurat's FindAllMarkers command with the parameters `only.pos = TRUE` and `logfc.threshold = 0.25`, requiring markers to be upregulated in the cluster and have a log fold change of at least .25. We found a second set of cluster markers through our custom function which searched through all genes and identified genes uniquely expressed in greater than 60% of cells in the cluster and in less than 10% of cells in every other cluster using Seurat's DotPlots. We searched for our identified markers in available zebrafish datasets (Fabian et al., 2022 [DOI](#); Farnsworth et al., 2020 [DOI](#); Lange et al., 2023 [DOI](#)), ZFIN (Howe et al., 2013 [DOI](#)), NCBI, medlineplus.gov, and genecards to give the clusters initial annotations. For each cluster, we examined multiple genes using DotPlots and FeaturePlots to propose the cluster identity.

Next, we identified one gene for each cluster which marked the cluster best (expressed in the most cells in the focal cluster and expressed in as few of the other clusters as possible) through consulting the two marker gene lists and examining markers with Dot Plots (Supplementary File 3, 4, 5). Using these markers, we completed a set of *in situ* hybridizations to hone our cluster annotations. Due to challenges culturing Gulf pipefish, we used both embryos and larvae from *Syngnathus leptorhynchus*, a pipefish from the same genus that lives in Oregon coastal habitats and allowed for easy collection, and from cultured Gulf pipefish for the cluster annotation *in situ* experiments. We caught pregnant male *Syngnathus leptorhynchus* using a beach seine near Coos Bay, Oregon under Oregon Department of Fisheries and Wildlife permit number 26987.

Syngnathus scovelli used for *in situ* experiments were purchased from Alyssa's Seahorse Saavy and Gulf Specimens Marine Lab then reared in our facility at 25 degrees C water and 25-28 PPT Salinity. We designed probes using NCBI Primer Blast with the Gulf pipefish genome assembly and produced these probes using Gulf pipefish embryonic cDNA pools. To create the probes, we completed two rounds of PCR. The first round used a gene specific forward primer with a reverse gene specific primer that had 10 nucleotides of T7 promotor sequence attached. PCR products were cleaned with Zymo clean and concentrator DNA kit and eluted in 15 ul of elution buffer. Round two of PCR used the same gene specific forward primer with a modified T7 promoter sequence (TGGACTAATACGACTCACTATAGGG) as the reverse primer, and finally the product was cleaned again with Zymo clean and concentrator DNA kit and eluted in 15 ul of elution buffer.

Round one PCR conditions are as followed: 95 degrees Celsius for 3:00 minutes, 40 cycles of denaturation (95 degrees for 30 seconds), annealing (30 seconds, annealing temperature varied by probe), and extension (72 degrees for 1:00 minute), and a final extension step of 3:00 minutes. Round two PCR conditions are as followed: 98 degrees Celsius for 30 seconds, 30 degrees for 10 seconds, 72 degrees for 50 seconds, then 35 rounds of denaturation (98 degrees for 10 seconds), annealing (50 degrees for 10 seconds), and extension (72 degrees for 50 seconds), and a final extension step of 10 minutes at 72 degrees. For round two PCRs with multiple bands (specifically, *ifitm5*), the band of the expected size was excised with a razorblade and DNA was extracted with Zymoclean gel DNA recovery kit.

The round two PCR product was sanger sequenced to confirm identity. All the marker genes primers and successful PCR conditions are in Supplementary File 6. Alignments of the probes to the unpublished *Syngnathus leptorhynchus* genome assembly are included in Supplementary File 7. The probes were transcribed with T7 polymerase for 2-6 hours then cleaned with Zymo RNA clean and concentrator and eluted into 30 ul of water. For the *in situ* hybridizations, we selected embryos and newly spawned larvae close to the developmental stage used in the atlas. We completed *in situ* hybridizations in keeping with (Thisse & Thisse, 2007 [link](#)), leaving the embryos in stain until background was observed. For spawned larvae, we completed a bleaching step (1% H2O2 and .5% KOH for 8 minutes) prior to the proteinase K digest. After imaging, we used the levels tool in Adobe Photoshop (v23.4.2) to white balance the pictures.

Single Cell KEGG Analysis

To identify pathways upregulated in cell clusters, we completed a KEGG analysis. We downloaded Gulf pipefish KEGG pathways (<https://www.kegg.jp> [link](#)). For the KEGG analysis, we used the marker genes identified from our FindAllMarkers list as the input genes. We converted these gene ids using keggConv to KEGG ids. We used a Wilcoxon enrichment test to ask whether cluster marker genes were enriched for each KEGG pathway.

Differentiation State Analysis

To assess whether proposed primordial cell clusters were composed of undifferentiated cells relative to other clusters from their lineages, we completed a differentiation analysis using CytoTRACE (v0.3.3, Gulati et al., 2020 [link](#)). Clusters from similar lineages (neural: 0, 3, 7, 8, 12, 22, 25, 33, and 35; muscle: 2, 10, 17, and 37; connective: 4, 5, 6, 9, 15, 16, 18, 24, 27, 28, and 29) were isolated using subset. Cell counts were gathered using as.Matrix(GetAssayData), then CytoTRACE was run on these counts. Data was plotted on a VlnPlot to show the variability.

Single Cell Atlas Network Analysis

To identify genetic networks present in our atlas, we completed a weighted gene network correlation analysis using WGCNA (v1.72-1, Langfelder & Horvath, 2008 [link](#)). We selected 3,000 variable features from the integrated assay of the single cell dataset for the WGCNA. We created an adjacency matrix from the data using bicor with a maxPOutliers of .05. To decide on a beta value

or the soft threshold power, we created an adjacency matrix plot using `pickSoftThreshold` and picked the threshold where the scale free topology model fit leveled off. We selected the value of two. Then, we raised the adjacency matrix to the power of two.

We calculated the dissimilarity matrix by calculating TOM similarity of the adjacency matrix and subtracting it from one. We then created a gene tree through `hclust` of the dissimilarity matrix with the average method. From the gene tree, we created modules using `cutreeDynamic` with `deepSplit` of two and the `minClusterSize` of 15 (setting the smallest cluster size to 15). We calculated module eigengenes using `moduleEigengenes` of the adjacency matrix then calculated the dissimilarity of the module eigengenes with `cor` of the module eigengenes subtracted from one and clustered the module eigengenes with `hclust`. We then chose a dissimilarity of 35% as the cut off for merging modules; however, no modules had dissimilarity scores with each other below 40%. Since the module eigengenes are calculated for each cell in the dataset, we used these values to calculate the t-statistic for each cell cluster (considered each sample) in the module (all of the module eigengene scores were used as the population).

We next tested the hypothesis that certain cell clusters were strongly associated with specific gene modules through a two-way permutation test (1000 permutations) and corrected p-values to control the False Discovery Rate (FDR). Additionally, we tested whether specific cell clusters drove underlying gene network structure by measuring network connectivity. The network connectivity was first calculated using the entire dataset, then each cell cluster was progressively dropped from the dataset and the connectivity was remeasured. This resulted in change in connectivity scores for each module-cluster pair. To assess whether these changes were significant, we completed 1000 permutations whereby cells were randomly dropped (the number of cells dropped was equal to the cell cluster size of the focal cluster), connectivity measured, and the change in connectivity recorded. The p-value is the number of instances where the change in connectivity is greater in the permutations than in the focal cell cluster run. P-values were corrected to control the FDR.

Using the module gene lists, we identified the number of genes from each module that were found in each KEGG pathway. Since the KEGG modules do not have p-values associated with the genes, we could not complete a Wilcoxon enrichment test. We instead removed any pathways where there were less than three genes present for any pathway and note that there is no statistical test run on these KEGG results. We visualized the networks using Cytoscape (v3.10.0).

In situ Hybridization

For genes chosen for further follow up analysis, we completed *in situ* hybridizations in Gulf pipefish (*Syngnathus scovelli*) or bay pipefish (*Syngnathus leptorhynchus*). Primer sequences were designed using NCBI Primer Blast with the Gulf Pipefish genome and synthesized using Gulf pipefish embryonic cDNA. The *sfrp1a* probe was synthesized using the PCR based probe preparation protocol described in the Single Cell Identification section. All other probes (*bmp4*, *dlx2a*, *dlx3b*, *fgf22*, *lhx6a*, *pitx2*, and *scpp1*) were prepared using TOPO cloning. Probe primer sequences as well as the species used for the *in situ* experiments are described in Supplementary file 6. Pregnant male bay pipefish were caught as described above. For Gulf pipefish embryos, we allowed the fish to mate in our facility then harvested embryos once they reached the appropriate stages. Fish were reared in 25 degrees C water with 25-28 ppt salinity. Embryos and larvae were fixed in 4% PFA, dehydrated through a series of PBT/MeOH washes, and stored in MeOH at -20C. 5-12 embryos were used for each probe. We completed *in situ* hybridizations in keeping with (Thisse & Thisse, 2007 [DOI](#)), leaving the embryos in stain until background was observed. After imaging, we used the levels tool in Adobe Photoshop (v23.4.2) to white balance the pictures.

Bone and Cartilage Staining

We used alcian and alizarin stains to mark cartilage and bones. Specifically, we assayed cartilage and bone development in siblings of the scRNAseq samples. These embryos were fixed in 4% PFA then stored at -20C in MeOH. We followed the protocol from Walker and Kimmel (2007) with minor alterations. We stored samples in 50% glycerol/0.1% KOH at 4C and imaged them in 100% glycerol. After imaging, we white balanced the photographs using Photoshop (v23.4.2) levels tool.

Gene Cluster Analysis

To examine close syngnathid outgroups, we downloaded the 2023 mandarin dragonet genome (GenBank assembly accession: GCA_027744825.1) and the 2024 cornetfish genome assembly (GenBank assembly accession: GCA_037954325.1) from NCBI. Since these genomes were unannotated, we manually identified *scpp* genes. We searched for *scpp* genes using BLASTN with medaka and additional fish sequences as the query. We also searched for these genes with mVISTA plots across conserved gene synteny regions (LAGAN alignment using translated anchoring) with medaka as the focal species. To identify *scpp* genes in additional species, we gathered cluster information from NCBI and ensembl. Additionally, we used mVISTA plots to further search for unannotated *scpp* genes.

Acknowledgements

We are indebted to Emily Rose for her collection of the Gulf pipefish specimens for the single cell libraries. We are grateful for everyone in the Cresko Lab for their ideas and participation in early morning pipefish collection trips. We are additionally thankful for Emily Beck for her assistance collecting pipefish. In particular, we are immensely appreciative of Mark Currey and Tiffany Thornton's valiant efforts to culture Gulf pipefish. Additionally, we are grateful for Balan Ramesh and Adam Jones for collaborating with the Cresko Lab to produce new Syngnathiformes genomes. We thank Tina Arredondo from the UO GC3F for her preparation of our single cell RNA sequencing libraries. This work was funded by the National Science Foundation Grant OPP-2015301 (to WAC, SB, and CMS), University of Oregon Research Excellence funds (WAC) and National Institute of Health Fellowship F31DE032559-02 (to HMM). Additionally, HMM was supported by the Genetics Training Program (NIH T32GM007413). Finally, VG received funding and mentorship from the Hui Undergraduate Research Scholars and MW was supported by the Knight Campus Undergraduate Scholars program at the University of Oregon.

Data Availability statement

All raw sequencing data associated with this study are published via NCBI (PRJNA1168967). The integrated single cell RNA sequencing atlas is also available through NCBI (GSE278814). The fasta file and updated Gulf pipefish annotation are stored on Dryad (10.5061/dryad.f7m0cfz60). Code used for the analysis is available on github (https://github.com/hopehealey/ssc_emb_atlas_scripts .

Supplementary figure captions

Figure S1: Two Gulf pipefish samples have similar contributions to cell clusters. This UMAP features cells plotted by sample (sample one in orange and sample two in cyan) in the first two UMAP dimensions.

Figure S2: QC metrics for both Gulf pipefish samples show similar sample quality after initial filtering. Sample quality was assessed using the number of genes identified (nFeatureRNA), the total RNA counts (nCountRNA), the percentage of mitochondrial reads (percentmt), and the likelihood of being a doublet (doubvalues). Gulf pipefish sample data post-filtering is displayed for each of these metrics.

