Single cell RNA sequencing of the *Strongylocentrotus purpuratus* larva reveals the blueprint of major cell types and nervous system of a non-chordate deuterostome

Periklis Paganos[1], Danila Voronov[1], Jacob Musser[2], Detlev Arendt[2] and Maria I. Arnone[1*]

[1]Stazione Zoologica Anton Dohrn, Department of Biology and Evolution of Marine Organisms, Villa Comunale, 80121, Naples, Italy

[2]European Molecular Biology Laboratory, Developmental Biology Unit, Meyerhofstrasse 1, 69117 Heidelberg, Germany

*corresponding author: miarnone@szn.it

Abstract

Identifying the molecular fingerprint of organismal cell types is key for understanding their function and evolution. Here, we use single cell RNA sequencing (scRNA-seq) to survey the cell types of the sea urchin early pluteus larva, representing an important developmental transition from non-feeding to feeding larva. We identify 21 distinct cell clusters, representing cells of the digestive, skeletal, immune, and nervous systems. Further subclustering of these reveal a highly detailed portrait of cell diversity across the larva, including the identification of neuronal cell types. We then validate important gene regulatory networks driving sea urchin development and reveal new domains of activity within the larval body. Focusing on neurons that co-express *Pdx-1* and *Brn1/2/4*, we identify an unprecedented number of genes shared by this population of neurons in sea urchin and vertebrate endocrine pancreatic cells. Using differential expression results from Pdx-1 knockdown experiments, we show that Pdx1 is necessary for the acquisition of the neuronal identity of these cells. We hypothesize that a network similar to the one orchestrated by Pdx1 in the sea urchin neurons was active in an ancestral cell type and then inherited by neuronal and pancreatic developmental lineages in sea urchins and vertebrates.

Introduction

Multicellular organisms consist of numerous cell types, specialized in performing different tasks that guide growth and survival. During embryonic development, cells go through rounds of proliferation, specification and differentiation into cell types with distinct functions. The information for this developmental diversification lies in the genome and the spatio-temporal expression of regulatory genes that specifies the molecular fingerprint of a given cell type (Fu et al., 2017).

The identity of each cell type is established, controlled and maintained by distinct Gene Regulatory Networks (GRNs). GRNs are logical maps of the regulatory inputs and outputs active in a cell at a given place and time, and are enacted by transcription factors, signaling molecules and terminal differentiation genes (Davidson et al., 2003, Davidson

and Erwin, 2006). GRNs have been studied in a variety of organisms ranging from plants to animals in order to analyze the gene interactions at a specific time and place during the life of an organism (Krouk et al., 2013), and have been used for understanding the relationship between genome, development and evolution (Davidson and Erwin, 2006). Therefore, understanding the genetic mechanisms that provide cell types with a specific identity, and the conservation of this identity across animal taxa, is essential for understanding cell type function and evolutionary history (Arendt, 2008, Arnone et al., 2016).

Until recently, most approaches for comparing cell types relied on the description of distinct morphological features, that is linked to the functionality of a given cell type, the identification of molecular markers, perturbation of gene expression and fate mapping. However, technological advances in microfluidics and nucleic acid barcoding now allow the high-through-put recognition of an organism's cell types at a single cell level. In particular, single cell RNA sequencing (scRNA-seq) technology, developed during the last decade, is a powerful method used to unravel the transcriptional content of individual cells, resulting in the identification of distinct cell types in an unbiased manner (Tang et al., 2009, Klein et al., 2015). ScRNA-seq involves dissociation of an organism, organ or tissue into single cells, isolation and capture of the single cells into droplets, specific barcoding of individual mRNAs, and sequencing of transcriptomic content of each cell. Computational analysis can then identify putative cell type families by clustering cells with similar transcriptional profiles. Further analysis of such cell clusters can lead to the identification of distinct cell types.

Within deuterostomes, echinoderms are a member of the phylogenetic sister group to chordates, making them an ideal model for understanding the origin and diversification of deuterostome and chordate cell types (Arnone et al., 2015). Sea urchin embryos and larvae have also been extensively used to unravel the general mechanisms of cell type specification and differentiation during development (Cameron and Davidson, 1991, Davidson et al., 1998, McClay, 2011, Lyons et al., 2012, McClay et al., 2020). The main reason for this lies in the ease with which different cell types and biological processes can be observed in the optically transparent embryos and larvae. Among the most well-studied sea urchin cell types are those comprising the nervous (Bisgrove and Burke, 1987, Burke et al., 2006a, McClay et al., 2018), immune (Rast et al., 2006, Ho et al., 2017) and digestive systems (Annunziata et al., 2014, Annunziata and Arnone, 2014, Perillo and Arnone, 2014, Perillo et al., 2016), and of both musculature (Andrikou et al., 2013, Andrikou et al., 2015) and skeleton (Okazaki, 1965, Duloquin et al., 2007, Rafiq et al., 2012, Sun and Ettensohn, 2017). For these, the developmental origins and gene regulatory wiring has been described in great detail, making the sea urchin an ideal model for GRN comparative analyses in development and evolution (Cary et al., 2020).

Here we take advantage of the detailed characterization of the sea urchin cell types performed over the years, the available cell type specific molecular markers, and the ease with which the sea urchin larvae are dissociated into single cells, to perform scRNA-seq and generate a comprehensive atlas of sea urchin larval cell type families. Our findings suggest that the larva consists of 21 genetically-distinct cell clusters, representing distinct cell type families (Shekhar and Menon, 2019), which we validate using fluorescent *in situ* hybridization (FISH) and immunohistochemistry (IHC). Based on previous studies tracing

87 developmental lineage, we assign cell type families to the specific embryonic germ layers
88 they derived from. Comparing the transcription factor content of the grouped cell clusters
89 reveals that most transcription factors are expressed pleiotropically independently of their
90 germ layer origins, yet tend to be cell type family-specific within a germ layer. In addition,
91 we illustrate how single-cell data complement and validate previously studied GRNs, and
92 also reveal novel cellular domains where these GRNs are likely also activated. Lastly, we
93 investigate neuronal diversity in the sea urchin larva, identifying 12 distinct neuronal cell
94 types. Among these, we recover a unique neurosecretory type controlled by *Sp-Pdx1* and
95 *Sp-Brn1/2/4* exhibiting a pancreatic-like gene expression signature (Perillo et al., 2018).
96 Our results confirm and extend this pancreatic-like signature, suggesting that an ancestral
97 neuron in early deuterostomes may have given rise to the endocrine cells in the vertebrate
98 pancreas. Supporting this, knockdown of *Sp-Pdx1* shows it is necessary for differentiation
99 of this pancreatic-like neuronal endocrine population, indicating it has an evolutionary
100 conserved role as a mediator of endocrine fate.

101

102 ## Results

103 ### Building a cell type atlas of the sea urchin larva with single-cell transcriptomics

104 Sea urchin early pluteus larvae were cultured and collected at 3 dpf. We performed single-
105 cell RNA sequencing on six samples from four independent biological replicates.
106 Individual samples were dissociated into single cells using a gentle enzyme-free
107 dissociation protocol and using the 10x Chromium scRNA-seq system (Figure 1A). In
108 total, transcriptomes from 19,699 cells were included in the final analysis. To identify sea
109 urchin larval cell types, we used Louvain graph clustering as implemented in the Seurat
110 pipeline (see methods). This revealed 21 genetically-distinct cell clusters (Figure 1B,
111 supplement 1A and B), each representing an individual cell type or set of closely related
112 cell types in the early pluteus larva.

113 Next, we set out to explore the identity of our initial 21 cell clusters. We first assigned
114 preliminary identities to each cluster based on the expression of previously described cell
115 type markers, benefiting from the unique and rich knowledge on sea urchin
116 developmental lineages: ciliary band (*Btub2*) (Harlow and Nemer, 1987), apical plate
117 (*Hbn*) (Burke et al., 2006a), aboral ectoderm (*Spec2a*) (Yuh et al., 2001), lower oral
118 ectoderm (*Bra*) (Wei et al., 2012), upper oral ectoderm (*Gsc*) (Wei et al., 2012), neurons
119 (*SynB*) (Burke et al., 2006a), esophageal muscles (*Mhc*) (Andrikou et al., 2013), coelomic
120 pouches (*Nan2*) (Juliano et al., 2010), blastocoelar cells (185/333) (Ho et al., 2017),
121 immune cells (*Gcm*) (Materna et al., 2013), skeleton (*Msp130*) (Harkey et al., 1992), anus
122 (*Hox11/13b*), intestine (*Cdx*), pyloric sphincter (*Pdx-1*), different stomach domains (*Chp,*
123 *ManrC1a, Endo16*) (Annunziata and Arnone, 2014), exocrine pancreas-like domain
124 (*Ptf1a*) (Perillo et al., 2016), cardiac sphincter (*Trop1*) (Yaguchi et al., 2017) and
125 esophagus (*Brn1/2/4*) (Cole and Arnone, 2009). Further, we grouped the cell clusters
126 according to their embryonic germ layer origin (Figure 1C) using knowledge from previous
127 lineage tracing experiments (Angerer and Davidson, 1984, Cameron et al., 1987).