Figure S3: Zebrafish muscle cell marker genes expression patterns correspond with annotated muscle cells in Gulf pipefish atlas. Feature plots of zebrafish muscle cell marker genes identified from the Daniocell scRNAseq database demonstrate Gulf pipefish expression. Annotated muscle cell clusters in pipefish are distinguished by a dashed outline.

Figure S4: Zebrafish muscle satellite cell marker genes expression patterns correspond with annotated muscle primordia cells in Gulf pipefish atlas. Feature plots of zebrafish muscle satellite cell marker genes identified from the Daniocell scRNAseq database demonstrate Gulf pipefish expression. Annotated muscle primordia cell cluster in pipefish is distinguished by a dashed outline.

Figure S5: Zebrafish cardiac muscle cell marker genes expression patterns correspond with annotated cardiac muscle cells in Gulf pipefish atlas. Feature plots of zebrafish cardiac muscle cell marker genes identified from the Daniocell scRNAseq database demonstrate Gulf pipefish expression. Annotated cardiac muscle cell cluster in pipefish is distinguished by a dashed outline.

Figure S6: Zebrafish retinal cell marker genes expression patterns correspond with annotated retinal cells in Gulf pipefish atlas. Feature plots of zebrafish retinal cell marker genes identified from the Daniocell scRNAseq database demonstrate Gulf pipefish expression. Annotated retinal cell cluster in pipefish is distinguished by a dashed outline.

Figure S7: Zebrafish neuronal cell marker genes expression patterns correspond with annotated neuronal cells in Gulf pipefish atlas. Feature plots of zebrafish neuronal cell marker genes identified from the Daniocell scRNAseq database demonstrate Gulf pipefish expression. Annotated neuronal cell clusters in pipefish are distinguished by a dashed outline.

Figure S8: Zebrafish chondroblast cell marker genes expression patterns correspond with annotated chondroblast cells in Gulf pipefish atlas. Feature plots of zebrafish chondroblast cell marker genes identified from the Daniocell scRNAseq database demonstrate Gulf pipefish expression. Annotated chondroblast cell cluster in pipefish is distinguished by a dashed outline.

Figure S9: Zebrafish osteoblast cell marker genes expression patterns correspond with annotated osteoblast cells in Gulf pipefish atlas. Feature plots of zebrafish osteoblast cell marker genes identified from the Daniocell scRNAseq database demonstrate Gulf pipefish expression. Annotated osteoblast cell cluster in pipefish is distinguished by a dashed outline.

Figure S10: Zebrafish notochord cell marker genes expression patterns correspond with annotated notochord cells in Gulf pipefish atlas. Feature plots of zebrafish notochord cell marker genes identified from the Daniocell scRNAseq database demonstrate Gulf pipefish expression. Annotated notochord cell cluster in pipefish is distinguished by a dashed outline.

Figure S11: Zebrafish fin cell marker genes expression patterns correspond with annotated fin cells in Gulf pipefish atlas. Feature plots of zebrafish fin cell marker genes identified from the Daniocell scRNAseq database demonstrate Gulf pipefish expression. Annotated fin cell cluster in pipefish is distinguished by a dashed outline.

Figure S12: Zebrafish mesenchymal cell marker genes expression patterns correspond with annotated mesenchymal cells in Gulf pipefish atlas. Feature plots of zebrafish mesenchymal cell marker genes identified from the Danio cell scRNAseq database demonstrate Gulf pipefish expression. Annotated mesenchymal cell clusters in pipefish are distinguished by a dashed outline.

Figure S13: Zebrafish tenocyte cell marker genes expression patterns correspond with annotated tenocyte cells in Gulf pipefish atlas. Feature plots of zebrafish tenocyte cell marker genes identified from the Danio cell scRNAseq database demonstrate Gulf pipefish expression. Annotated tenocyte and ligament cell cluster in pipefish is distinguished by a dashed outline.

Figure S14: Zebrafish erythrocyte cell marker genes expression patterns correspond with annotated erythrocyte cells in Gulf pipefish atlas. Feature plots of zebrafish erythrocyte cell marker genes identified from the Danio cell scRNAseq database demonstrate Gulf pipefish expression. Annotated erythrocyte cell clusters in pipefish are distinguished by a dashed outline.

Figure S15: Zebrafish endothelial cell marker genes expression patterns correspond with annotated endothelial cells in Gulf pipefish atlas. Feature plots of zebrafish endothelial cell marker genes identified from the Danio cell scRNAseq database demonstrate Gulf pipefish expression. Annotated endothelial cell cluster in pipefish is distinguished by a dashed outline.

Figure S16: Zebrafish gut and liver cell marker genes expression patterns correspond with annotated gut and liver cells in Gulf pipefish atlas. Feature plots of zebrafish gut and liver cell marker genes identified from the Danio cell scRNAseq database demonstrate Gulf pipefish expression. Annotated gut and liver cell cluster in pipefish is distinguished by a dashed outline.

Figure S17: Zebrafish epidermal cell marker genes expression patterns correspond with annotated epidermal cells in Gulf pipefish atlas. Feature plots of zebrafish epidermal cell marker genes identified from the Danio cell scRNAseq database demonstrate Gulf pipefish expression. Annotated epidermal cell cluster in pipefish is distinguished by a dashed outline.

Figure S18: Zebrafish glial cell marker genes expression patterns correspond with annotated glial cells in Gulf pipefish atlas. Feature plots of zebrafish glial cell marker genes identified from the Danio cell scRNAseq database demonstrate Gulf pipefish expression. Annotated glial cell cluster in pipefish is distinguished by a dashed outline.

Figure S19: Zebrafish immune cell marker genes expression patterns correspond with annotated immune cells in Gulf pipefish atlas. Feature plots of zebrafish immune cell marker genes identified from the Danio cell scRNAseq database demonstrate Gulf pipefish expression. Annotated immune cell clusters in pipefish are distinguished by a dashed outline.

Figure S20: Zebrafish pigment cell marker genes expression patterns correspond with annotated pigment cells in Gulf pipefish atlas. Feature plots of zebrafish pigment cell marker genes identified from the Danio cell scRNAseq database demonstrate Gulf pipefish expression. Annotated pigment cell clusters in pipefish are distinguished by a dashed outline.

Figure S21: Gulf pipefish cluster 0 markers. Feature plots of top markers for cluster 0 identified from the marker gene lists.

Figure S22: Gulf pipefish cluster 1 markers. Feature plots of top markers for cluster 1 identified from the marker gene lists.

Figure S23: Gulf pipefish cluster 2 markers. Feature plots of top markers for cluster 2 identified from the marker gene lists.

Figure S24: Gulf pipefish cluster 3 markers. Feature plots of top markers for cluster 3 identified from the marker gene lists.

Figure S25: Gulf pipefish cluster 4 markers. Feature plots of top markers for cluster 4 identified from the marker gene lists.

Figure S26: Gulf pipefish cluster 5 markers. Feature plots of top markers for cluster 5 identified from the marker gene lists.

Figure S27: Gulf pipefish cluster 6 markers. Feature plots of top markers for cluster 6 identified from the marker gene lists.

Figure S28: Gulf pipefish cluster 7 markers. Feature plots of top markers for cluster 7 identified from the marker gene lists.

Figure S29: Gulf pipefish cluster 8 markers. Feature plots of top markers for cluster 8 identified from the marker gene lists.

Figure S30: Gulf pipefish cluster 9 markers. Feature plots of top markers for cluster 9 identified from the marker gene lists.

Figure S31: Gulf pipefish cluster 10 markers. Feature plots of top markers for cluster 10 identified from the marker gene lists.

Figure S32: Gulf pipefish cluster 12 markers. Feature plots of top markers for cluster 12 identified from the marker gene lists.

Figure S33: Gulf pipefish cluster 13 markers. Feature plots of top markers for cluster 13 identified from the marker gene lists.

Figure S34: Gulf pipefish cluster 14 markers. Feature plots of top markers for cluster 14 identified from the marker gene lists.

Figure S35: Gulf pipefish cluster 15 markers. Feature plots of top markers for cluster 15 identified from the marker gene lists.

Figure S36: Gulf pipefish cluster 16 markers. Feature plots of top markers for cluster 16 identified from the marker gene lists.

Figure S37: Gulf pipefish cluster 17 markers. Feature plots of top markers for cluster 17 identified from the marker gene lists.

Figure S38: Gulf pipefish cluster 18 markers. Feature plots of top markers for cluster 18 identified from the marker gene lists.

Figure S39: Gulf pipefish cluster 19 markers. Feature plots of top markers for cluster 19 identified from the marker gene lists.

Figure S40: Gulf pipefish cluster 20 markers. Feature plots of top markers for cluster 20 identified from the marker gene lists.

Figure S41: Gulf pipefish cluster 21 markers. Feature plots of top markers for cluster 21 identified from the marker gene lists.

Figure S42: Gulf pipefish cluster 22 markers. Feature plots of top markers for cluster 22 identified from the marker gene lists.

Figure S43: Gulf pipefish cluster 23 markers. Feature plots of top markers for cluster 23 identified from the marker gene lists.

Figure S44: Gulf pipefish cluster 24 markers. Feature plots of top markers for cluster 24 identified from the marker gene lists.

Figure S45: Gulf pipefish cluster 25 markers. Feature plots of top markers for cluster 25 identified from the marker gene lists.

Figure S46: Gulf pipefish cluster 26 markers. Feature plots of top markers for cluster 26 identified from the marker gene lists.

Figure S47: Gulf pipefish cluster 27 markers. Feature plots of top markers for cluster 27 identified from the marker gene lists.

Figure S48: Gulf pipefish cluster 28 markers. Feature plots of top markers for cluster 28 identified from the marker gene lists.

Figure S49: Gulf pipefish cluster 29 markers. Feature plots of top markers for cluster 29 identified from the marker gene lists.

Figure S50: Gulf pipefish cluster 30 markers. Feature plots of top markers for cluster 30 identified from the marker gene lists.

Figure S51: Gulf pipefish cluster 31 markers. Feature plots of top markers for cluster 31 identified from the marker gene lists.

Figure S52: Gulf pipefish cluster 32 markers. Feature plots of top markers for cluster 32 identified from the marker gene lists.

Figure S53: Gulf pipefish cluster 33 markers. Feature plots of top markers for cluster 33 identified from the marker gene lists.

Figure S54: Gulf pipefish cluster 34 markers. Feature plots of top markers for cluster 34 identified from the marker gene lists.

Figure S55: Gulf pipefish cluster 35 markers. Feature plots of top markers for cluster 35 identified from the marker gene lists.

Figure S56: Gulf pipefish cluster 36 markers. Feature plots of top markers for cluster 36 identified from the marker gene lists.

Figure S57: Gulf pipefish cluster 37 markers. Feature plots of top markers for cluster 37 identified from the marker gene lists.

Figure S58: Marker genes define cell identity and are a resource for cell cluster exploration. This dotplot shows the marker genes on the x-axis, and the cell cluster on the y-axis. The size of each dot indicates the percentage of cells that express the gene in each cluster. The darkness of each dot represents the average expression of the gene.

Figure S59: *slc25a4*, a marker for cluster 2, is expressed in muscle cells in bay pipefish. Atlas expression is limited to cluster 2 (A). *In situ* hybridization experiments revealed muscle staining of *slc25a4* in bay pipefish (B-D). Wild caught bay pipefish at the onset of craniofacial elongation were used for the experiment. Embryos were mounted laterally to highlight expression domains.

Figure S60: *scg2a*, a marker for cluster 3, is expressed in the brain of bay pipefish. Atlas expression is limited to cluster 3 (A). *In situ* hybridization experiments revealed brain staining of *scg2a* in bay pipefish (B-D). Wild caught bay pipefish at the onset of craniofacial elongation were used for the experiment. Embryos were mounted laterally to highlight expression domains.

Figure S61: *prdm16*, a marker for cluster 5, is expressed in the head and fins of bay pipefish. Strong expression of *prdm16* in cluster 5, osteochondrogenic mesenchyme (A). *In situ* hybridization experiments revealed mesenchymal staining of *prdm16* in the fins and face of bay pipefish (B-D). Wild caught bay pipefish at the onset of craniofacial elongation were used for the experiment. Embryos were mounted laterally (B, D, E) and with a dorsal view (C) to highlight expression domains.