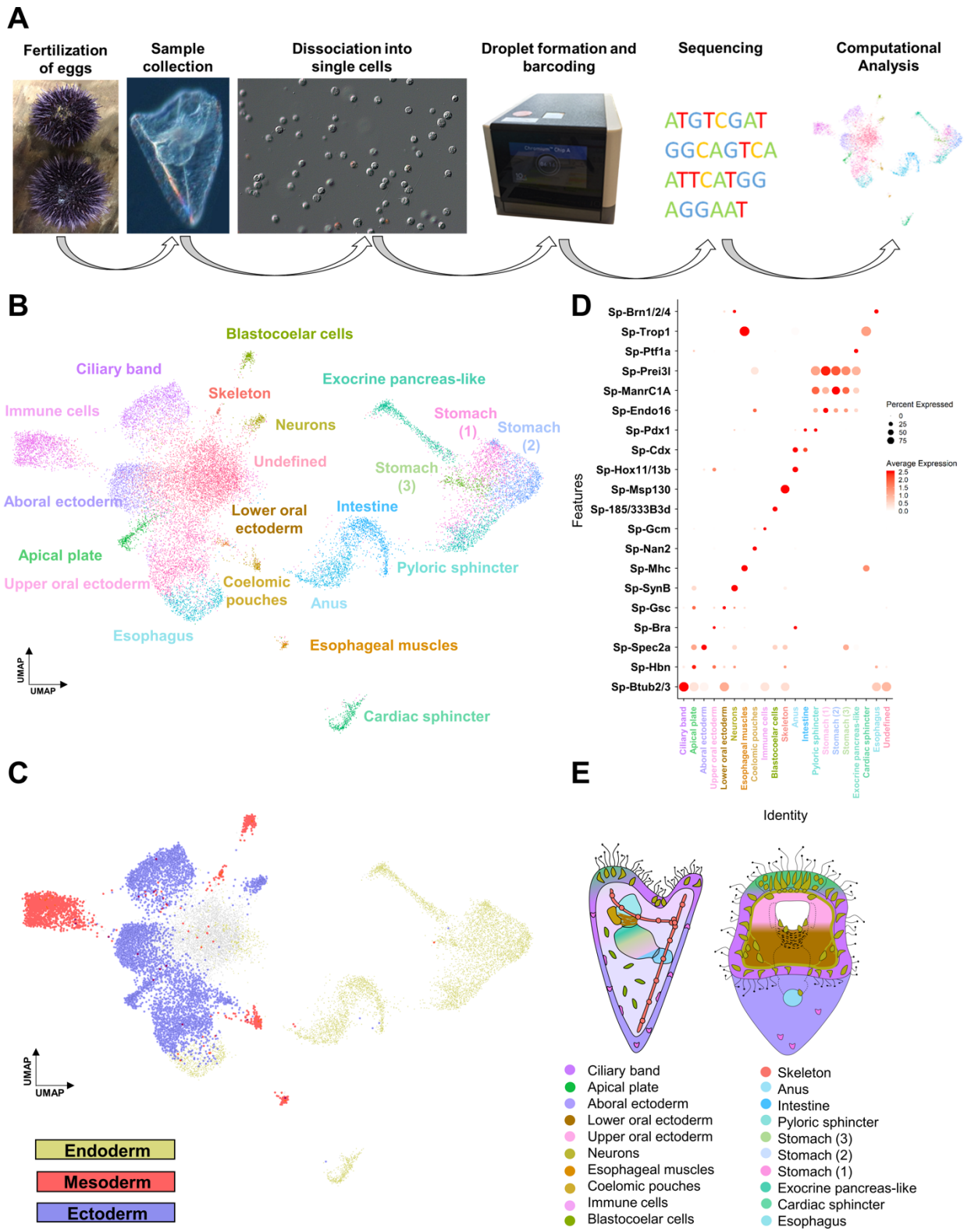


Figure 1. Cell type family atlas of the 3 dpf *S. purpuratus* larva. A) Single-cell RNA sequencing pipeline from gamete fertilization to computational analysis. **B)** UMAP showing 3 dpf larval cells colored by their assignment to the initial set of 21 distinct cell clusters. **C)** UMAP with cells colored by germ layer of origin: endoderm (yellow), mesoderm (red), and ectoderm (blue).

D) Dotplot of gene markers specific to cell clusters. **E)** Illustration depicting location of cell type families on different larval domains. Color-code is the same as in Figure 1B.

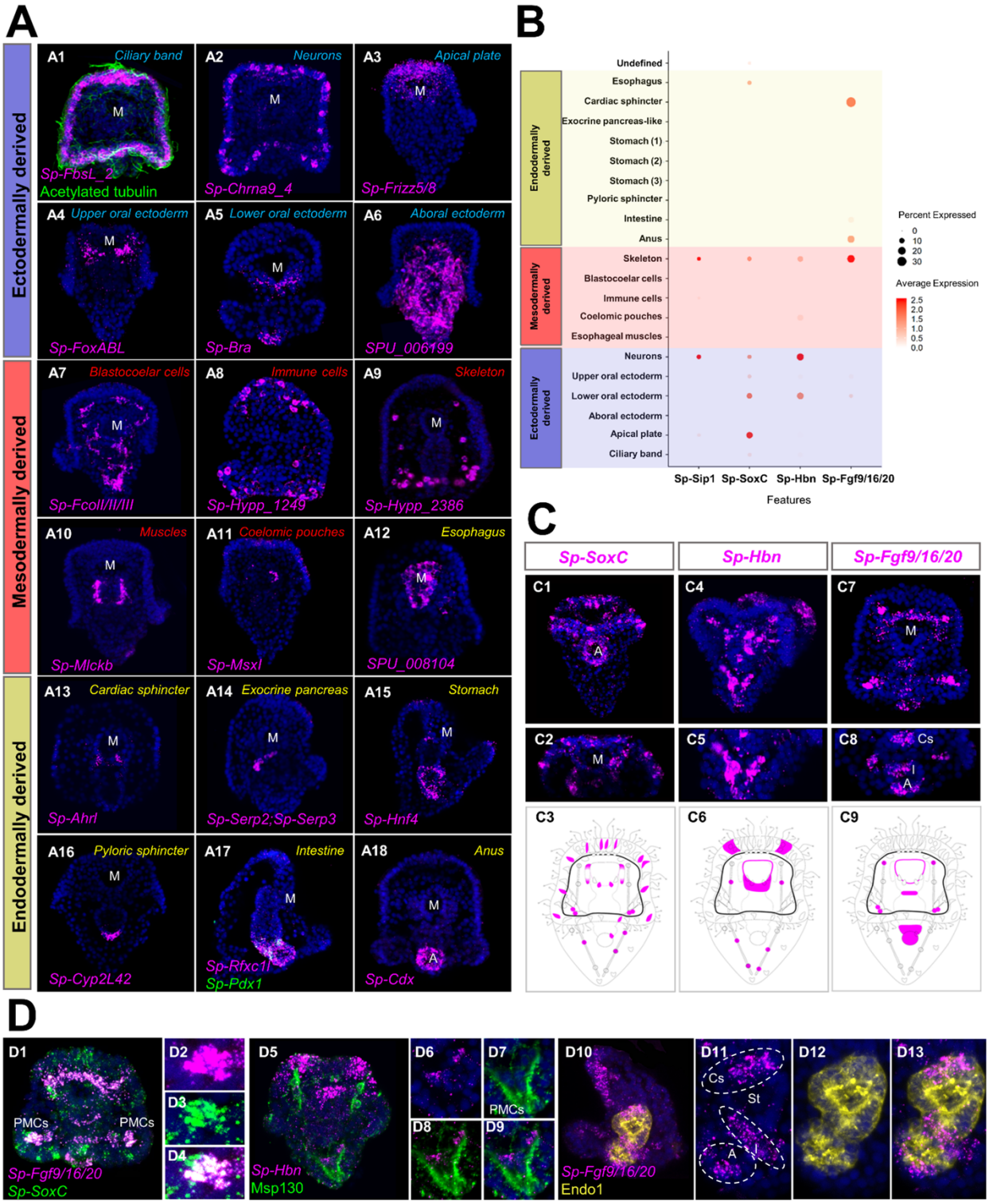


Figure 2. Validation of scRNA-seq predictions and novel expression domains. A) FISH of *S. purpuratus* 3 dpf larvae with antisense probes for *Sp-FbsL_2* (A1), *Sp-Chrna9_4* (A2), *Sp-Frizz5/8* (A3), *Sp-FoxABL* (A4), *Sp-Bra* (A5), *SPU_006199* (A6), *Sp-FcoII/II/III* (A7), *Sp-*

Hypp_1249 (A8), *Sp-Hypp_2386* (A9), *Sp-Mlckb* (A10), *Sp-MsxI* (A11), *SPU_008104* (A12), *Sp-AhrI* (A13), *Sp-Serp2; Sp-Serp3* (A14), *Sp-Hnf4* (A15), *Sp-Cyp2L42* (A16), *Sp-Rfxc1I* (A17), *Sp-Pdx1* (A17) and *Sp-Cdx* (A18). Color-code indicates germ layer embryonic origin: endoderm (yellow), mesoderm (red), ectoderm (blue). Immunofluorescent detection of acetylated tubulin in ciliary band (green). **B)** Dotplot of *Sp-Sip1*, *Sp-SoxC*, *Sp-Hbn* and *Sp-Fgf9/16/20* expression showing previously described and novel expression domains. **C)** FISH of *S. purpuratus* 3 dpf larvae with antisense probes for *Sp-SoxC* (C1-C2), *Sp-Hbn* (C4-C5) and *Sp-Fgf9/16/20* (C7-C8). Illustrations depicting all the expression domains of *Sp-SoxC* (C3), *Sp-Hbn* (C6) and *Sp-Fgf9/16/20* (C9). Illustrations depicting all the expression domains of *Sp-SoxC* (C3), *Sp-Hbn* (C6) and *Sp-Fgf9/16/20* (C9). Nuclei are labelled with DAPI (in blue). All images are stacks of merged confocal Z sections. **D)** FISH of *S. purpuratus* 3 dpf larvae with antisense probes for *Sp-SoxC* and Sp-Fgf9/16/20 (D1-D4), for *Sp-Hbn* combined with immunohistochemical detection for the skeletal cells marker Msp130 (D5-D9) and *Sp-Fgf9/16/20* with immunohistochemical detection for the midgut and hindgut protein Endo1 (D10-D13). A, Anus; Cs, Cardiac sphincter; I, Intestine; M, Mouth; PMCs, Primary mesenchyme cells; St, Stomach.

To validate these identities, we identified all genes expressed in each cell type, totaling in 15,578 WHL genes (transcriptome models) and 12,924 genes (SPU gene models) (Supplementary file 1), and performed *in situ* hybridization on a selected set of these together in combination with previously described markers. Based on this, we mapped 20 of the 21 clusters to distinct larval domains and confirmed their identity (Figure 1E). Notably, the resulting expression patterns validated the initial predictions (Figure 2- Figure supplement 2), verifying the high quality of the single-cell dataset. Importantly, this approach identified various new markers for each cell type family, including *Sp-FbsL_2* (ciliary band; Figure 2A1*)*, *Sp-hypp_2386* (skeletal cells; Figure 2A9), and *Sp-Serp2/3* (exocrine pancreas-like cells; Figure 2A14). The 21st cluster, which had a poorly-defined molecular signature as judged by the low number and expression level of total and marker genes combined with the lack of specific localization, likely represents not fully differentiated cells (Figure 1- Figure supplement D & F) and we refer to this cluster as undefined.

Our scRNA-seq analysis and *in situ* hybridization protocol unraveled novel expression domains for several previously described cell type markers. For instance, the transcription factors *Sp-SoxC* and *Sp-Hbn*, previously described in early neuronal specification (Garner et al., 2016, Wei et al., 2016, Yaguchi et al., 2016), were predicted by our scRNA-seq analysis to also be expressed in skeletal cells (Figure 2B &,). Likewise, the FGF signaling ligand, *Sp-Fgf9/16/20*, is known to be involved in skeletal formation and is expressed in specific populations of PMCs (Adomako-Ankomah and Ettensohn, 2014). ScRNA-seq indicates it is also expressed in oral ectoderm, cardiac sphincter, intestine and anus (Figure 2B). Using *in situ* hybridization and immunohistochemical detection methods we confirmed the predictions and we found *Sp-SoxC* expressed in *Sp-Fgf9/16/20* positive PMCs and *Sp-Hbn* expressed in skeletal cells located in the vertex (Figures 2C1-3, 2C4-6 2D1-4 & 2D5-9). Moreover, we found *Sp-Fgf9/16/20* transcripts localized in distinct gut domains corresponding to the cardiac sphincter, intestinal and anal regions (Figure 2C7-9, 2D10-13).

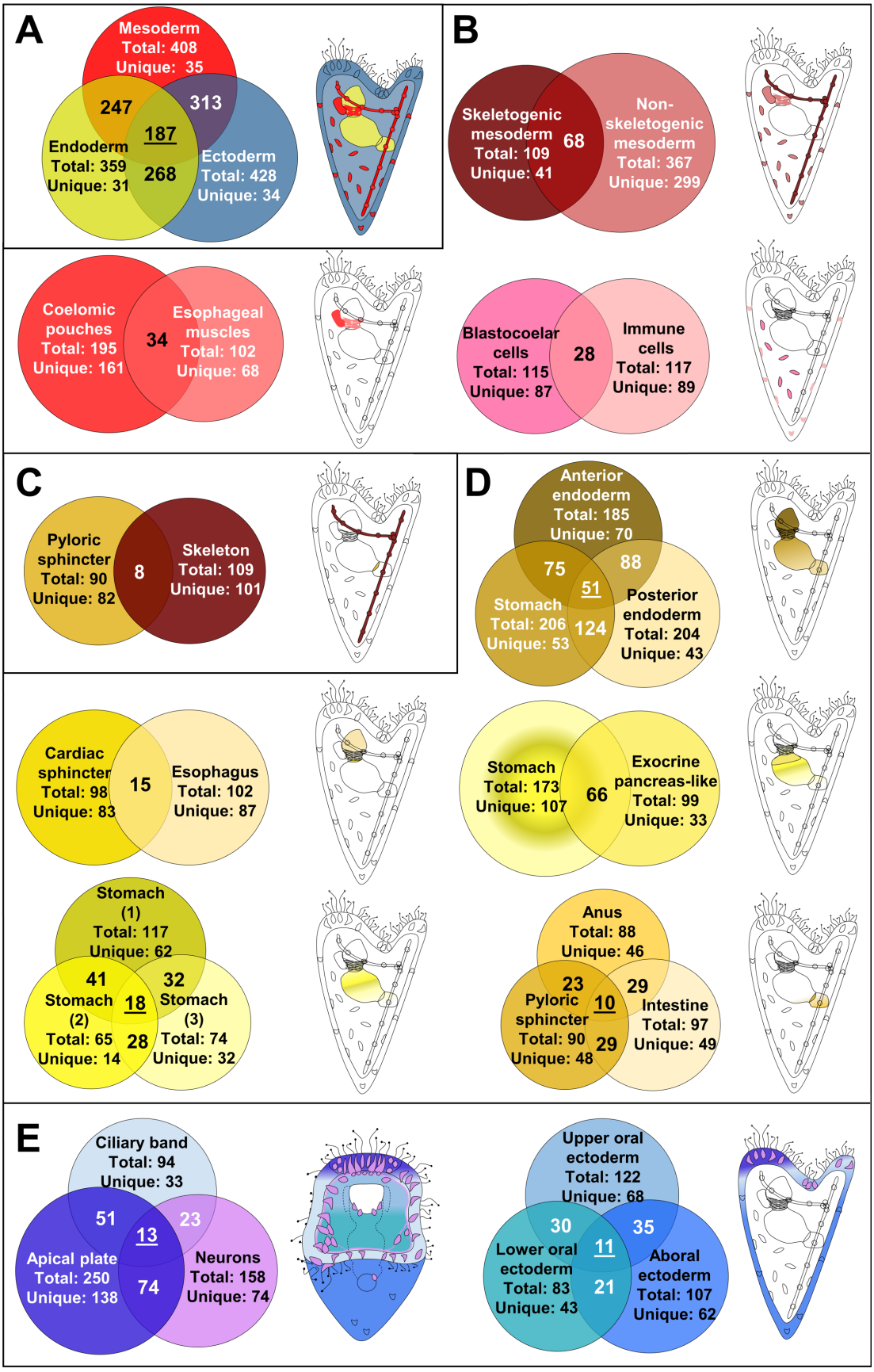
We compared the limits of detection by *in situ* hybridization versus single cell RNA sequencing, using the coelomic pouch cell cluster as a case study. The coelomic pouches are derived from the mesoderm and the left coelomic pouch contributes to the formation

of the rudiment and juvenile sea urchin after metamorphosis (Strathmann, 1987, Smith et al., 2008). The formation of the coelomic pouches is complex, and includes contributions from the small micromeres, and a mesodermal cell population that is set aside during early development (Pehrson and Cohen, 1986, Strathmann, 1987). Previous attempts to characterize this population had involved screening of genes active in germline determination and maintenance in other species and revealed that, while some germ line-specific transcripts and proteins were found exclusively expressed in the small micromeres and the coelomic pouch of the sea urchin embryo (Juliano et al., 2006), the majority of the genes tested by *in situ* hybridization were not enriched in this cell type. Interestingly, plotting the Juliano and coauthors' gene list, alongside previously described coelomic pouch specific genes (Luo and Su, 2012, Martik and McClay, 2015), they were all found in our analysis to be expressed in the same cell cluster (Figure 2- Figure supplement 3). This suggests a higher sensitivity of single cell RNA sequencing compared to the *in situ* hybridization, adding crucial missing information on the molecular fingerprint of a complex cell type.

Lastly, to determine which cells in the larva were undergoing active proliferation, we plotted expression of cell division markers in sea urchin, including *pcna*, DNA polymerases, DNA ligases, condensins, and centromere proteins (Perillo et al., 2020). The majority of cell proliferation genes were found to be enriched in the ciliary band, apical plate, coelomic pouch, immune and skeletal cell clusters (Figure 2- Figure supplement 4A). We also observed Cdk genes enriched in several endodermally derived cell populations (Figure 2-supplement 4A). Validating this, we observed S-phase cells in ciliary band, apical plate, oral ectoderm, endodermal and skeletal cells using Edu pulse labelling (Figure 2- Figure supplement 4B). In contrast, we did not observe Edu fluorescence in cell clusters that lacked expression of proliferation markers, such as the aboral ectoderm, suggesting that indeed they may not be proliferating.

Shared lineage information of the larval cell type families

We next compared transcription factor profiles of early larval cell type families. In general, cell populations derived from the same germ layer shared more factors with each other than with cell populations from other germ layers (Figure 3, Figure 3-Figure supplement 1), consistent with the finding that cell type expression programs often retain information about their developmental lineage (Sladitschek et al., 2020). Further analysis however revealed very few transcription factors specific to cells from a single germ layer, contrasting with many regulators expressed in derivatives of more than one germ layer (Figure 3A), and nearly one third (n = 187) shared by cell type families from all three layers. Unexpectedly, mesodermal cell populations share expression of more transcription factors with ectodermal than with endodermal cell populations, even though they are more closely linked to endodermal lineage in development. We also noted that cell type families derived from the same germ layer share up to one third of transcription factors, while the majority are cell type-specific (Figure 3B, 3D & 3E). In general, neighboring cell populations and those with common developmental origins share a larger number of TFs (Figure 3D), compared to cell populations with different developmental histories (Figure 3C).


A
Mesoderm
Total: 408
Unique: 35
247
313
187
Endoderm
Total: 359
Unique: 31
268
Ectoderm
Total: 428
Unique: 34
B
Skeletogenic mesoderm
Total: 109
Unique: 41
68
Non-skeletogenic mesoderm
Total: 367
Unique: 299
Coelomic pouches
Total: 195
Unique: 161
34
Esophageal muscles
Total: 102
Unique: 68
Blastocoelar cells
Total: 115
Unique: 87
28
Immune cells
Total: 117
Unique: 89
C
Pyloric sphincter
Total: 90
Unique: 82
8
Skeleton
Total: 109
Unique: 101
D
Anterior endoderm
Total: 185
Unique: 70
75
88
51
Stomach
Total: 206
Unique: 53
124
Posterior endoderm
Total: 204
Unique: 43
Cardiac sphincter
Total: 98
Unique: 83
15
Esophagus
Total: 102
Unique: 87
Stomach
Total: 173
Unique: 107
66
Exocrine pancreas-like
Total: 99
Unique: 33
Stomach (1)
Total: 117
Unique: 62
41
32
Stomach (2)
Total: 65
Unique: 14
18
28
Stomach (3)
Total: 74
Unique: 32
Anus
Total: 88
Unique: 46
23
29
Pyloric sphincter
Total: 90
Unique: 48
10
29
Intestine
Total: 97
Unique: 49
E
Ciliary band
Total: 94
Unique: 33
51
23
13
Apical plate
Total: 250
Unique: 138
74
Neurons
Total: 158
Unique: 74
Upper oral ectoderm
Total: 122
Unique: 68
30
35
11
Lower oral ectoderm
Total: 83
Unique: 43
21
Aboral ectoderm
Total: 107
Unique: 62

228

Figure 3. Regulatory states of the 3 dpf *S. purpuratus* larva. A) Comparison of the transcription factor content per germ layer. Venn diagram showing the shared and unique transcription factors per germ layer. Ectodermally derived cell type families are shown in blue, mesodermally derived in red, and endodermally derived in yellow. **B)** Comparison of the transcription factor content across mesodermal lineages and cell type families. Venn diagram showing the shared and unique transcription factors per comparison. **C)** Transcription factor content comparison of pyloric sphincter (endodermally derived) and skeletal cells (mesodermally derived), used as a negative control of our comparison. **D)** Comparison of the transcription factor content per endodermal lineage and endodermally derived cell type families. Venn diagram showing the shared and unique transcription factors per comparison. **E)** TF signature comparison of ectodermally derived cell type families. Venn diagram showing the shared and unique transcription factors per comparison. Cartoons indicated the relative position of each cell type family/lineage. Mesodermal cell type families/lineages are shown in shades of red, endodermal ones in shades of yellow and endodermal ones in shades of blue.