Figure S62: *elnb*, a marker for cluster 6, is expressed in the head of bay pipefish. Strong expression of *elnb* in cluster 6, osteochondrogenic mesenchyme (A). *In situ* hybridization experiments revealed mesenchymal staining of *elnb* in the face of bay pipefish (B-D). Wild caught bay pipefish at the onset of craniofacial elongation were used for the experiment. Embryos were mounted dorsally to showcase the craniofacial staining (B) and laterally (C-E) to show the lack of staining in the ceratohyal, dorsal fin, and caudal fin.

Figure S63: *fndc9*, a marker for cluster 7, is expressed in the brain of Gulf pipefish. Expression of *fndc9* is limited to cluster 7, brain cells, in the atlas (A). *In situ* hybridization experiments revealed brain staining of *fndc9* in Gulf pipefish (B-C). One day post spawn Gulf pipefish larvae were used for this experiment. Embryos were mounted to highlight two views of the brain staining, one dorsal (B) and one lateral (C).

Figure S64: *insm1b*, a marker for cluster 8, is expressed in the brain of Gulf pipefish. Strong expression of *insm1b* is observed in cluster 8, brain progenitor cells, in the atlas (A). *In situ* hybridization experiments revealed brain staining of *insm1b* in Gulf pipefish (B). Five day post spawn Gulf pipefish larvae were used for this experiment. Embryos were mounted laterally to highlight the brain staining.

Figure S65: *tnmd*, a marker for cluster 9, is expressed in tendons and ligaments of bay pipefish. Strong expression of *tnmd* is observed in cluster 9, tenocyte and ligament cells, in the atlas (A). *In situ* hybridization experiments revealed tendon and ligament staining of *tnmd* in bay pipefish (B-E). Wild caught embryos from bay pipefish at the onset of craniofacial elongation were used for the experiment. Embryos were mounted laterally (B,E) and ventrally (C,D) to showcase tendon and ligament staining.

Figure S66: *myf5*, a marker for cluster 10, is expressed in the muscle of bay pipefish. Strong expression of *myf5* is observed in cluster 10, muscle progenitor cells, in the atlas (A). *In situ* hybridization experiments revealed muscle staining of *myf5* in bay pipefish (B-D). Wild caught embryos from bay pipefish at the onset of craniofacial elongation were used for the experiment. Embryos were mounted laterally to highlight muscle staining.

Figure S67: *hpbdb*, a marker for cluster 13, is expressed in somites of bay pipefish. Strong expression of *hpbdb* is observed in cluster 13, fibroblasts, in the atlas (A). *In situ* hybridization experiments revealed somite staining of *hpbdb* in bay pipefish (B). Wild caught embryos from bay pipefish at the mid-somitogenesis stage were used for the experiment. Embryos were mounted laterally to highlight muscle staining.

Figure S68: *ifitm5*, a marker for cluster 18, is expressed in bay pipefish osteoblasts. Specific expression of *ifitm5* is observed in cluster 18, osteoblasts, in the atlas (A). *In situ* hybridization experiments revealed bone staining of *ifitm5* in bay pipefish (B-E). Wild caught embryos from bay pipefish at the onset of craniofacial elongation were used for the experiment. Embryos were mounted laterally to highlight muscle staining.

Figure S69: *plaat4*, a marker for cluster 23, is weakly expressed in bay pipefish epidermal cells. Specific expression of *plaat4* is observed in cluster 23, epidermal cells, in the atlas (A). *In situ* hybridization experiments revealed epidermal staining of *plaat4* in bay pipefish (B-D). Wild caught embryos from bay pipefish at the onset of craniofacial elongation were used for the experiment. Embryos were mounted ventrally (B) and laterally (C,D) to highlight epidermal staining.

Figure S70: *hyal6*, a marker for cluster 33, is strongly expressed in bay pipefish hair cells. Specific expression of *hyal6* is observed in cluster 33, hair cells, in the atlas (A). *In situ* hybridization experiments revealed staining of *hyal6* in bay pipefish nares and ears (B-C). Wild caught embryos from bay pipefish from one day post spawn larvae were used for the experiment. Embryos were mounted dorsally (B) and laterally (C) to highlight staining in nares and ears respectively.

Figure S71: *soat2*, a marker for cluster 36, is strongly expressed in bay pipefish muscle cells. Expression of *soat2* is observed in cluster 36, muscle cells, in the atlas (A). *In situ* hybridization experiments revealed staining of *soat2* in bay pipefish muscle and fins (B-E). Wild caught embryos from bay pipefish at the onset of craniofacial elongation were used for the experiment. Embryos were mounted laterally to highlight staining in muscle and fin.

Figure S72: KEGG analysis reveals pathways significantly enriched in cell clusters. KEGG pathways are listed on the x-axis. For each cluster, the number of genes that belong to the specific KEGG pathway is represented by the height of each bar. Cell clusters are labeled by both color and their designated number. For clarity, only cell clusters with more than three annotated genes in each pathway are shown and certain pathways involving infection and metabolism are removed.

Figure S73: Clusters 8,10, and 16 are composed of largely undifferentiated cells. Differentiation state of cell clusters was measured using CytoTRACE, where an output of 1 indicates an undifferentiated cell and a value of 0 suggests a cell is fully differentiated. Cells are plotted on a violin plot with individual cell predictions shown as dots. Panels show results from muscle (A), neural (B), and connective (C) clusters.

Figure S74: Cell clusters drive gene module connectivity. The heatmap shows the change in gene network connectivity when individual cell clusters are removed. Cell clusters are in the columns and gene modules are in rows. The coloration scale is based on the change in connectivity when a cell cluster is removed, with yellow indicating a large change in connectivity. The modules and clusters are clustered using the Pearson distance method. Asterisks indicate cluster-module pairs that are significant using a one way permutation test and FDR correction. The size of the gene modules are shown to the right.

Figure S75: Gene module expression comparison with the Pigment Cluster reveals potential elusive cell types within the cluster. Genes within the network are combined into one feature then plotted on a feature plot. In this case, UMAP dimensions one and two are the x and y-axis accordingly, dots represent each pigment cell, and the color of the dots represents the expression of the module. Cells that express gene modules are either colored red or green (when modules are plotted on separate plots). When plotted on the same plot, cells that express both gene modules are yellow. Panels A through J show each gene module comparison.

Figure S76: Conserved fin *prdm16* expression domains in stickleback and pipefish, but fishes have expression in different regions of the head. The gene *prdm16* expression was probed using *in situ* hybridization. Embryos used in this assay were similarly staged (9dpf in threespine stickleback and mid-craniofacial elongation in bay pipefish). Stickleback (A-C) expressed *prdm16* in the lower jaw, brain, gill arches, and fins. Bay pipefish also have fin staining (F), but craniofacial expression is limited to the mesenchyme above the ethmoid region.

Figure S77: Conserved expression domains of *fgf22* in pipefish and stickleback fish.

Expression of *fgf22* was probed using *in situ* hybridization. Embryos used in this assay were similarly staged (9dpf threespine stickleback and 10dpf Gulf pipefish). Staining in the jaws (A,E), brain (B,F), gill arches (C,G), and fins (D, H) suggests similar *fgf22* expression domains in both species.

Figure S78: Broad *bmp4* craniofacial expression in stickleback, but *bmp4* is restricted to ethmoid and ceratohyal regions in pipefish. Expression of *bmp4* was probed using *in situ* hybridization. Embryos used in this assay were similarly staged (9dpf threespine stickleback and mid-craniofacial elongation staged bay pipefish). Stickleback have *bmp4* staining in tooth germs (A), the jaws (A-B), and gill arches (C). However, *bmp4* expression is restricted in bay pipefish to mesenchyme above the ethmoid and ceratohyal.

Figure S79: *sfrp1a* has conserved fin expression in pipefish and stickleback, but craniofacial expression differs between the species. Expression of *sfrp1a* was probed using *in situ* hybridization. Embryos used in this assay were similarly staged (9dpf threespine stickleback and mid-craniofacial elongation staged bay pipefish). Both species had *sfrp1a* staining in the fins (D,H). However, stickleback have broad craniofacial expression (in lower jaws, A-B, and gill arches, C) while expression is limited to mesenchyme above the ethmoid (E), the jaw joint (F), and around the ceratohyal (G) in bay pipefish.

Figure S80: *scpp* gene cluster analysis identifies most *scpp* losses are not unique to syngnathids. A) denotes the Syngnathoidei lineage which contains syngnathids and blue spotted cornetfish (among other fish) to highlight that this clade contains various tooth alterations. Syngnathids lack teeth completely and cornetfish have small teeth limited to the lower jaw. B) denotes solely the syngnathid clade, which completely lacks teeth. The phylogenetic tree was time calibrated using Hughes et al. 2018 [\[1\]](#) and Stiller et al. 2022 [\[2\]](#).

Figure S81: 3dpf Zebrafish do not express lectin genes in epidermal cells. This dotplot shows the marker genes on the x-axis and y-axis, organized by cell type. The size of the dot represents the percentage of cells that express the gene in each cluster. The darkness of the dot represents the average expression of the genes.

Supplementary file captions

SupplementaryFile1_scRNAseq_cellranger_info.csv, a table containing quality metrics for the single cell libraries.

SupplementaryFile2_cell_cluster_identity_and_sizes.csv, a table containing the number of cells in each cell cluster and cluster identities.

SupplementaryFile3_seurat_findallmarker_output.csv, a file containing marker genes identified using Seurat FindAllMarkers for each cluster.

SupplementaryFile4_dotplotmethod_marker_genes.csv, a file containing marker genes identified using the DotPlot method.

SupplementaryFile5_select_marker_gene_detailed_information.xlsx, a file containing additional information on the marker gene identified for every cluster.

SupplementaryFile6_insitu_probe_sequences_and_prep_information.xlsx, a list of the *in situ* hybridization probes used in this study, the conditions used to prepare the probes, and the staging/sample information for the embryos.

SupplementaryFile7_insitu_probe_alignment.zip, a zipped folder containing alignments of *in situ* hybridization probes with the unpublished *Syngnathus leptorhynchus* genome.

SupplementaryFile8_WGCNA_module_number_to_color_translation_table.csv, genetic networks were initially labeled with colors, we converted these labels to numeric annotations for simplicity using this conversion table. The table also contains the number of genes in each network.

SupplementaryFile9_WGCNA_networks_with_gene_descriptions.xlsx, a file containing all the genetic networks, the genes inside of them, and additional information for the genetic networks highlighted in this paper.

SupplementaryFile10_WGCNA_mod_association_wClus_tstat.csv, this file contains the t-statistics derived for each module-cell cluster pair. The cell clusters are in the rows and the gene modules are in the columns.

SupplementaryFile11_WGCNA_associat_mods_w_clus_tstat_pvalues_corrected.csv, p-values for the t-statistics of the strength of association between gene modules and cell clusters. P-values are corrected for multiple testing hypotheses using *fdr*. The cell clusters are in the rows and the gene modules are in the columns.

SupplementaryFile12_WGCNA_connectivity_change.csv, the change in connectivity for gene modules when individual cell clusters are removed. The cell clusters are in the columns and the gene networks are in the rows.

SupplementaryFile13_Connectivity_Permutation_Pvalues_Corrected.csv, the p-values for the change in connectivity are found in this csv. The cell clusters are in the columns and the gene networks are in rows.

References

- Ahi E. P., Kapralova K. H., Pálsson A., Maier V. H., Gudbrandsson J., Snorrason S. S., Jónsson Z. O., Franzdóttir S. R (2014) **Transcriptional dynamics of a conserved gene expression network associated with craniofacial divergence in Arctic charr** *EvoDevo* **5**:1–19
- Ahi E. P., Lecaudey L. A., Ziegelbecker A., Steiner O., Glabonjat R., Goessler W., Hois V., Wagner C., Lass A., Sefc K. M (2020) **Comparative transcriptomics reveals candidate carotenoid color genes in an East African cichlid fish** *BMC Genomics* **21**:1–15
- Anderson A. P., Jones A. G (2019) **Choosy Gulf pipefish males ignore age but prefer active females with deeply keeled bodies** *Animal Behaviour* **155**:37–44
- Arendt D. *et al.* (2016) **The origin and evolution of cell types** *Nat Rev Genet* **17**:744–757
- Bergen D. J. M. *et al.* (2022) **Regenerating zebrafish scales express a subset of evolutionary conserved genes involved in human skeletal disease** *BMC Biology* **20**:1–25
- Bergert B. A., Wainwright P. C (1997) **Morphology and kinematics of prey capture in the syngnathid fishes *Hippocampus erectus* and *Syngnathus floridae*** *Marine Biology* **127**:563–570
- Berglund A., Rosenqvist G., Svensson I (1986) **Reversed sex roles and parental energy investment in zygotes of two pipefish (Syngnathidae) species** *Marine Ecology Progress Series* **29**:209–215
- Bjork B. C., Turbe-Doan A., Prysak M., Herron B. J., Beier D. R (2010) **Prdm16 is required for normal palatogenesis in mice** *Human Molecular Genetics* **19**:774–789
- Brown G. D., Willment J. A., Whitehead L (2018) **C-type lectins in immunity and homeostasis** *Nature Reviews Immunology* **18**:374–380
- Butler A., Hoffman P., Smibert P., Papalexi E., Satija R (2018) **Integrating single-cell transcriptomic data across different conditions, technologies, and species** *Nature Biotechnology* **36**:411–420
- Cäestro C., Yokoi H., Postlethwait J. H (2007) **Evolutionary developmental biology and genomics** *Nature Reviews Genetics* **8**:932–942
- Carcupino M (2002) **Functional significance of the male brood pouch in the reproductive strategies of pipefishes and seahorses: a morphological and ultrastructural comparative study on three anatomically different pouches** *Journal of Fish Biology* **61**:1465–1480
- Chari T. *et al.* (2021) **Whole-animal multiplexed single-cell RNA-seq reveals transcriptional shifts across *Clytia* medusa cell types** *Science Advances* **7**
- Chen J. W., Galloway J. L (2014) **The development of zebrafish tendon and ligament progenitors** *Development* **141**:2035–2045

- Chen Y. P., Zhang Y., Jiang T. X., Barlow A. J., St. Amand T. R., Hu Y., Heaney S., Francis-West P., Chuong C. M., Maas R. (2000) **Conservation of early odontogenic signaling pathways in Aves** *PNAS* **97**:10044–10049
- Crump J. G., Maves L., Lawson N. D., Weinstein B. M., Kimmel C. B (2004) **An essential role for Fgfs in endodermal pouch formation influences later craniofacial skeletal patterning** *Development* **131**:5703–5716
- Denisenko E. *et al.* (2020) **Systematic assessment of tissue dissociation and storage biases in single-cell and single-nucleus RNA-seq workflows** *Genome Biology* **21**:1–25
- Ding H. L., Clouthier D. E., Artinger K. B (2013) **Redundant roles of PRDM family members in zebrafish craniofacial development** *Developmental Dynamics* **242**:67–79
- Docheva D., Hunziker E. B., Fässler R., Brandau O (2005) **Tenomodulin Is Necessary for Tenocyte Proliferation and Tendon Maturation** *Molecular and Cellular Biology* **25**:699–705
- Du J., Miller A. J., Widlund H. R., Horstmann M. A., Ramaswamy S., Fisher D. E (2003) **MLANA/MART1 and SILV/PMEL17/GP100 Are Transcriptionally Regulated by MITF in Melanocytes and Melanoma** *The American Journal of Pathology* **163**:333–343
- Dudley J. S., Hannaford P., Dowland S. N., Lindsay L. A., Thompson M. B., Murphy C. R., Van Dyke J. U., Whittington C. M. (2021) **Structural changes to the brood pouch of male pregnant seahorses (*Hippocampus abdominalis*) facilitate exchange between father and embryos** *Placenta* **114**:115–123
- Fabian P., Tseng K.-C., Thiruppathy M., Arata C., Chen H.-J., Smeeton J., Nelson N., Crump J. G (2022) **Lifelong single-cell profiling of cranial neural crest diversification in zebrafish** *Nature Communications* **13**:1–13
- Farnsworth D. R., Saunders L. M., Miller A. C (2020) **A single-cell transcriptome atlas for zebrafish development** *Developmental Biology* **459**:100–108
- Farrell J. A., Wang Y., Riesenfeld S. J., Shekhar K., Regev A., Schier A. F (2018) **Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis** *Science* **360**
- Feregrino C., Sacher F., Parnas O., Tschopp P (2019) **A single-cell transcriptomic atlas of the developing chicken limb** *BMC Genomics* **20**:1–15
- Flammang B. E., Ferry-Graham L. A., Rinewalt C., Ardizzone D., Davis C., Trejo T (2009) **Prey capture kinematics and four-bar linkages in the bay pipefish, *Syngnathus leptorhynchus*** *Zoology* **112**:86–96
- Fuess L. E., Bolnick D. I (2023) **GBE Single-Cell RNA Sequencing Reveals Microevolution of the Stickleback Immune System** *Genome Biology and Evolution* **15**
- Fuiten A. M., Cresko W. A (2021) **Evolutionary divergence of a *Hoxa2b* hindbrain enhancer in syngnathids mimics results of functional assays** *Development Genes and Evolution* **231**:57–71
- Gibert Y., Samarut E., Ellis M. K., Jackman W. R., Laudet V (2019) **The first formed tooth serves as a signalling centre to induce the formation of the dental row in zebrafish** *Proc Biol Sci* **286**

- Goldsmith M. I., Fisher S., Waterman R., Johnson S. L (2003) **Saltatory control of isometric growth in the zebrafish caudal fin is disrupted in long fin and rapunzel mutants** *Developmental Biology* **259**:303–317
- Gulati G. S. *et al.* (2020) **Single-cell transcriptional diversity is a hallmark of developmental potential** *Science* **367**:405–411
- Hafemeister C., Satija R (2019) **Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression** *Genome Biology* **20**
- Hain D. *et al.* (2022) **Molecular diversity and evolution of neuron types in the amniote brain** *Science* **377**
- Harada A., Shiota R., Okubo R., Yorifuji M., Sogabe A., Motomura H., Hiroi J., Yasumasu S., Kawaguchi M (2022) **Brood pouch evolution in pipefish and seahorse based on histological observation** *Placenta* **120**:88–96
- Harlin-Cognato A., Hoffman E. A., Jones A. G (2006) **Gene cooption without duplication during the evolution of a male-pregnancy gene in pipefish** *PNAS* **103**:19407–19412
- Healey H. M., Bassham S., Cresko W. A (2022) **Single-cell Iso-Sequencing enables rapid genome annotation for scRNAseq analysis** *Genetics* **220**
- Hong X., Wang Y., Wang K., Wei C., Li W., Yu L., Xu H., Zhu J., Zhu X., Liu X (2023) **Single-Cell Atlas Reveals the Hemocyte Subpopulations and Stress Responses in Asian Giant Softshell Turtle during Hibernation** *Biology* **12**
- Howe D. G. *et al.* (2013) **ZFIN, the Zebrafish Model Organism Database: Increased support for mutants and transgenics** *Nucleic Acids Research* **41**:D854–D860
- Hughes L. C. *et al.* (2018) **Comprehensive phylogeny of ray-finned fishes (Actinopterygii) based on transcriptomic and genomic data** *PNAS* **115**:6249–6254
- Jain I., Stroka C., Yan J., Huang W. M., Iovine M. K (2007) **Bone growth in zebrafish fins occurs via multiple pulses of cell proliferation** *Developmental Dynamics* **236**:2668–2674
- Johnson S. L., Nguyen A. T. N., Lister J. A (2011) **mitfa is required at multiple stages of melanocyte differentiation but not to establish the melanocyte stem cell** *Developmental Biology* **350**:405–413
- Jungerson H. F. E. (1910) **Ichthyotomical Contributions: The structure of the Aulostomidae, Syngnathidae, and Solenostomidae**
- Kaneda-Nakashima K. *et al.* (2022) **Role of Mel1/Prdm16 in bone differentiation and morphology** *Experimental Cell Research* **410**
- Kawasaki K (2009) **The SCPP gene repertoire in bony vertebrates and graded differences in mineralized tissues** *Development Genes and Evolution* **219**:147–157
- Kimura T., Takehana Y., Naruse K (2017) **Pnp4a Is the causal gene of the medaka iridophore Mutant guanineless G3: Genes, Genomes, Genetics** **7**:1357–1363
- Koiwai K., Kondo H., Hirono I (2023) **scRNA-seq Analysis of Hemocytes of Penaeid Shrimp Under Virus Infection** *Marine Biotechnology* **25**:488–502

- Kuo R. I., Tseng E., Eory L., Paton I. R., Archibald A. L., Burt D. W (2017) **Normalized long read RNA sequencing in chicken reveals transcriptome complexity similar to human** *BMC Genomics* **18**:1–19
- Kvarnemo C., Mobley K. B., Partridge C., Jones A. G., Ahnesjö I (2011) **Evidence of paternal nutrient provisioning to embryos in broad-nosed pipefish *Syngnathus typhle*** *Journal of Fish Biology* **78**:1725–1737
- Lamason R. L. *et al.* (2005) **Genetics: SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans** *Science* **310**:1782–1786
- Lange M. *et al.* (2023) **ZebraHub – Multimodal Zebrafish Developmental Atlas Reveals the State-Transition Dynamics of Late-Vertebrate Pluripotent Axial Progenitors** *BioRxiv*
- Langfelder P., Horvath S (2008) **WGCNA: An R package for weighted correlation network analysis** *BMC Bioinformatics* **9**:1–13
- Leerberg D. M., Hopton R. E., Draper B. W (2019) **Fibroblast growth factor receptors function redundantly during zebrafish embryonic development** *Genetics* **212**:1301–1319
- Leysen H., Roos G., Adriaens D (2011) **Morphological variation in head shape of pipefishes and seahorses in relation to snout length and developmental growth** *Journal of Morphology* **272**:1259–1270
- Li H (2018) **Minimap2: Pairwise alignment for nucleotide sequences** *Bioinformatics* **34**:3094–3100
- Lin Q. *et al.* (2016) **The seahorse genome and the evolution of its specialized morphology** *Nature* **540**:395–399
- Liu Z. *et al.* (2016) **The channel catfish genome sequence provides insights into the evolution of scale formation in teleosts** *Nature Communications* **7**
- Manning C. G., Foster S. J., Vincent A. C. J (2019) **A review of the diets and feeding behaviours of a family of biologically diverse marine fishes (Family Syngnathidae)** *Reviews in Fish Biology and Fisheries* **29**:197–221
- McCarthy N., Sidik A., Bertrand J. Y., Eberhart J. K (2016) **An Fgf-Shh signaling hierarchy regulates early specification of the zebrafish skull** *Developmental Biology* **415**:261–277
- Melamed P., Xue Y., Poon J. F. D., Wu Q., Xie H., Yeo J., Foo T. W. J., Chua H. K (2005) **The male seahorse synthesizes and secretes a novel C-type lectin into the brood pouch during early pregnancy** *The FEBS Journal* **272**:1221–1235
- Miao M., Bruce A. E. E., Bhanji T., Davis E. C., Keeley F. W (2007) **Differential expression of two tropoelastin genes in zebrafish** *Matrix Biology* **26**:115–124
- Miyake A., Itoh N (2013) **Fgf22 regulated by Fgf3/Fgf8 signaling is required for zebrafish midbrain development** *Biology Open* **2**:515–524
- Mori S., Nakamura T (2022) **Redeployment of odontode gene regulatory network underlies dermal denticle formation and evolution in suckermouth armored catfish** *Scientific Reports* **12**