To further characterize the regulatory profile of larval cell type families we set out to identify the expression profiles of members of major transcription factor families (Figure 4 and Figure 4 supplement 1). In *S. purpuratus*, most homeobox transcription factors were previously found expressed at the gastrula stage (2 dpf), with several members expressed in domains derived from all three germ layers (Howard-Ashby et al., 2006). Our single cell analysis, although at a later developmental time point, supports these findings, and further refines our understanding of their expression to specific cell type families. In the early pluteus larva, most homeobox class transcription factors are enriched in ectodermally derived cell populations, such as the apical plate and neurons. In contrast, ANTP Class and HNF class transcription factors are enriched in endodermal derivatives (Figure 4). Other major transcription factor families, such as the Forkhead, Ets, and Zinc-finger families, members of which are expressed throughout sea urchin embryogenesis (Tu et al., 2006, Rizzo et al., 2006, Materna et al., 2006), are also expressed across a spectrum of cell populations. Forkhead and zinc-finger transcription factors are highly expressed in specific cell type families of all three germ layer derivatives, whereas Ets family TFs are enriched in ectodermal and mesodermal derivatives (Figure 4).

The active regulatory state of a given cell type family is an immediate consequence of the gene regulatory network active at this time point. Previous research in sea urchin has described in detail many regulatory networks active during embryonic and larval development. Our scRNA-seq data is broadly consistent with previous studies, yet also identifies new domains and cell populations in which these regulatory networks may be active. For instance, we plotted all transcription factors known to be active in specifying coelomic pouch cells (Martik and McClay, 2015). Our data confirmed their co-expression in coelomic pouch, but also revealed their co-expression in the apical plate (Figure 5A). Similarly, plotting genes involved in the aboral ectoderm gene regulatory network (Ben-Tabou de-Leon et al., 2013), we found all genes in both the aboral ectoderm cluster as well as in the apical plate cells (Figure 5B). On the other hand, plotting members of the pre-gastrula skeletogenic mesoderm regulatory network (Figure 5C) revealed most were still active in the pluteus larva and specific to skeletal cells (Oliveri et al., 2008, Shashikant and Ettensohn, 2019, Khor et al., 2019, Ettensohn, 2020).

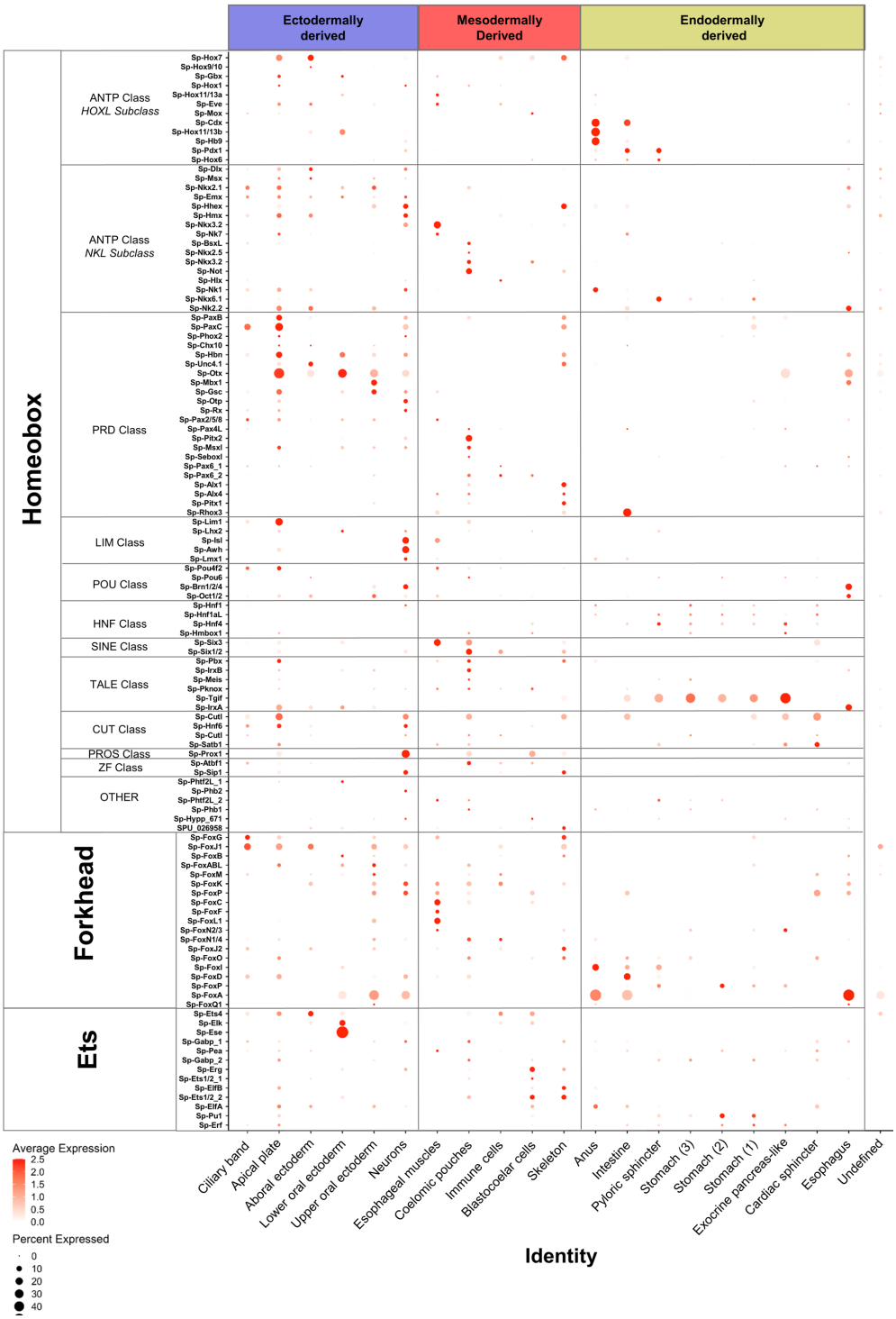
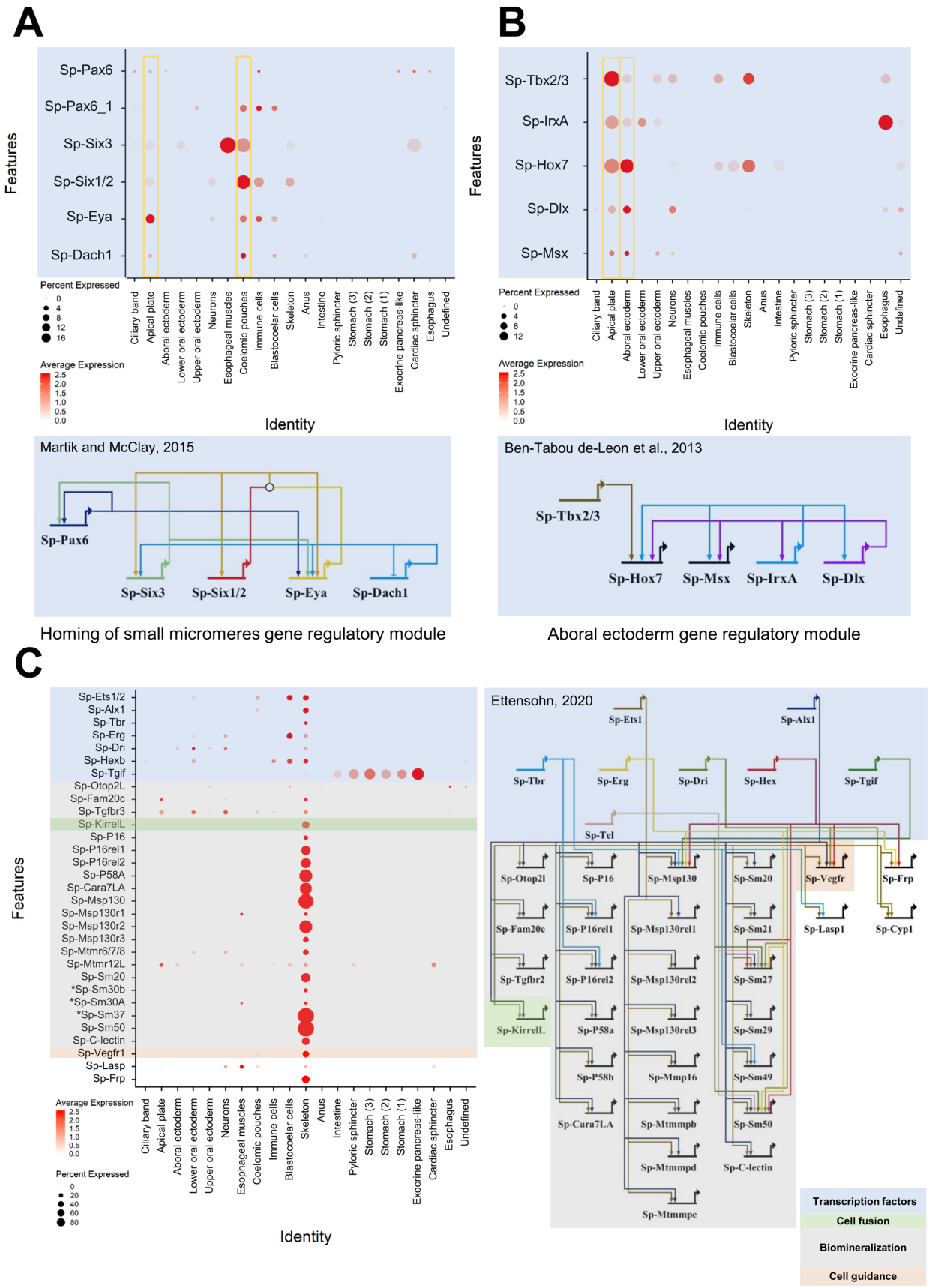


Figure 4. Localization of major transcription factor family members. Dotplot showing the average scaled expression of members of the Homeobox, Forkhead and Ets transcription factor families. The developmental origins of each cell type family are shown in blue for ectodermally derived, red for mesodermally derived and yellow for endodermally derived ones.