- Moriyama Y., Ito F., Takeda H., Yano T., Okabe M., Kuraku S., Keeley F. W., Koshiba-Takeuchi K (2016) **Evolution of the fish heart by sub/neofunctionalization of an elastin gene** *Nature Communications* **7**:1–10
- Morrison J. A. *et al.* (2017) **Single-cell transcriptome analysis of avian neural crest migration reveals signatures of invasion and molecular transitions** *ELife* **6**
- Parker J., Guslund N. C., Jentoft S., Roth O (2022) **Characterization of Pipefish Immune Cell Populations Through Single-Cell Transcriptomics** *Frontiers in Immunology* **13**
- Partridge C., Shardo J., Boettcher A (2007) **Osmoregulatory role of the brood pouch in the euryhaline Gulf pipefish, *Syngnathus scovelli*** *Comparative Biochemistry and Physiology - A Molecular and Integrative Physiology* **147**:556–561
- Potts H. G., Lemieux M. E., Rice E. S., Warren W., Choudhury R. P., Mommersteeg M. T. M (2022) **Discordant Genome Assemblies Drastically Alter the Interpretation of Single-Cell RNA Sequencing Data Which Can Be Mitigated by a Novel Integration Method** *Cells* **11**
- Qu M. *et al.* (2021) **Seadragon genome analysis provides insights into its phenotype and sex determination locus** *Science Advances* **7**
- Ramesh B., Small C. M., Healey H., Johnson B., Barker E., Currey M., Bassham S., Myers M., Cresko W. A., Jones A. G (2023) **Improvements to the Gulf pipefish *Syngnathus scovelli* genome** *GigaByte* **2023**
- Ripley J. L., Foran C. M (2006) **Differential parental nutrient allocation in two congeneric pipefish species (*Syngnathidae*: *Syngnathus* spp)** *Journal of Experimental Biology* **209**:1112–1121
- Ripley J. L., Williams P. S., Foran C. M (2010) **Morphological and quantitative changes in paternal brood-pouch vasculature during embryonic development in two syngnathus pipefishes** *Journal of Fish Biology* **77**:67–79
- Rose E., Johnson B. D., Armitage V., Jones A. G (2023) **The response of the brood pouch transcriptome to synthetic estrogen exposure in the Gulf pipefish (*Syngnathus scovelli*)** *Frontiers in Marine Science* **10**:1–12
- Roth O., Klein V., Beemelmans A., Scharsack J. P., Reusch T. B. H (2012) **Male pregnancy and biparental immune priming** *American Naturalist* **180**:802–814
- Roth O. *et al.* (2020) **Evolution of male pregnancy associated with remodeling of canonical vertebrate immunity in seahorses and pipefishes** *PNAS* **117**:9431–9439
- Royan M. R., Siddique K., Csucs G., Puchades M. A., Nourizadeh-Lillabadi R., Bjaalie J. G., Henkel C. V., Weltzien F. A., Fontaine R (2021) **3D Atlas of the Pituitary Gland of the Model Fish Medaka (*Oryzias latipes*)** *Frontiers in Endocrinology* **12**
- Salamanca-Díaz D. A., Schulreich S. M., Cole A. G., Wanninger A (2022) **Single-Cell RNA Sequencing Atlas From a Bivalve Larva Enhances Classical Cell Lineage Studies** *Frontiers in Ecology and Evolution* **9**
- Schneider R. F., Gunter H. M., Salewski I., Woltering J. M., Meyer A (2023) **Growth dynamics and molecular bases of evolutionary novel jaw extensions in halfbeaks and needlefishes (*Beloniformes*)** *Molecular Ecology* **32**:5798–5811

- Schneider R. F., Woltering J. M., Adriaens D., Roth O (2023) **A comparative analysis of the ontogeny of syngnathids (pipefishes and seahorses) reveals how heterochrony contributed to their diversification** *Developmental Dynamics* **252**:553–588
- Shema E., Bernstein B. E., Buenrostro J. D (2019) **Single-cell and single-molecule epigenomics to uncover genome regulation at unprecedented resolution** *Nature Genetics* **51**:19–25
- Shull L. C., Lencer E. S., Kim H. M., Goyama S., Kurokawa M., Costello J. C., Jones K., Artinger K. B (2022) **PRDM paralogs antagonistically balance Wnt/ β -catenin activity during craniofacial chondrocyte differentiation** *Development* **149**
- Shull L. C., Sen R., Menzel J., Goyama S., Kurokawa M., Artinger K. B (2020) **The conserved and divergent roles of Prdm3 and Prdm16 in zebrafish and mouse craniofacial development** *Developmental Biology* **461**:132–144
- Sire J. Y., Delgado S. C., Girondot M (2008) **Hen's teeth with enamel cap: From dream to impossibility** *BMC Evolutionary Biology* **8**:1–11
- Small C. M., Bassham S., Catchen J., Amores A., Fuiten A. M., Brown R. S., Jones A. G., Cresko W. A (2016) **The genome of the Gulf pipefish enables understanding of evolutionary innovations** *Genome Biology* **17**
- Small C. M., Harlin-Cognato A. D., Jones A. G (2013) **Functional similarity and molecular divergence of a novel reproductive transcriptome in two male-pregnant Syngnathus pipefish species** *Ecology and Evolution* **3**:4092–4108
- Small C. M., Healey H. M., Currey M. C., Beck E. A., Catchen J., Lin A. S. P., Cresko W. A., Bassham S (2022) **Leafy and weedy seadragon genomes connect genic and repetitive DNA features to the extravagant biology of syngnathid fishes** *PNAS* **119**
- Soldatov R. *et al.* (2019) **Spatiotemporal structure of cell fate decisions in murine neural crest** *Science* **364**
- Sommer S., Whittington C. M., Wilson A. B (2012) **Standardised classification of pre-release development in male-brooding pipefish, seahorses, and seadragons (Family Syngnathidae)** *BMC Developmental Biology* **12**:12–15
- Songco-Casey J. O., Coffing G. C., Piscopo D. M., Pungor J. R., Kern A. D., Miller A. C., Niell C. M (2022) **Cell types and molecular architecture of the Octopus bimaculoides visual system** *Current Biology : CB* **32**
- Steger J., Cole A. G., Denner A., Lebedeva T., Genikhovich G., Ries A., Reischl R., Taudes E., Lassnig M., Technau U (2022) **Single-cell transcriptomics identifies conserved regulators of neuroglandular lineages** *Cell Reports* **40**
- Stiller J., Short G., Hamilton H., Saarman N., Longo S., Wainwright P., Rouse G. W., Simison W. B (2022) **Phylogenomic analysis of Syngnathidae reveals novel relationships, origins of endemic diversity and variable diversification rates** *BMC Biology* **20**:1–21
- Stock D. W., Jackman W. R., Trapani J (2006) **Developmental genetic mechanisms of evolutionary tooth loss in cypriniform fishes** *Development* **133**:3127–3137

- Stölting K. N., Wilson A. B (2007) **Male pregnancy in seahorses and pipefish: Beyond the mammalian model** *BioEssays* **29**:884–896
- Szabo-Rogers H. L., Geetha-Loganathan P., Nimmagadda S., Fu K. K., Richman J. M (2008) **FGF signals from the nasal pit are necessary for normal facial morphogenesis** *Developmental Biology* **318**:289–302
- Tardaguila M. *et al.* (2018) **SQANTI: Extensive characterization of long-read transcript sequences for quality control in full-length transcriptome identification and quantification** *Genome Research* **28**
- Teichmann S. A., Babu M. M (2004) **Gene regulatory network growth by duplication** *Nature Genetics* **36**:492–496
- Tengfei D. *et al.* (2021) **Multiple Fetal Nutritional Patterns Before Parturition in Viviparous Fish *Sebastes schlegelii* (Hilgendorf, 1880)** *Frontiers in Marine Science* **7**
- Thisse B., Thisse C. (2004) **Fast Release Clones: A High Throughput Expression Analysis** *ZFIN Direct Data Submission*
- Thisse C., Thisse B (2007) **High-resolution in situ hybridization to whole-mount zebrafish embryos** *Nature Protocols* **3**:59–69
- Tokita M., Chaeychomsri W., Siruntawineti J (2013) **Developmental basis of toothlessness in turtles: Insight into convergent evolution of vertebrate morphology** *Evolution* **67**:260–273
- Ton M. L. N., Guibentif C., Göttgens B. (2020) **Single cell genomics and developmental biology: moving beyond the generation of cell type catalogues** *Current Opinion in Genetics and Development* **64**:66–71
- Ton M. L. N. *et al.* (2023) **An atlas of rabbit development as a model for single-cell comparative genomics** *Nature Cell Biology* **25**:1061–1072
- Tseng E (2021) **cDNA cupcake** *GitHub*
- Tucker A. S., Headon D. J., Schneider P., Ferguson B. M., Overbeek P., Tschopp J., Sharpe P. T (2000) **Edar/Eda interactions regulate enamel knot formation in tooth morphogenesis** *Development* **127**:4691–4700
- Tucker A., Sharpe P (2004) **The cutting-edge of mammalian development; how the embryo makes teeth** *Nature Reviews Genetics* **5**:499–508
- Uniken Venema W. T. C. *et al.* (2022) **Gut mucosa dissociation protocols influence cell type proportions and single-cell gene expression levels** *Scientific Reports* **12**:1–14
- Van Wassenbergh S., Roos G., Genbrugge A., Leysen H., Aerts P., Adriaens D., Herrel A. (2009) **Suction is kid's play: extremely fast suction in newborn seahorses** *Biology Letters* **5**:200–203
- Van Wassenbergh S., Strother J. A., Flammang B. E., Ferry-Graham L. A., Aerts P. (2007) **Extremely fast prey capture in pipefish is powered by elastic recoil** *Journal of The Royal Society Interface* **5**:285–296

- Verreijdt L., Debais-Thibaud M., Borday-Birraux V., Van Der Heyden C., Sire J. Y., Huysseune A. (2006) **Expression of the dlx gene family during formation of the cranial bones in the zebrafish (Danio rerio): Differential involvement in the visceral skeleton and braincase** *Developmental Dynamics* **235**:1371–1389
- Vonk A. C. *et al.* (2023) **Single-cell analysis of lizard blastema fibroblasts reveals phagocyte-dependent activation of Hedgehog-responsive chondrogenesis** *Nature Communications* **14**:1–20
- Wagner A (2011) **The molecular origins of evolutionary innovations** *Trends in Genetics* **27**:397–410
- Wagner D. E., Weinreb C., Collins Z. M., Briggs J. A., Megason S. G., Klein A. M (2018) **Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo** *Science* **360**:981–987
- Wagner G. P., Lynch V. J (2010) **Evolutionary novelties** *Current Biology* **20**:R48–R52
- Walshe J., Mason I (2003) **Fgf signalling is required for formation of cartilage in the head** *Developmental Biology* **264**:522–536
- Wang W. *et al.* (2024) **The TET-Sall4-BMP regulatory axis controls craniofacial cartilage development** *Cell Reports* **43**
- Wetzel J. T., Wourms J. P (2004) **Embryogenesis in the Dwarf Seahorse, Hippocampus zosterae (Syngnathidae)** *Gulf and Caribbean Research* **16**:27–35
- Whittington C. M., Friesen C. R (2020) **The evolution and physiology of male pregnancy in syngnathid fishes** *Biological Reviews* **95**:1252–1272
- Whittington C. M., Griffith O. W., Qi W., Thompson M. B., Wilson A. B (2015) **Seahorse Brood Pouch Transcriptome Reveals Common Genes Associated with Vertebrate Pregnancy** *Molecular Biology and Evolution* **32**:3114–3131
- Whittington C. M., Griffith O. W., Qi W., Thompson M. B., Wilson A. B (2015) **Seahorse Brood Pouch Transcriptome Reveals Common Genes Associated with Vertebrate Pregnancy** *Molecular Biology and Evolution* **32**
- Williams R. M., Candido-Ferreira I., Repapi E., Gavriouchkina D., Senanayake U., Ling I. T. C., Telenius J., Taylor S., Hughes J., Sauka-Spengler T (2019) **Reconstruction of the Global Neural Crest Gene Regulatory Network In Vivo** *Developmental Cell* **51**:255–276
- Williams R. M., Candido-Ferreira I., Repapi E., Gavriouchkina D., Senanayake U., Ling I. T. C., Telenius J., Taylor S., Hughes J., Sauka-Spengler T (2019) **Reconstruction of the Global Neural Crest Gene Regulatory Network In Vivo** *Developmental Cell* **51**:255–276
- Wolf M. *et al.* (2024) **Near chromosome-level and highly repetitive genome assembly of the snake pipefish Entelurus aequoreus (Syngnathiformes: Syngnathidae)** *GigaByte* **2024**:1–13
- Woronowicz K. C., Schneider R. A (2019) **Molecular and cellular mechanisms underlying the evolution of form and function in the amniote jaw** *EvoDevo* **10**:1–21
- Wourms J. P (1981) **Viviparity: The Maternal-Fetal Relationship in Fishes** *American Zoologist* **21**:473–515

Woych J. *et al.* (2022) **Cell-type profiling in salamanders identifies innovations in vertebrate forebrain evolution** *Science* **377**

Young M. D., Behjati S (2020) **SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data** *GigaScience* **9**:1–10

Zhang Y., Ravi V., Qin G., Dai H., Zhang H., Han F., Wang X., Liu Y (2020) **Comparative genomics reveal shared genomic changes in syngnathid fishes and signatures of genetic convergence with placental mammals** *National Science Review* **7**:964–977

Zhou J., Gao Y., Lan Y., Jia S., Jiang R (2013) **Pax9 regulates a molecular network involving Bmp4, Fgf10, Shh signaling and the Osr2 transcription factor to control palate morphogenesis** *Development* **140**:4709–4718

Author information

Hope M Healey

Institute of Ecology and Evolution, University of Oregon, Eugene, United States
ORCID iD: [0000-0001-9978-1553](https://orcid.org/0000-0001-9978-1553)

Hayden B Penn

Institute of Ecology and Evolution, University of Oregon, Eugene, United States
ORCID iD: [0009-0000-3090-5423](https://orcid.org/0009-0000-3090-5423)

Clayton M Small

Institute of Ecology and Evolution, University of Oregon, Eugene, United States, School of Computer and Data Science, University of Oregon, Eugene, United States
ORCID iD: [0000-0003-1615-7590](https://orcid.org/0000-0003-1615-7590)

Susan Bassham

Institute of Ecology and Evolution, University of Oregon, Eugene, United States
ORCID iD: [0000-0002-7309-2095](https://orcid.org/0000-0002-7309-2095)

Vithika Goyal

Institute of Ecology and Evolution, University of Oregon, Eugene, United States
ORCID iD: [0009-0009-8601-8474](https://orcid.org/0009-0009-8601-8474)

Micah A Woods

Institute of Ecology and Evolution, University of Oregon, Eugene, United States
ORCID iD: [0009-0004-2156-3352](https://orcid.org/0009-0004-2156-3352)

William A Cresko

Institute of Ecology and Evolution, University of Oregon, Eugene, United States, Knight Campus for Accelerating Scientific Impact, University of Oregon, Eugene, United States
ORCID iD: [0000-0002-3496-8074](https://orcid.org/0000-0002-3496-8074)

For correspondence: wcresko@uoregon.edu

Editors

Reviewing Editor

Frank Chan

University of Groningen, Groningen, Netherlands

Senior Editor

Claude Desplan

New York University, New York, United States of America

Reviewer #1 (Public review):

Syngnathid fishes (seahorses, pipefishes, and seadragons) present very particular and elaborated features among teleosts and a major challenge is to understand the cellular and molecular mechanisms that permitted such innovations and adaptations. The study provides a valuable new resource to investigate the morphogenetic basis of four main traits characterizing syngnathids, including the elongated snout, toothlessness, dermal armor and male pregnancy. More particularly, the authors have focused on a late stage of pipefish organogenesis to perform single-cell RNA-sequencing (scRNA-seq) completed by in situ hybridization analyses to identify molecular pathways implicated in the formation of the different specific traits.

The first set of data explores the scRNA-seq atlas composed of 35,785 cells from two samples of gulf pipefish embryos that authors have been able to classify into major cell types characterizing vertebrate organogenesis, including epithelial, connective, neural and muscle progenitors. To affirm identities and discover potential properties of clusters, authors primarily use KEGG analysis that reveals enriched genetic pathways in each cell types. After revisions, the authors have provided extended supplementary files to well interpret the dataset and some statements have been clarified. I thank the authors for the revisions/completions of ISH results compared to initial submission.

To conclude, the scRNA-seq dataset in this unconventional model organism will be useful for the community and will provide clues for future research to understand the extraordinary evolution of the Syngnathidae family.

<https://doi.org/10.7554/eLife.97764.2.sa3>

Reviewer #2 (Public review):

Summary:

The authors present the first single-cell atlas for syngathid fishes, providing a resource for future evolution & development studies in this group.

Strengths:

The concept here is simple and I find the manuscript to be well written. I like the in situ hybridization of marker genes >> this is really nice. I also appreciate the gene co-expression analysis to identify modules of expression. There are no explicit hypotheses tested in the manuscript, but the discovery of these cell types should have value in this organism and in the determination of morphological novelties in seahorses and their relatives.

Weaknesses:

I think there are a few computational analyses that might improve the generality of the results.

(1) The cell types: The authors use marker gene analysis and KEGG pathways to identify cell types. I'd suggest a tool like SAMap (<https://elifesciences.org/articles/66747>) which compares single cell data sets from distinct organisms to identify 'homologous' cell types -- I imagine the zebrafish developmental atlases could serve as a reasonable comparative reference.

(2) Trajectory analyses: Authors suggest that their analyses might identify progenitor cell states and perhaps related differentiated states. They might explore cytoTRACE and/or pseudotime-based trajectory analyses to more fully delineate these ideas.

(3) Cell-cell communication: I think it's very difficult to identify 'tooth primordium' cell types, because cell types won't be defined by organ in this way. for instance dental glia will cluster with other glia, dental mesenchyme will likely cluster with other mesenchymal cell types. so the histology and ISH is most convincing in this regard. having said this, given the known signaling interactions in the developing tooth (and in development generally) the authors might explore cell-cell communication analysis (e.g., CellChat) to identify cell types that may be interacting.

Comments on revisions:

I feel essentially the same about this manuscript. it's a useful resource for future experimental forays into this unique system. The team made improvements to deal with comments from other reviewers related to quality of confirmatory in situ hybridization. This is good.

Regarding their response that one can't use CellChat if you're not working in mice or human, this is inaccurate. the assumption one makes is that ligand-receptor pairs and signaling pathways have conserved functions across animals (vertebrates). It's the same assumption the authors make when using the KEGG pathway to score enrichment of pathways in clusters. CellChat used in fishes in Johnson et al 2023 Nature Communications | (2023) 14:4891.

<https://doi.org/10.7554/eLife.97764.2.sa2>

Reviewer #3 (Public review):

Summary:

This study established a single-cell RNA sequencing atlas of pipefish embryos. The results obtained identified unique gene expression patterns for pipefish-specific characteristics, such as *fgf22* in the tip of the palatoquadrate and Meckel's cartilage, broadly informing the genetic mechanisms underlying morphological novelty in teleost fishes. The data obtained are unique and novel, potentially important in understanding fish diversity. Thus, I would enthusiastically support this manuscript if the authors improve it to generate stronger and more convincing conclusions than the current forms.

Weakness:

Regarding the expression of *sfrp1a* and *bmp4* dorsal to the elongating ethmoid plate and surrounding the ceratohyal: Are their expression patterns spatially extended or broader compared to the pipefish ancestor? Is there a much closer species available to compare gene expression patterns with pipefish? Did the authors consider using other species closely related to pipefish for ISH? *Sfrp1a* and *bmp4* may be expressed in the same regions of much more closely related species without face elongation. I understand that embryos of such species are not always accessible, but it is also hard to argue responsible genes for a specific

phenotype by only comparing gene expression patterns between distantly related species (e.g., pipefish vs. zebrafish). Due to the same reason, I would not directly compare/argue gene expression patterns between pipefish and mice, although I should admit that mice gene expression patterns are sometimes helpful to make a hypothesis of fish evolution. Alternatively, can the authors conduct ISH in other species of pipefish? If the expression patterns of *sfrp1a* and *bmp4* are common among fishes with face elongation, the conclusion would become more solid. If these embryos are not available, is it possible to reduce the amount of Wnt and BMP signal using Crispr/Cas, MO, or chemical inhibitor? I do think that there are several ways to test the Wnt and/or BMP hypothesis in face elongation.

<https://doi.org/10.7554/eLife.97764.2.sa1>

Author response:

The following is the authors' response to the original reviews.

Public Reviews:

Reviewer #1 (Public Review):

Syngnathid fishes (seahorses, pipefishes, and seadragons) present very particular and elaborated features among teleosts and a major challenge is to understand the cellular and molecular mechanisms that permitted such innovations and adaptations. The study provides a valuable new resource to investigate the morphogenetic basis of four main traits characterizing syngnathids, including the elongated snout, toothlessness, dermal armor, and male pregnancy. More particularly, the authors have focused on a late stage of pipefish organogenesis to perform single-cell RNA-sequencing (scRNA-seq) completed by in situ hybridization analyses to identify molecular pathways implicated in the formation of the different specific traits.

The first set of data explores the scRNA-seq atlas composed of 35,785 cells from two samples of gulf pipefish embryos that authors have been able to classify into major cell types characterizing vertebrate organogenesis, including epithelial, connective, neural, and muscle progenitors. To affirm identities and discover potential properties of clusters, authors primarily use KEGG analysis that reveals enriched genetic pathways in each cell types. While the analysis is informative and could be useful for the community, some interpretations appear superficial and data must be completed to confirm identities and properties. Notably, supplementary information should be provided to show quality control data corresponding to the final cell atlas including the UMAP showing the sample source of the cells, violin plots of gene count, UMI count, and mitochondrial fraction for the overall

dataset and by cluster, and expression profiles on UMAP of selected markers characterizing cluster identities.

We thank the reviewer for these suggestions, and have added several figures and supplemental files in response. We added a supplemental UMAP showing the sample that each cell originated (S1). We also added supplemental violin plots for each sample showing the gene count, unique molecular identifier (UMI) count, mitochondrial fraction, and the doublet scores (S2). We added feature plots of zebrafish marker genes for these major cell types and marker genes identified from our dataset to the supplement (S3:S57). We also provided two supplemental files with marker genes. These changes should clarify the work that went into labeling the clusters. Although some of the cluster labels are general, we decided it would be unwise to label clusters with speculated specific annotations. We only gave specific annotations to clusters with concrete markers and/or in situ hybridization (ISH)

results that cemented an annotation. As shown in the new supplemental figures and files, certain clusters had clear, specific markers while others did not. Therefore, we used caution when we annotated clusters without distinct markers.

The second set of data aims to correlate the scRNA-seq analysis with in situ hybridizations (ISH) in two different pipefish (gulf and bay) species to identify and characterize markers spatially, and validate cell types and signaling pathways active in them. While the approach is rational, the authors must complete the data and optimize labeling protocols to support their statements. One major concern is the quality of ISH stainings and images; embryos show a high degree of pigmentation that could hide part of the expression profile, and only subparts and hardly detectable tissues/stainings are presented. The authors should provide clear and good-quality images of ISH labeling on whole-mount specimens, highlighting the magnification regions and all other organs/structures (positive controls) expressing the marker of interest along the axis. Moreover, ISH probes have been designed and produced on gulf pipefish genome and cDNA respectively, while ISH labeling has been performed indifferently on bay or gulf pipefish embryos and larvae. The authors should specify stages and species on figure panels and should ensure sequence alignment of the probe-targeted sequences in the two species to validate ISH stainings in the bay pipefish. Moreover, spatiotemporal gene expression being a very dynamic process during embryogenesis, interpretations based on undefined embryonic and larval stages of pipefish development and compared to 3dpf zebrafish are insufficient to hypothesize on developmental specificities of pipefish features, such as on the absence of tooth primordia that could represent a very discrete and transient cell population. The ISH analyses would require a clean and precise spatiotemporal expression comparison of markers at the level of the entire pipefish and zebrafish specimens at well-defined stages, otherwise, the arguments proposed on teleost innovations and adaptations turn out to be very speculative.

We are appreciative of the reviewer's feedback. We primarily used the *in situ* hybridization (ISH) data as supplementary to the scRNAseq library and we are aware that further evidence is necessary to identify origins of syngnathid's evolutionary novelties. Our goal was to provide clues for the developmental genetic basis of syngnathid derived features. We hope that our study will inspire future investigations and are excited for the prospect that future research could include this reviewer's ideas.