A
Features
Sp-Pax6
Sp-Pax6_1
Sp-Six3
Sp-Six1/2
Sp-Eya
Sp-Dach1
Percent Expressed
Average Expression
Ciliary band
Apical plate
Aboral ectoderm
Lower oral ectoderm
Upper oral ectoderm
Neurons
Esophageal muscles
Coelomic pouches
Immune cells
Blastocoelar cells
Skeleton
Anus
Intestine
Pyloric sphincter
Stomach (3)
Stomach (2)
Stomach (1)
Exocrine pancreas-like
Cardiac sphincter
Esophagus
Undefined
Identity
Martik and McClay, 2015
Sp-Pax6
Sp-Six3
Sp-Six1/2
Sp-Eya
Sp-Dach1
Homing of small micromeres gene regulatory module
B
Sp-Tbx2/3
Sp-IrxA
Sp-Hox7
Sp-Dlx
Sp-Msx
Ben-Tabou de-Leon et al., 2013
Sp-Tbx2/3
Sp-Hox7
Sp-Msx
Sp-IrxA
Sp-Dlx
Aboral ectoderm gene regulatory module
C
Sp-Ets1/2
Sp-Alx1
Sp-Tbr
Sp-Erg
Sp-Dri
Sp-Hexb
Sp-Tgif
Sp-Otop2L
Sp-Fam20c
Sp-Tgfbr3
Sp-KirrelL
Sp-P16
Sp-P16rel1
Sp-P16rel2
Sp-P58A
Sp-Cara7LA
Sp-Msp130
Sp-Msp130r1
Sp-Msp130r2
Sp-Msp130r3
Sp-Mtmr6/7/8
Sp-Mtmr12L
Sp-Sm20
*Sp-Sm30b
*Sp-Sm30A
*Sp-Sm37
Sp-Sm50
Sp-C-lectin
Sp-Vegfr1
Sp-Lasp
Sp-Frp
Ettensohn, 2020
Sp-Ets1
Sp-Alx1
Sp-Tbr
Sp-Erg
Sp-Dri
Sp-Hex
Sp-Tgif
Sp-Tel
Sp-Otop2l
Sp-P16
Sp-Msp130
Sp-Sm20
Sp-Vegfr
Sp-Frp
Sp-Fam20c
Sp-P16rel1
Sp-Msp130rel1
Sp-Sm21
Sp-Lasp1
Sp-Cyp1
Sp-Tgfbr2
Sp-P16rel2
Sp-Msp130rel2
Sp-Sm27
Sp-KirrelL
Sp-P58a
Sp-Msp130rel3
Sp-Sm29
Sp-P58b
Sp-Mmp16
Sp-Sm49
Sp-Cara7LA
Sp-Mtmmpb
Sp-Sm50
Sp-Mtmmpd
Sp-C-lectin
Sp-Mtmmpe
Transcription factors
Cell fusion
Biomineralization
Cell guidance

280

Figure 5. Validation of preexisting GRNs and putative novel function of specific gene regulatory modules. A) Dotplot showing the mRNA localization of genes involved in the homing of small micromeres to the coelomic pouch and novel apical plate domain. **B)** Dotplot of aboral ectoderm regulatory module genes showing novel apical plate expression. **C)** Pre-gastrula gene regulatory network enriched in skeletal cells of the sea urchin pluteus larva. Asterisks indicate larval genes involved in biomineralization, putative members of this GRN.

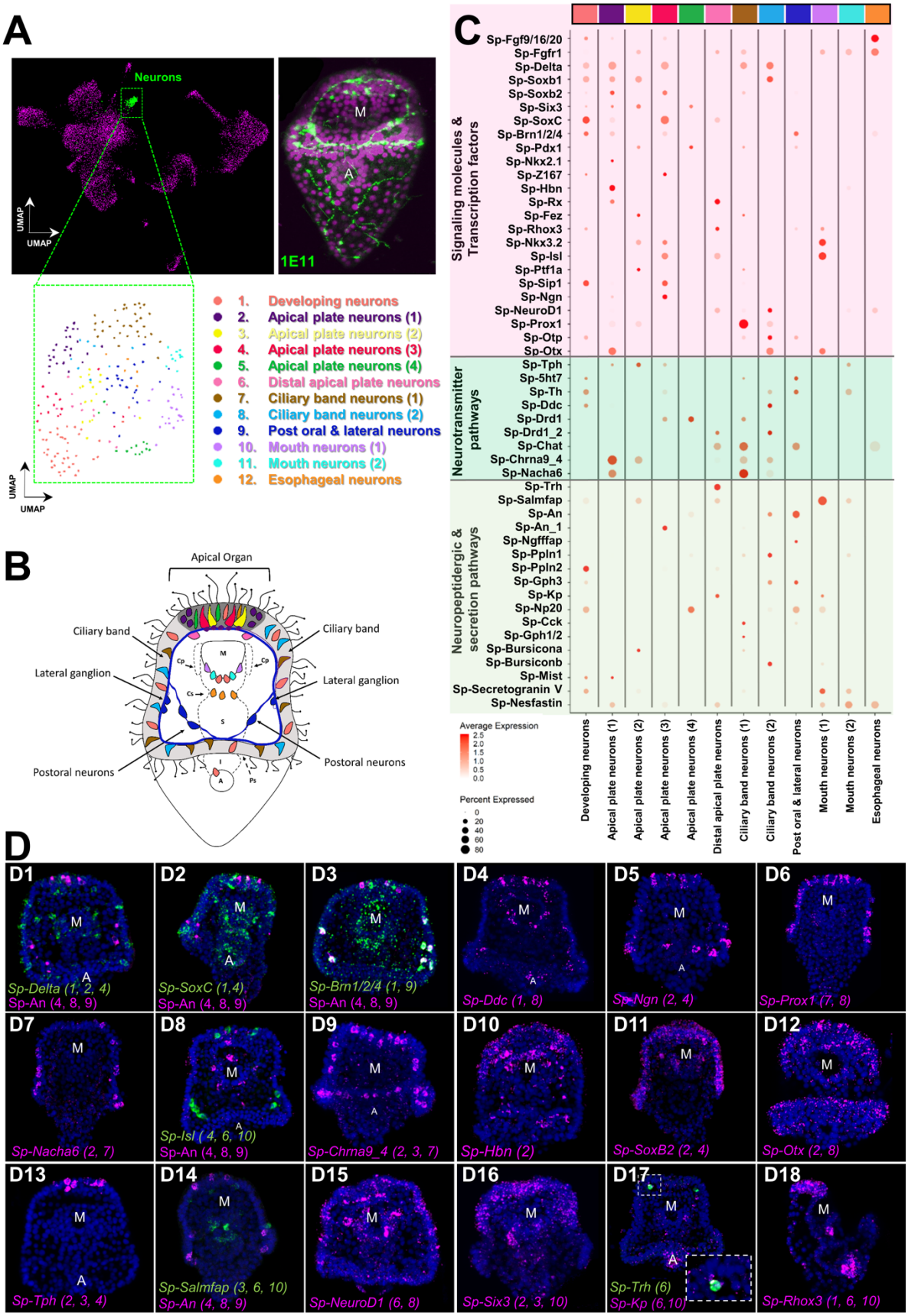
These findings illustrate the immediate benefit of our dataset to drastically expand our knowledge of larval regulatory networks. Finally, our scRNA-seq recreates a nearly identical 3 dpf endoderm expression pattern atlas as that published previously by our group using more traditional methods (Annunziata et al., 2014), providing additional information on each gene's average expression level and the percentage of cells expressing each marker (Figure 5- Figure supplement 1).

Unravelling the neuronal diversity and molecular signature of the nervous system

The classification of neuronal cell type diversity is an important step for unravelling the evolution and function of the nervous system. The sea urchin free swimming larva is equipped with a nervous system consisting of interconnected ganglia (Burke et al., 2006a) that allows the animal to respond to environmental stimuli and coordinate its swimming behavior (Soliman, 1983, Katow et al., 2010). Several neuronal types, including apical and ciliary band neurons, as well as neurons along the digestive tube, have been previously identified and their specification described in detail (Burke et al., 2006a, Burke et al., 2006b, Wei et al., 2009, Wei et al., 2011, Garner et al., 2016, Wei et al., 2016, McClay et al., 2018, Perillo et al., 2018, Wood et al., 2018).

Our initial clustering analysis resolved single clusters for neuronal cells, as well as for PMCs and immune cells. However, expression of known markers suggested the presence of distinct subclusters in each of these cell type family groups. In order to investigate this, we independently performed subclustering and re-analysis of the neuronal, immune, and PMC cells. Subclustering of each of these initial major clusters revealed 12 neuronal, 8 immune, and 5 PMC subclusters, each likely representing distinct cell types (Figure 6A, Figure 6- Figure Supplement 1 & 2). Two of the immune subclusters expressed polyketide synthase 1 (*Sp-Pks1*), suggesting these represent sea urchin pigment cell populations (Calestani and Rogers, 2010). We also found a subcluster of immune cells that expresses the membrane attack complex/perforin family gene (*Sp-MacpfA2*), suggesting this corresponds to immune system globular cells (Figure 6-supplement 2). Notably, our finding of 5 PMC subclusters corroborates previous reports showing five distinct groups of PMC cells along the syncytium (Sun and Ettensohn, 2014) (Figure 6- Figure Supplement 1).

To annotate the 12 neuronal cell types revealed via subclustering we took advantage of the extensive previous work investigating neurogenesis and neuronal differentiation in sea urchin. Plotting neuronal markers, we resolved unique molecular signatures for each subcluster and assigned each a putative identity and location in the larva (Figures 6B-D). To validate this, we then conducted in-situ hybridization experiments for gene markers labeling these specific neuronal populations (Figure 6D), including genes encoding transcription factors (*SoxC*, *Delta*, *Ngn*, *Prox1*, *Isl*, *Hbn*, *SoxB2*, *Otx*, *NeuroD1*, *Six3*), and members of neurotransmitter (*Ddc*, *Nacha6*, *Chrna9_4*, *Tph*), and neuropeptidergic signaling pathways (*An*, *Salmfap*, *Trh*).