All of the developmental stages and species information for the embryos used were in the figure captions as well as in supplemental file 6. Because we primarily used wild caught embryos, we did not have specific ages of most embryos. Syngnathid species are challenging to culture in the laboratory, and extracting embryos requires euthanizing the father which makes it difficult to obtain enough embryos for ISH. In addition, embryos do not survive long when removed from the brood pouch prematurely. We supplemented our ISH with bay pipefish caught off the Oregon coast because these fish have large broods. Wild caught pregnant male bay pipefish were immediately euthanized, and their broods were fixed. Because we did not have their age, we classified them based on developmental markers such as presence of somites and the extent of craniofacial elongation. Although these classification methods are not ideal, they are consistent with the syngnathid literature (Sommer et al. 2012). Since the embryos used for the ISH were primarily wild caught, we had a few different developmental stages represented in our ISH data. For our tooth primordia search, we used embryos from the same brood (therefore, same stage) for these experiments.

We understand the concern for the degree of pigmentation in the samples. We completed numerous bleach trials before embarking on the *in situ* hybridization experiments. After completing a bleach trial with a probe created from the gene *tnmd* for ISH_, we noticed that the bleached embryos were missing expression domains found in the unbleached embryos. We were, therefore, concerned that using bleached embryos for our experiments would

result incorrect conclusions about the expression domains of these genes. We sparingly used bleaching at older stages, hatched larvae, where it was fundamentally necessary to see staining. As stated above, the primary goal of this manuscript was to generate and annotate the first scRNA-seq atlas in a syngnathid, and the ISHs were utilized to support inferred cluster annotations only through a positive identification of marker gene expression in expected tissues/cells. Therefore, the obscuring of gene expression by pigmentation would have resulted in the absence of evidence for a possible cluster annotation, not an incorrect annotation.

For the ease of viewing the ISHs, we improved annotations and clarity. We increased the brightness and contrast of images. In the original submission, we had to lower the image resolution to make the submission file smaller. We hope that these improvements plus the true image quality improves clarity of ISH results. We also included alignments in our supplementary files of bay pipefish sequences to the Gulf pipefish probes to showcase the high degree of sequence similarity.

Sommer, S., Whittington, C. M., & Wilson, A. B. (2012). Standardised classification of pre-release development in male-brooding pipefish, seahorses, and seadragons (Family Syngnathidae). *BMC Developmental Biology*, 12, 12–15.

To conclude, whereas the scRNA-seq dataset in this unconventional model organism will be useful for the community, the spatiotemporal and comparative expression analyses have to be thoroughly pushed forward to support the claims. Addressing these points is absolutely necessary to validate the data and to give new insights to understand the extraordinary evolution of the Syngnathidae family.

We really appreciate the reviewer's enthusiasm for syngnathid research, and hope that the additional files and explanation of the supporting role of the ISHs have adequately addressed their concerns. We share the reviewer's enthusiasm and are excited for future work that can extend this study.

Reviewer #2 (Public Review):

Summary:

The authors present the first single-cell atlas for syngnathid fishes, providing a resource for future evolution & development studies in this group.

Strengths:

The concept here is simple and I find the manuscript to be well written. I like the in situ hybridization of marker genes - this is really nice. I also appreciate the gene co-expression analysis to identify modules of expression. There are no explicit hypotheses tested in the manuscript, but the discovery of these cell types should have value in this organism and in the determination of morphological novelties in seahorses and their relatives.

We are grateful for this reviewer's appreciation of the huge amount of work that went into this study, and we agree that the in situ hybridizations (ISHs) support the scRNAseq study as we intended. We appreciate that the reviewer thinks that this work will add value to the syngnathid field.

Weaknesses:

I think there are a few computational analyses that might improve the generality of the results.

(1) The cell types: The authors use marker gene analysis and KEGG pathways to identify cell types. I'd suggest a tool like SAMap (<https://elifesciences.org/articles/66747>) which compares single-cell data sets from distinct organisms to identify 'homologous' cell types - I imagine the zebrafish developmental atlases could serve as a reasonable comparative reference.

We appreciate the reviewer's request, and in fact we would have loved to integrate our dataset with zebrafish. However, syngnathid's unique craniofacial development makes it challenging to determine the appropriate stage for comparison. While 3 days post fertilization (dpf) zebrafish data were appropriate for comparisons of certain cell types (e.g. epidermal cells), it would have been problematic for other cell types (e.g. osteoblasts) that are not easily detectable until older zebrafish stages. Therefore, determining equivalent stages between these species is difficult and contains potential for error. Future research should focus on trying to better match stages across syngnathids and zebrafish (and other fish species such as stickleback). Studies of this nature promise to uncover the role of heterochrony in the evo-devo of syngnathid's unique snouts.

(2) Trajectory analyses: The authors suggest that their analyses might identify progenitor cell states and perhaps related differentiated states. They might explore cytoTRACE and/or pseudotime-based trajectory analyses to more fully delineate these ideas.

We thank the reviewer for this suggestion! We added a trajectory analysis using cytoTRACE to the manuscript. It complemented our KEGG analysis well (L172-175; S73) and has improved the manuscript.

(3) Cell-cell communication: I think it's very difficult to identify 'tooth primordium' cell types, because cell types won't be defined by an organ in this way. For instance, dental glia will cluster with other glia, and dental mesenchyme will likely cluster with other mesenchymal cell types. So the histology and ISH is most convincing in this regard. Having said this, given the known signaling interactions in the developing tooth (and in development generally) the authors might explore cell-cell communication analysis (e.g., CellChat) to identify cell types that may be interacting.

We agree! It would have been a wonderful addition to the paper to include a cell-cell communication analysis. One limitation of CellChat is that it only includes mouse and human orthologs. Given concerns of reviewer #3 for mouse-syngnathid comparisons, we decided to not pursue CellChat for this study. We are looking forward to future cell communication resources that include teleost fishes.

Reviewer #3 (Public Review):

Summary:

This study established a single-cell RNA sequencing atlas of pipefish embryos. The results obtained identified unique gene expression patterns for pipefish-specific characteristics, such as fgf22 in the tip of the palatoquadrate and Meckel's cartilage, broadly informing the genetic mechanisms underlying morphological novelty in teleost fishes. The data obtained are unique and novel, potentially important in understanding fish diversity. Thus, I would enthusiastically support this manuscript if the authors improve it to generate stronger and more convincing conclusions than the current forms.

Thank you, we appreciate the reviewer's enthusiasm!

Weaknesses:

*Regarding the expression of *sfrp1a* and *bmp4* dorsal to the elongating ethmoid plate and surrounding the ceratohyal: are their expression patterns spatially extended or broader compared to the pipefish ancestor? Is there a much closer species available to compare gene expression patterns with pipefish? Did the authors consider using other species closely related to pipefish for ISH? *Sfrp1a* and *bmp4* may be expressed in the same regions of much more closely related species without face elongation. I understand that embryos of such species are not always accessible, but it is also hard to argue responsible genes for a specific phenotype by only comparing gene expression patterns between distantly related species (e.g., pipefish vs. zebrafish). Due to the same reason, I would not directly compare/argue gene expression patterns between pipefish and mice, although I should admit that mice gene expression patterns are sometimes helpful to make a hypothesis of fish evolution. Alternatively, can the authors conduct ISH in other species of pipefish? If the expression patterns of *sfrp1a* and *bmp4* are common among fishes with face elongation, the conclusion would become more solid. If these embryos are not available, is it possible to reduce the amount of Wnt and BMP signal using Crispr/Cas, MO, or chemical inhibitor? I do think that there are several ways to test the Wnt and/or BMP hypothesis in face elongation.*

We appreciate the reviewer's suggestion, and their recognition for challenges within this system. In response to this comment, we completed further *in situ* hybridization experiments in threespine stickleback, a short snouted fish that is much more closely related to syngnathids than is zebrafish, to make comparisons with pipefish craniofacial expression patterns (S76-S79). We added ISH data for the signaling genes (*fgf22*, *bmp4*, and *sfrp1a*) as well as *prdm16*. Through adding this additional ISH results, we speculated that craniofacial expression of *bmp4*, *sfrp1a*, and *prdm16* is conserved across species. However, compared to the specific ceratohyal/ethmoid staining seen in pipefish, stickleback had broad staining throughout the jaws and gills. These data suggest that pipefish have co-opted existing developmental gene networks in the development of their derived snouts. We added this interpretation to the results and discussion of the manuscript (L244-L248; L262-277; L444-470).

Recommendations for the authors:

Reviewing Editor (Recommendations for the Authors):

We hope that the eLife assessment, as well as the revisions specified here, prove helpful to you for further revisions of your manuscript.

Revisions considered essential:

*(1) Marker genes and single-cell dataset analyses. While these analyses have been performed to a good standard in broad terms, there is a majority view here that cell type annotations and trajectory analyses can be improved. In particular, there is question about the choice of marker genes for the current annotation. For one it can depend on the use of single marker genes (see *tnnt1* example for clusters 17 and 31). Here, we recommend incorporating results from SAMap and trajectory analysis (e.g., cytoTRACE or standard pseudotime).*

Because of the reviewer comments, we became aware that we insufficiently communicated how cell clusters were annotated. We did mention in the manuscript that we did not use single marker genes to annotate clusters, but instead we used multiple marker genes for each cluster for the annotation process. We used both marker genes derived from our dataset and marker genes identified from zebrafish resources for cluster annotation. We chose single marker genes for each cluster for visualization purposes and for *in situ* hybridizations. However, it is clear from the reviewers' comments that we needed to make more clear how

the annotations were performed. To make this effort more clear in our revision, we included two new supplementary files – one with Seurat derived marker genes and one with marker genes derived from our DotPlot method. We also included extensive supplementary figures highlighting different markers. Using DanioCell, we identified 6 zebrafish markers per major cell type and showed their expression patterns in our atlas with FeaturePlots. We also included feature plots of the top 6 marker genes for each cluster. We hope that the addition of these 40+ plots (S3:S57) to the supplement fully addresses these concerns.

We appreciated the suggestion of cytortrace from reviewer #2! We ran cytortrace on three major cell lineages (neural, muscle, and connective; S73) which complemented our KEGG analysis in suggesting an undifferentiated fate for clusters 8, 10, and 16. We chose to not run SAMap because it is a scRNA-seq library integration tool. Although we compared our lectin epidermal findings to 3 dpf zebrafish scRNA-seq data, we did not integrate the datasets out of concern that we could draw erroneous conclusions for other cell types. Future work that explores this technical challenge may uncover the role of heterochrony in syngnathid craniofacial development. We detail these changes more fully in our responses to reviewers.

(2) The claims regarding evolutionary novelty and/or the genes involved are considered speculative. In part, this comes from relying too heavily on comparisons against zebrafish, as opposed to more closely related species. For example, the discussion regarding C-type lectin expression in the epidermis and KEGG enrichment (lines 358 - 364) seems confusing. Another good example here is the discussion on sfrp1a (lines 258 - 261). Here, the text seems to suggest craniofacial sfrp1a expression (or specifically ethmoid expression?) is connected to the development of the elongated snout in pipefish. However, craniofacial expression of sfrp1a is also reported in the arctic charr, which the authors grouped into fishes with derived craniofacial structures. Separately, sfrp2 expression was also reported in stickleback fish, for example. Do these different discussions truly support the notion that sfrp1a expression is all that unique in pipefish, rather than that pipefish and zebrafish are only distantly related and that sfrp1a was a marker gene first, and co-opted gene second? The authors should respond to the comments in the public review related to this aspect, and include more informative comparison and discussion.

A much more nuanced discussion with appropriate comparisons and caveats would be strongly recommended here.

We appreciate this insight and used it as a motivator to complete and add select comparative ISH data to this manuscript. We added in situ hybridization experiments from stickleback fish for craniofacial development genes (*sfrp1a*, *prdm16*, *bmp4*, and *fgf22*; S76-S79). After adding stickleback ISH to the manuscript, we were able to make comparisons between pipefish and stickleback patterns and draw more informed conclusions (L244-L248; L262-277; L444-470). We added additional nuance to the discussion of the head, tooth (L485-489), and male pregnancy (L358-L391) sections to address concerns of study limitations. We describe in more detail these additional data in response to reviewers.

(3) In situ hybridization results: as already included above, there is generally weak labeling of species, developmental stages, and other markings that can provide context. The collective feeling here is that as it is currently presented, the ISH results do not go too far beyond simply illustrative purposes. To take these results further, more detailed comparison may be needed. At a minimum, far better labeling can help avoid making the wrong impression.