A
Neurons
UMAP
UMAP
M
A
1E11
1. Developing neurons
2. Apical plate neurons (1)
3. Apical plate neurons (2)
4. Apical plate neurons (3)
5. Apical plate neurons (4)
6. Distal apical plate neurons
7. Ciliary band neurons (1)
8. Ciliary band neurons (2)
9. Post oral & lateral neurons
10. Mouth neurons (1)
11. Mouth neurons (2)
12. Esophageal neurons
B
Apical Organ
Ciliary band
Ciliary band
Lateral ganglion
Lateral ganglion
Postoral neurons
Postoral neurons
C
Signaling molecules & Transcription factors
Sp-Fgf9/16/20
Sp-Fgfr1
Sp-Delta
Sp-Soxb1
Sp-Soxb2
Sp-Six3
Sp-SoxC
Sp-Brn1/2/4
Sp-Pdx1
Sp-Nkx2.1
Sp-Z167
Sp-Hbn
Sp-Rx
Sp-Fez
Sp-Rhox3
Sp-Nkx3.2
Sp-Isl
Sp-Ptf1a
Sp-Sip1
Sp-Ngn
Sp-NeuroD1
Sp-Prox1
Sp-Otp
Sp-Otx
Neurotransmitter pathways
Sp-Tph
Sp-5ht7
Sp-Th
Sp-Ddc
Sp-Drd1
Sp-Drd1_2
Sp-Chat
Sp-Chrna9_4
Sp-Nacha6
Neuropeptidergic & secretion pathways
Sp-Trh
Sp-Salmfap
Sp-An
Sp-An_1
Sp-Ngfffap
Sp-Ppln1
Sp-Ppln2
Sp-Gph3
Sp-Kp
Sp-Np20
Sp-Cck
Sp-Gph1/2
Sp-Bursicona
Sp-Bursiconb
Sp-Mist
Sp-Secretogranin V
Sp-Nesfastin
Average Expression
Percent Expressed
Developing neurons
Apical plate neurons (1)
Apical plate neurons (2)
Apical plate neurons (3)
Apical plate neurons (4)
Distal apical plate neurons
Ciliary band neurons (1)
Ciliary band neurons (2)
Post oral & lateral neurons
Mouth neurons (1)
Mouth neurons (2)
Esophageal neurons
D
D1 Sp-Delta (1, 2, 4) Sp-An (4, 8, 9)
D2 Sp-SoxC (1,4) Sp-An (4, 8, 9)
D3 Sp-Brn1/2/4 (1, 9) Sp-An (4, 8, 9)
D4 Sp-Ddc (1, 8)
D5 Sp-Ngn (2, 4)
D6 Sp-Prox1 (7, 8)
D7 Sp-Nacha6 (2, 7)
D8 Sp-Isl (4, 6, 10) Sp-An (4, 8, 9)
D9 Sp-Chrna9_4 (2, 3, 7)
D10 Sp-Hbn (2)
D11 Sp-SoxB2 (2, 4)
D12 Sp-Otx (2, 8)
D13 Sp-Tph (2, 3, 4)
D14 Sp-Salmfap (3, 6, 10) Sp-An (4, 8, 9)
D15 Sp-NeuroD1 (6, 8)
D16 Sp-Six3 (2, 3, 10)
D17 Sp-Trh (6) Sp-Kp (6,10)
D18 Sp-Rhox3 (1, 6, 10)

327

Figure 6. Neuronal complexity of the 3 dpf *S. purpuratus* larva. A) (From left to right and top to bottom) UMAP highlighting the neurons cluster (green), immunohistochemical detection for the paneuronal sea urchin marker 1E11 (green), UMAP showing the 12 distinct neuronal subclusters. **B)** Schematic representation of the 3 dpf pluteus larva showing the localization of neuronal subclusters (colors as in A). **C)** Dotplot of signaling molecules, transcription factors, and neurotransmitters involved in sea urchin neuronal function and neurogenesis (colors as in A). **D)** FISH of *S. purpuratus* 3 dpf larvae with antisense probes for the neuronal genes *Sp-Delta* (D1), *Sp-SoxC* (D2), *Sp-Brn1/2/4* (D3), *Sp-Ddc* (D4), *Sp-Ngn* (D5), *Sp-Prox1* (D6), *Sp-Nacha6* (D7), *Sp-Isl* (D8), *Sp-An* (D8 and D14), *Sp-Chrna9_4* (D9), *Sp-Hbn* (D10), *Sp-SoxB2* (D11), *Sp-Otx* (D12), *Sp-Tph* (A13), *Sp-Salmfap* (D14), *Sp-NeuroD1*(D15), *Sp-Six3* (D16), *Sp-Trh* (D17), *Sp-Kp* (D17) and *Sp-Rhox3* (D18). FISH shown in figures D1-3 are paired with immunohistochemical detection of the neuropeptide Sp-An. Nuclei are labelled with DAPI (in blue). All images are stacks of merged confocal Z sections. A, anus; M, mouth.

The sea urchin larva neuronal differentiation proceeds via stepwise differentiation, including transient expression of the Notch ligand Delta, followed by expression of the transcription factors *SoxC* and *Brn1/2/4* (Garner et al., 2016). During the final stages of neurogenesis, the transcription factors *Sip1*, *Z167*, *Ngn* and *Otp* regulate differentiation of diverse neuronal populations, including apical and ciliary band neurons (Wei et al., 2016, McClay et al., 2018). In our data, we observed *Sp-Delta*, *Sp-SoxC* and *Sp-Brn1/2/4,* as well different combinations of the transcription factors mentioned above, co-localize in three neuronal populations (subclusters 1, 2 and 4), indicating neuronal differentiation is taking place in those three subclusters (Figure 6C). Interestingly, in one of these populations (subcluster 2) we found expression of the transcription factors *Sp-Rx*, *Sp-Hbn* (Figure 6D10) and *Sp-Six3* (Figure 6D16), which are known to be expressed in the periphery of the larva's apical domain (Burke et al., 2006a, Wei et al., 2009). This suggests that this population is located in the periphery of the apical plate and not within the apical organ. In the apical domain, we also detected a subcluster (number 6), which coexpress *Sp-Trh* and *Sp-Salmfap* neuropeptides (Wood et al., 2018), as well as *Sp-Kp* (Kissepeptin) (Figure 6D17).

In total, we identified three neuronal subclusters located in the apical domain (subclusters 2, 3 and 4) of the larva that express Tryptophan hydroxylase (*Tph*), which encodes a key enzyme in the serotonin biosynthesis pathway, suggesting these represent serotonergic neurons in the larva. Within the ciliary band, which comprises the larva's peripheral nervous system (Slota et al., 2020), we identified two distinct cholinergic subclusters (7 and 8) expressing the enzyme involved in acetylcholine biosynthesis (*Sp-Chat*) (Figure 6C), one of which (subcluster 8) expresses also two nicotinic acetylcholine receptors (*Nacha6*, *Chrna9*). Moreover, we identified a neuronal subcluster in close proximity with the ciliary band, which corresponds to the lateral and post-oral neurons (subcluster 9). This population has been previously characterized by our group and was found to co-express *Sp-Pdx1*, *Sp-Brn1/2/4*, and the neuropeptide Sp-An (Perillo et al., 2018). Using gene markers that mark differentiated neurons expressed in the rim of the larva's mouth, including *Sp-Nkx3.2* (Wei et al., 2011), *Sp-Isl* (Perillo et al., 2018), the neuropeptide *Sp-Salmfap* (Wood et al., 2018), and the enzyme *tyrosine hydroxylase* (*Sp-Th*) involved in the dopaminergic pathway, we identified two distinct mouth neurons subtypes (subclusters 10 and 11; Figure 6C). Lastly, we found one neuronal population that, based on its molecular signature, and in particular the co-presence of *Sp-Fgfr1* and *Sp-Fgf9/16/20* (Figure 6-Figure supplement 3)*,* could correspond to neurons associated to

375 the cardiac sphincter within the esophageal endoderm (subcluster 12). Overall, our
376 subclustering analysis increases the resolution of the different neuronal subtypes present
377 at this developmental stage, describing new neuronal subtypes and providing novel
378 markers and gene candidates for future studies of these cell types and their gene
379 regulatory networks.

380 **Characterizing a neuroendocrine neuronal population controlled by Pdx-1**

381 Previous studies from our group suggested that the nervous system of the sea urchin
382 larva displays a strong pre-pancreatic signature, with neurons expressing genes that are
383 involved in endocrine cell differentiation in the vertebrate pancreas (Perillo et al., 2018).
384 To investigate this, we focused on the post-oral and lateral neuron subcluster, which co-
385 expresses *Sp-Pdx1*, *Sp-Brn1/2/4* and *Sp-An* (Figure 6). Double immunohistochemical
386 staining of the neuronal marker 1E11 and Sp-An shows that these neurons lie in close
387 proximity to the ciliary band, and project axons towards both the apical plate and ciliary
388 band (Figure 7 A1-A3). Double FISH of *Sp-An* and *Sp-FbsL_2,* a ciliary band marker
389 revealed by this study*,* as well as double IHC of Sp-An and acetylated tubulin (labeling
390 cilia), further indicate their distribution relative to the ciliary band (Figure 7 A4-A6).
391 Moreover, we found that the post-oral Sp-An neurons are found in close proximity to cells
392 of both ciliary band subclusters (Figure 7 A7-A9), and project axons to the cell bodies of
393 the *Sp-Prox1* positive neurons (Figure 7 A9). We also observed close proximity between
394 An positive neurons in the post oral arms with the cells of the ventral-lateral cluster of
395 PMCs (Figure 7 A10), and with immune globular cells (Figure 7 A11). Next, we set out to
396 validate whether novel genes predicted by our single cell analysis to be expressed in this
397 neuronal population can be validated *in vivo*. Among the genes predicted to be expressed
398 in this population are the transcription factors *Sp-Nk1* and *Sp-Otp,* as well as the
399 catecholaminergic and cholinergic neuronal markers Sp-Th and Sp-Chat (Slota and
400 McClay, 2018), respectively. Double fluorescent *in situ* hybridization of Sp-Nk1 and the
401 neuropeptide Sp-An, as well as fluorescent *in situ* hybridization of *Sp-Otp* combined with
402 the immunohistochemical detection of Sp-An, reveal co-localization of these three genes
403 in the post-oral and lateral neuronal population (Figure 7 A13-A14), verifying the single-
404 cell data. Additionally, double immunostainings of the anti-Sp-An with anti-Th (Figure
405 7A9) and anti-Chat antibodies suggest that these two key enzymes, involved in different
406 neurotransmitter biosynthesis pathways, are co-produced in the Sp-*Pdx1/Sp-Brn1/2/4*
407 neurons (Figure 7 A12 and A15).

408 The vertebrate orthologue of *Sp-Pdx1* is essential for pancreas development, β-cell
409 differentiation, and maintaining mature β-cell function (Kaneto et al., 2007). As previously
410 described by our group, knockdown of the pancreatic transcription factor Sp-Pdx1 results
411 in severe downregulation of the Sp-An neuropeptide, compromising the neuroendocrine
412 fate of this neuronal type (Perillo et al., 2018). To further characterize the *Sp-Pdx1*/*Sp-*
413 *Brn1/2/4* neuronal population, we performed a comprehensive analysis of genes involved
414 in pancreatic development and β-cell differentiation, as well as gene markers related to
415 neuroendocrine fate. We identified a total of 20 genes, all involved in the formation and
416 proper function of vertebrate endocrine pancreas, which are differentially enriched in the
417 *Sp-Pdx1*/*Sp-Brn1/2/4* neurons (Figure 7B).

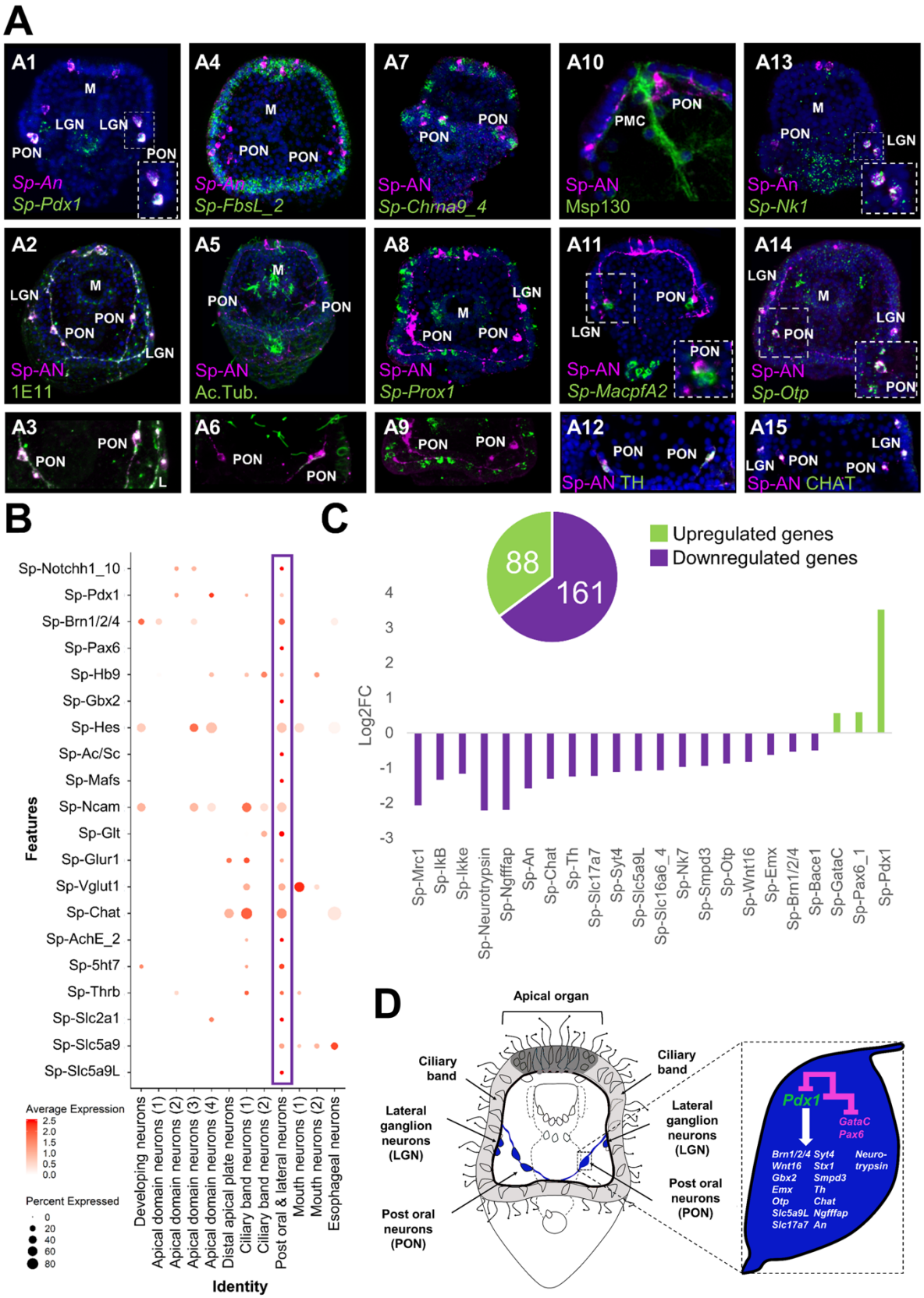


Figure 7. ScRNA-seq reveals a Pdx-1 dependent neuroendocrine cell type. **A)** Molecular

420 characterization of a *Sp-Pdx1*/*Sp-Brn1/2/4* double positive neuronal population. A1) Double FISH
421 of *S. purpuratus* 3 dpf larvae with specific antisense probes for *Sp-Pdx1* and *Sp-An*. A2) Double
422 immunohistochemical detection of the Sp-An and Synaptotagmin (1E11) proteins. A3) Close up
423 caption of the Sp-An PON neurons shown in A2. A4) Double FISH of *S. purpuratus* 3 dpf larvae
424 with specific antisense probes for *Sp-FbsL_2* and Sp-An. Double immunohistochemical detection
425 of the Sp-An and acetylated tubulin proteins. A6) Close up caption of the Sp-An PON neurons
426 shown in A5. A7) FISH of *S. purpuratus* 3 dpf larvae with a specific antisense probe for *Sp-*
427 *Chrna9_4* paired with immunodetection of Sp-An. A8) FISH of *S. purpuratus* 3 dpf larvae with a
428 specific antisense probe for *Sp-Prox1* paired with immunohistochemical detection of Sp-An. A9)
429 Close up caption of the Sp-An PON neurons shown in A8. A10) Double immunohistochemical
430 staining for the neuropeptide Sp-An and the skeletal cells marker Msp130. A11) FISH of *S.*
431 *purpuratus* 3 dpf larvae with a specific antisense probe for *Sp-MacpfA2* paired with
432 immunohistochemical detection of Sp-An. A12) Double immunohistochemical staining for the
433 neuropeptide Sp-An and the enzyme Sp-TH. A13) Double FISH of *S. purpuratus* 3 dpf larvae with
434 specific antisense probes for *Sp-Nk1* and Sp-An. A14) FISH of *S. purpuratus* 3 dpf larvae with a
435 specific antisense probe for *Sp-Otp* paired with immunohistochemical detection of Sp-An. Double
436 immunohistochemical staining for the neuropeptide Sp-An and the enzyme Sp-Chat. Nuclei are
437 labelled with DAPI (in blue). All images are stacks of merged confocal Z sections. LGN, lateral
438 ganglion neurons; M, Mouth; PON, Post-oral neurons. **B)** Dotplot of genes important in endocrine
439 pancreas differentiation and function in vertebrates. **C)** Bar plot of selected Sp-Pdx1 target genes
440 in the *Sp-Pdx1*/*Sp-Brn1/2/4* positive population as revealed by differential RNA sequencing
441 analysis of *Sp-Pdx1* knockdown larvae. **D)** Schematic representation of the Sp-Pdx1/Sp-Brn1/2/4
442 neurons genetic wiring as revealed by the combination of scRNA-seq and differential RNA-seq
443 analysis after *Sp-Pdx1* knockdown, highlighting an important role of *Sp-Pdx1* in the differentiation
444 of this cell type.

445 Next, we intersected our scRNA-seq with bulk RNA sequencing data derived from Pdx1
446 morphants assayed at the same developmental stage in a previous study (Annunziata
447 and Arnone, 2014). By coupling knowledge of cell type-specific expression programs with
448 genes differentially expressed in the Pdx1 knockdown mutants, we were able to identify
449 and refine likely gene targets specific to the *Sp-Pdx1*/*Sp-Brn1/2/4* neuronal cell type. In
450 total, we found 249 genes belonging to the *Sp-Pdx1*/*Sp-Brn1/2/4* neuron subcluster (9)
451 that were differentially expressed in the *Pdx1* knockdown dataset, with 65% of the targets
452 being downregulated (Figure 7C). Among the downregulated genes, we found key
453 transcription factors involved in neuronal differentiation, including *Sp-Brn1/2/4* and *Sp-*
454 *Otp*, as well as terminal differentiation genes important in neuronal signaling, such as *Sp-*
455 *An*, *Sp-Ngfffap*, *Sp-Th* and *Sp-Chat* (Figure 7C).

456

457 ## Discussion

458 Cell type identity is determined by the differential use of genomic information among cells.
459 Unraveling the distinct transcriptomic signatures of cell types yields valuable insight into
460 their function, as well as their evolutionary and developmental origins. In recent years,
461 single cell transcriptomics has emerged as a powerful and unbiased approach for
462 characterizing cell type diversity across a wide variety of animal taxa, with studies
463 spanning from insects (Davie et al., 2018, Severo et al., 2018, Cho et al., 2020), to
464 cnidarians (Sebe-Pedros et al., 2018, Siabert et al. 2019), planarians (Swapna et al. 2018,
465 Zeng et al. 2018) and sea squirts (Sharma et al., 2019, Cao et al., 2019), as well as

466 vertebrates such as zebrafish (Wagner et al., 2018, Chestnut et al., 2020), mice
467 (Nestorowa et al., 2016, Jung et al., 2019, Ximerakis et al., 2019, Yu et al., 2019, Qi et
468 al., 2020) and humans (Yu et al., 2019, Qi et al., 2020, Esaulova et al., 2020, Zhao et al.,
469 2020).

470 The sea urchin embryo has served as a valuable model for understanding cell type
471 molecular specification and differentiation via gene regulatory networks. Despite this,
472 knowledge of later stages of development, including larval cell types, is limited. Here we
473 used single cell RNA sequencing to generate a detailed atlas of cell types of the early
474 pluteus larva, and to unravel the neuronal diversity at this critical stage that marks the
475 end of embryogenesis and the beginning of the larval life cycle.