Based on the reviewers' comments, we made changes to improve ISH clarity and add select comparative ISH findings. ISH was used to further interpretation of the scRNAseq atlas. All the developmental stages and species information for the embryos used were in the figure

captions as well as in supplemental file 4. Since we primarily used wild caught embryos, we did not have specific ages of most embryos. The technical challenges of acquiring and staging *Syngnathus* embryos are detailed above. Because we did not have their age, we classified them based on developmental markers (such as presence of somites and the extent of craniofacial elongation). Although these classification methods are not ideal, they are consistent with the syngnathid literature (Sommer et al. 2012).

We followed reviewer #1's recommendations by adding an annotated graphic of a pipefish head, aligning bay and Gulf pipefish sequences for the probe regions, expanding out our supplemental figures for ISH into a figure for each probe, and improving labeling. These changes improved the description of the ISH experiments and have increased the quality of the manuscript.

We would have loved to complete detailed comparative studies as suggested, but doing such a complete analysis was not feasible for this study. Therefore, we completed an additional focused analysis. We followed reviewer #3's idea and added ISHs from threespine stickleback, a short snouted fish, for 4 genes (*sfrp1a*, *prdm16*, *fgf22*, and *bmp4*). While more extensive ISHs tracking all marker genes through a variety of developmental stages in pipefish and stickleback would have provided crucial insights, we feel that it is beyond the scope of this study and would require a significant amount of additional work. We, thus, primarily interpreted the ISH results as illustrative data points in our discussion. As we state in the response to reviewer 1, the generation and annotation of the first scRNA-seq atlas in a syngnathid is the primary goal of this manuscript. The ISHs were utilized primarily to support inferred cluster annotations if a positive identification of marker gene expression in expected tissues/cells occurred.

Reviewer #1 (Recommendations For The Authors):

While the scRNA-seq dataset offers a valuable resource for evo-devo analyses in fish and the hypotheses are of interest, critical aspects should be strengthened to support the claims of the study.

Concerning the scRNA-seq dataset, the major points to be addressed are listed below:

- Supplementary file 3 reports the single markers used to validate cluster annotations. To confirm cluster identities, more markers specific to each cluster should be highlighted and presented on the UMAP.

We recognize the reviewer's concern and had in reality used numerous markers to annotate the clusters. Based upon the reviewer's comment we decided to make this clear by creating feature plots for every cluster with the top 6 marker genes. These plots showcase gene specificity in UMAP space. We also added feature plots for zebrafish marker genes for key cell types. Through these changes and the addition of 54 supplementary figures (S3:S57), we hope that it is clear that numerous markers validated cluster identity.

*For example, as clusters 17 and 37 share the same *tnnt1* marker, which other markers permit to differentiate their respective identity.*

This is a fair point. Cluster 17 and 37 both are marked by a *tnnt1* ortholog.

Different paralogous co-orthologs mark each cluster (cluster 17: *LOC125989146*; cluster 37: *LOC125970863*). In our revision to the above comment, additional (6) markers per cluster were highlighted which should remedy this concern.

- L146: the low number of identified cartilaginous cells (only 2% of total connective tissue cells) appears aberrant compared to bone cell number, while Figure 1 presents a

welldeveloped cartilaginous skeleton with poor or no signs of ossification. Please discuss this point.

We also found this to be interesting and added a brief discussion on this subject to the results section (L147-L149). Single cell dissociations can have variable success for certain cell types. It is possible that the cartilaginous cells were more difficult to dissociate than the osteoblast cells.

- L162: pax3a/b are not specific to muscle progenitors as the genes are also expressed in the neural tube and neural crest derivatives during organogenesis. Please confirm cluster 10 identity.

Thank you for the reminder, we added numerous feature plots that explored zebrafish (from Daniocell) and pipefish markers (identified in our dataset). Examining zebrafish satellite muscle markers (*myog*, *pabpc4*, and *jam2a*) shows a strong correspondence with cluster #10.

- L198: please specify in the text the pigment cell cluster number.

We completed this change.

- L199: it is not clear why considering module 38 correlated to cluster 20 while modules 2/24 appear more correlated according to the p-value color code.

We thank the reviewer for pointing this confusing element out! Although the t-statistic value for module 38 (3.75) is lower than the t-statistics for modules 2 and 24 (5.6 and 5.2, respectively), we chose to highlight module 38 for its ‘connectivity dependence’ score. In our connectivity test, we examined whether removing cells from a specific cell cluster reduced the connectivity of a gene network. We found that removing cluster 20 led to a decrease in module 38’s connectivity (-.13, p=0) while it led to an increase in modules 2 and 24’s connectivity (.145, p=1; .145, p=9.14; our original supplemental files 9-10). Therefore, the connectivity analysis showed that module 38’s structure was more dependent on cluster 20 than in comparison with modules 2 and 24. Although you highlighted an interesting quandary, we decided that this is tangential to the paper and did not add this discussion to the manuscript.

- Please describe in the text Figure 4A.

Completed, we thank the reviewer for catching this!

Concerning embryo stainings, the major points to be addressed are listed below:

- Figure 1: please enhance the light/contrast of figures to highlight or show the absence of alcian/alizarin staining. Mineralized structures are hardly detectable in the head and slight differences can be seen between the two samples. The developmental stage should be added. Please homogenize the scale bar format (remove the unit on panels E and, G as the information is already in the text legend). It would be useful to illustrate the data with a schematic view of the structures presented in panels B, and E, and please annotate structures in the other panels.

We thank the reviewer for these suggestions to improve our figure. We increased the brightness and contrast for all our images. We also added an illustration of the head with labels of elements. As discussed, we used wild caught pregnant males and, therefore, do not know the exact age of the specimens. However, we described the developmental stage based

on morphological observations. Slight differences in morphology between samples is expected. We and others have noticed that

developmental rate varies, even within the same brood pouch, for syngnathid embryos. We observed several mineralization zones including in the embryos including the upper and lower jaws, the mes(ethmoid), and the pectoral fin. We recognize the cartilage staining is more apparent than the bone staining, though increasing image brightness and contrast did improve the visibility of the mineralization front.

- All ISH stainings and images presented in Figures 4-6/ Figures S2-3 should be revised according to comments provided in the public review.

We thank the reviewer for providing thorough comments, we provided an in-depth response to the public review. We made several improvements to the manuscript to address their concerns.

- Figure 4: Figure 4B should be described before 4C in the text or inverse panels / L222 the Meckel's cartilage is not shown on Figure 4C. The schematic views in H should be annotated and the color code described / the ISH data must be completed to correlate spatially clusters to head structures.

We thank the reviewer for pointing this out, we fixed the issues with this figure and added annotations to the head schematics.

- Figure 5: typo on panels 'alician' = alcian.

We completed this change.

- Figures S2-3: data must be better presented, polished / typo in captions 'relavant' = relevant.

Thank you for this critique, we created new supplementary figures to enhance interpretation of the data (S59-S71). In these new figures, we included a feature plot for each gene and respective ISHs.

- Figure S3: soat2 = no evidence of muscle marker neither by ISH presented nor in the literature.

We realized this staining was not clear with the previous S2/S3 figures. Our new changes in these supplementary figures based on the reviewer's ideas made these ISH results clearer. We observed soat2 staining in the sternohyoideus muscle (panel B in S71).

Other points:

- The cartilage/bone developmental state (Alcian/alizarin staining) and/or ISH for classical markers of muscle development (such as pax3/myf5) could be used to clarify the This could permit the completion of a comparative analysis between the two species and the interpretation of novel and adaptative characters.

We appreciate this idea! We thought deeply about a well characterized comparative analysis between pipefish and zebrafish for this study. We discussed our concerns in our public response to reviewer 2. We found that it was challenging to stage match all cell types, and were concerned that we could make erroneous conclusions. For example, our pipefish samples were still inside the male brood pouch and possessed yolk sacs. However, we found osteoblast cells in our scRNAseq atlas, and in alizarin staining. Although zebrafish literature

notes that the first zebrafish bone appears at 3 dpf (Kimmel et al. 1995), osteoblasts were not recognized until 5 dpf in two scRNAseq datasets (Fabian et al. 2022; Lange et al. 2023). A 5dpf zebrafish is considered larval and has begun hunting. Therefore, we chose to not integrate our data out of concern that osteoblast development may occur at different timelines between the fishes.

Fabian, P., Tseng, K.-C., Thiruppathy, M., Arata, C., Chen, H.-J., Smeeton, J., Nelson, N., & Crump, J. G. (2022). Lifelong single-cell profiling of cranial neural crest diversification in zebrafish. *Nature Communications* 2022 13:1, 13(1), 1–13.

Lange, M., Granados, A., VijayKumar, S., Bragantini, J., Ancheta, S., Santhosh, S., Borja, M., Kobayashi, H., McGeever, E., Solak, A. C., Yang, B., Zhao, X., Liu, Y., Detweiler, A. M., Paul,

S., Mekonen, H., Lao, T., Banks, R., Kim, Y.-J., ... Royer, L. A. (2023). Zebrahub – Multimodal Zebrafish Developmental Atlas Reveals the State-Transition Dynamics of Late-Vertebrate Pluripotent Axial Progenitors. *BioRxiv*, 2023.03.06.531398.

Kimmel, C., Ballard, S., Kimmel, S., Ullmann, B., Schilling, T. (1995). Stages of Embryonic Development of the Zebrafish. *Developmental Dynamics* 203:253-310.

| *'in situs' in the text should be replaced by 'in situ experiments'.*

We made this change (L395, L663, L666, L762).

| *- Lines 562-565: information on samples should be added at the start of the result section to better apprehend the following scRNA-seq data.*

We thank the reviewer for pointing out this issue. Although we had a few sentences on the samples in the first paragraph of the result section, we understand that it was missing some critical pieces of information. Therefore, we added these additional details to the beginning of the results section (L126-L132).

| *- Lines 629-665: PCR with primers designed on gulf pipefish genome could be performed in parallel on bay and gulf cDNA libraries, and amplification products could be sequenced to analyze alignment and validate the use of gulf pipefish ISH probes in bay pipefish embryos. Probe production could also be performed using gulf primers on bay pipefish cDNA pools.*

After the submission of this manuscript, a bay pipefish genome was prepared by our laboratory. We used this genome to align our probes, these alignments demonstrate strong sequence conservation between the species. We included these alignments in our supplemental files.

| *- L663: the bleaching step must be optimized on pipefish embryos.*

We understand this concern and had completed several bleach optimization experiments prior to publication. Although we found that bleaching improved visibility of staining, we noticed with the probe *tnmd* that bleached embryos did not have complete staining of tendons and ligaments. The unbleached embryos had more extensive staining than the bleached embryos. We were concerned that bleaching would lead to failures to detect expression domains (false negatives) important for our analysis. Therefore, we did not use bleaching with our in situs experiments (except with hatched fish with a high degree of pigmentation).

| *- Indicate the number of specimens analyzed for each labeling condition.*

We thank the reviewer for noticing this issue. We added this information to the methods (L766-767).

| - Describe the fixation and pre-treatment methods previous to ISH and skeleton stainings

We thank the reviewer for pointing out this issue, we added these descriptions (L765-766; L772-774).

Reviewer #3 (Recommendations For The Authors):

(1) *If *sfrp1a* expression is observed also in other fish species with derived craniofacial structures, it's important to discuss this more in the Discussion. This could be a common mechanism to modify craniofacial structures, although functional tests are ultimately required (but not in this paper, for sure). Can lines 421-428 involve the statement "a prolonged period of chondrocyte differentiation" underlies craniofacial diversity?*

This is a great idea, and we added a sentence that captures this ethos (L451-452).

(2) *Lines 334-346 need to be rephrased. It's hard to understand which genes are expressed or not in pipefish and zebrafish. Did "23 endocytosis genes" show significant enrichment in zebrafish epidermis, or are they expressed in zebrafish epidermis?*

We thank the reviewer for this comment, we re-phrased this section for clarity (L365-368).

| (3) *Figure 4 is missing the "D" panel and two "E" panels.*

We thank the reviewer for noticing this, we fixed this figure.

| (4) *Line 302: "whole-mount" or "whole mount"*

We thank the reviewer for the catch!

<https://doi.org/10.7554/eLife.97764.2.sa0>