476 **Cellular diversity of the *S. purpuratus* pluteus larva**

477 Conducting scRNA-seq on isolated *S. purpuratus* early pluteus cells we initially identified
478 21 genetically distinct cell clusters from 19.699 cells (Figure 1 A), that consist of
479 derivatives from all three germ layers. Similar ectodermal, mesodermal and endodermal
480 representatives have also been reported at a single cell resolution in previous studies,
481 carried out at either earlier developmental stages or different sea urchin species,
482 highlighting the power of this technique to capture fine and dynamic developmental
483 processes during sea urchin development (Foster et al. 2020, Massri et al. 2021). The
484 cells comprising the 21 cell type families are expressing a total of 15,578 WHL genes
485 (transcriptome models) and 12,924 genes (SPU gene models). Notably, the
486 computationally identified number of cells per cluster did not correlate well with their
487 actual distribution in the larva. This could be a bias linked to the dissociation process, as
488 for instance skeletal structures due to their calcite and syncytial nature are the last to
489 dissociate and hence cells can get trapped within the debris and thus be underestimated
490 in our datasets. Nonetheless, the cell type families and cell types identified in our study
491 include all known larval cell types, suggesting we obtained a sufficient number of cells to
492 comprehensively survey cell diversity at this developmental stage. Further, the total
493 number of genes expressed in our data was 15.579, in relatively close agreement to the
494 16.500 genes expressed at the end of *S. purpuratus* embryogenesis (Tu et al., 2014).
495 Our results reveal a rich tapestry of cell type families within the early pluteus larva. In
496 particular, our study reveals unprecedented transcriptional diversity among cells of the
497 pluteus larva that relates to feeding and digestion, as seen by the large representation of
498 such clusters in our dataset. This includes two distinct oral ectoderm cell clusters, as well
499 as distinct cell expression programs for the esophagus, cardiac and pyloric sphincters,
500 exocrine pancreas-like cells, three distinct stomach cell clusters, intestine and anus. We
501 also identified the mesodermally derived esophageal muscle cell type that ensures the
502 proper function of the digestive apparatus by regulating the flow of the water containing
503 food within the different compartments of the gut.

504 Cell type family tree reconstruction reveals two well supported groups (Figure 3 – Figure
505 supplement 1). One group represents the endoderm. It comprises a core of closely related
506 cell type families (stomach cells 1,2,3, pyloric and cardiac sphincter, and exocrine
507 pancreas-like cells) and the more distant subgroup of intestinal and anal cells. The
508 coherence of this group is also apparent from the transcription factor heatmap (Figure 4–
509 Figure supplement 2), which reveals numerous factors shared between subsets or the

entire endodermal group such as *Klf15* that is known to control lipid and glucose metabolism in the liver (Oishi and Manabe 2018). The other group represents ectoderm. It contains apical plate, ciliary band, the aboral and upper oral ectoderm cells, and neurons. Again, this group manifests in shared transcription factor expression including *Klf2/4, Tbx2/3,* and *SoxB1* and *SoxB2*. Unexpectedly, this group firmly comprises some cell type families that are usually not considered ectodermal, such as the immune cells and the esophageal cells (Figure 3 – Figure supplement 1). Outside these two groups, the remaining cell type families do not assemble into a well-supported group. Instead, they exhibit molecular affinities to either the endodermal or ectodermal clades.

Beyond our initial cell clusters, we used subclustering to uncover diversity among neuronal, PMC, and immune cells in the pluteus larva. Among immune cells, the presence of two pigment cell subclusters is in line with findings by Perillo and colleagues that revealed two such populations (Perillo et al., 2020). However, in our study we found that only one cluster is *Sp-Gcm* positive, in contrast to their findings which found both of their clusters expressed *Sp-Gcm*. This difference could be a result of the different approaches used to identify the different cell types or due to transient expression of *Sp-Gcm* in the additional pigment cell type.

The larval nervous system has been among the first echinoderm cell types to be characterized at a molecular level and yet the exact number of neuronal subtypes is still not clear. Extensive work has been done on identifying the molecular pathways guiding neuronal specification and the genes active during neuronal differentiation, however this is limited to describing general neuronal categories. Most of the current information on the different neuronal types relies on detection of specific neuropeptides, neurotransmitters and enzymes involved in their biosynthesis. The most recent estimate of neuronal diversity used neuropeptidergic content to identify seven distinct neuronal types (Wood et al., 2018). Our study supports and refines this earlier work by providing a comprehensive and unbiased survey of neuronal diversity in the early pluteus larva of *S. purpuratus.*

Neurons in *S. purpuratus* arise from two ectodermal neurogenic regions (ciliary band and apical domain) and from the anteriormost part of the foregut (Garner et al., 2016, McClay et al., 2018, Wei et al., 2011), which is derived from endoderm (mouth neurons and esophageal neurons, see Figure 6A). Different neuronal types from these domains arise at different developmental time points, although by 3 dpf most larval neurons are thought to be present and patterning diverse larval domains. Notably, our initial clustering analysis recovered neurons as a single cluster, suggesting they share a common molecular signature regardless of their developmental origin. Subclustering of these revealed twelve distinct neuronal populations, which we were able to trace back on the larva. By doing so, we identified i) one cluster consisting of neurons that apparently are undergoing differentiation as judged by the combinatorial expression of Delta, Soxc, Brn1/2/4 and terminal differentiation gene markers, ii) four associated with the apical domain, one of which is expressing the same combination of TFs mentioned above, suggesting it might also be undergoing differentiation, iii) one matching the distal apical plate neurons, iv) two corresponding to ciliary band neurons, v) two located in the rim of the mouth, vi) one associated with esophageal structures, and vii) one corresponding to the post-oral and lateral neurons. Thus, our study greatly enhances our knowledge of neuronal diversity in

the *S. purpuratus* larva, nearly doubling the number of known neuronal cell types at this developmental stage.

Lastly, we also found one cell cluster (undefined) that could not be traced back to a specific domain. Based on the numbers and distributions of the features and transcript numbers that are comparable to the ones of real cell type families, such as the esophageal cell type family, we speculate that this represents a real cell type family and not empty droplets. (Figure1- Figure supplement 1E). Overall, this cluster exhibited greatest transcriptional similarity to cell types derived from the ectoderm (Figure 1- Figure supplement 1F), suggesting it may be of ectodermal origin. Consistent with this, a recent study revealed a similar uncharacterized ectodermal cell type (Perillo et al., 2020), further suggesting this cell population exists and is not solely an artifact of our analysis. Taking into account the great plasticity and regeneration capability of the sea urchin larva, it is also possible that this non-differentiated ectodermal cell type is a progenitor population in stasis, waiting to be activated. Future studies are necessary to validate its identity, function, and origin during larval development.

Sp-Pdx1 as a regulator of neuroendocrine fate

Morphogenesis and organogenesis rely on the hierarchical control of gene expression as encompassed in the GRNs. Although sea urchins diverged from chordates prior to the origin of the pancreas, we have previously demonstrated that the neurogenic and neuronal territories of the sea urchin embryo and larva have a strong pancreatic-like molecular signature (Perillo et al., 2018). Interestingly, one neuronal population was found to express *Sp-Pdx1* and *Sp-Brn1/2/4*, as well as the echinoderm-specific neuropeptide An. *Pdx1* in mammals is essential for proper pancreatic formation, β-cell differentiation, and regulation of the mature β-cells physiology and function (Hui and Perfetti, 2002). On the other hand, the vertebrate orthologue of *Sp-Brn1/2/4*, *Brn4* is expressed in the α pancreatic cells, where it acts as a key differentiation factor of this lineage (Hussain et al., 2002).

The gene regulatory cascade leading to the endocrine pancreas formation in vertebrates has been described in great detail and several transcription factors (Hes, Ascl1, Pax6, Hb9, Mafs, Gbx2), and signaling molecules (Notch) have been characterized as essential for this process (Ishibashi et al., 1995, de la Pompa et al., 1997, Li et al., 1999, Iso et al., 2003, Mizusawa et al., 2004, Zaret and Grompe, 2008, Arkhipova et al., 2012, Hart et al., 2013, Tritschler et al., 2017, Mitchell et al., 2017, Buckle et al., 2018) . (, de la Pompa et al., 1997, Iso et al., 2003).

Here, we dissected the molecular fingerprint of the *Sp-Pdx1*/*Sp-Brn1/2/4* neuronal type and identified the presence of genes involved in pancreatic development as well as of genes known to be expressed in both endocrine pancreatic cells and neurons. From our scRNA-seq analysis it is evident that *Sp-Notch*, *Sp-Pax6*, *Sp-Hb9*, *Sp-Hes*, *Sp-Ac/Sc* (the orthologue of ASCL1), *Sp-Mafs* and the recently re-annotated *Sp-Gbx2* (previously annotated as *Sp-Nk7*- https://new.echinobase.org) are all expressed in the *Sp-Pdx1*/*Sp-Brn1/2/4* neurons. This suggests that these neurons of a non-chordate deuterostome have a gene regulatory machinery similar to the endocrine pancreas cells.

597 It has also been demonstrated that both endocrine pancreas and neuronal cells share
598 similar features and are able to produce and respond to several neuronal genes and
599 neurotransmitters. Interestingly, we were able to identify these shared components in our
600 *Sp-Pdx1*/*Sp-Brn1/2/4* neurons. For instance, the neural cell adhesion molecule *Ncam*
601 known to be produced in the nervous system and endocrine cells of the rat is also
602 expressed in this cluster (Langley et al., 1989). Moreover, these neurons also express
603 genes encoding members of the glutamate signaling pathway (*Sp-Glt*: glutamate
604 synthase; *Sp-Vglut1*: glutamate transporter; *Sp-Glur1*: glutamate receptor), which in
605 mammals are involved in glucose-responsive insulin secretion (Gonoi et al., 1994,
606 Maechler and Wollheim, 1999), and tyrosine hydroxylase (*Th*), the rate-limiting enzyme
607 of catecholamine biosynthesis, which is present in the endocrine pancreas of multiple
608 species (Teitelman et al., 1993, Iturriza and Thibault, 1993). Furthermore, *Sp-Pdx1*/*Sp-*
609 *Brn1/2/4* neurons express choline acetyltransferase (*Chat*), which has been found to be
610 highly expressed in human pancreatic islets and may be essential for the stimulation of
611 insulin secretion by the neighboring β-cells (Rodriguez-Diaz et al., 2011).

612 Additionally, serotonergic signaling is believed to be involved in the regulation of insulin
613 secretion as several serotonin receptors have been found to be expressed in human
614 pancreatic islets (Amisten et al., 2015). Transcripts of the sea urchin serotonin receptor
615 *Sp-5ht7* are present in *Sp-Pdx1*/*Sp-Brn1/2/4* neuronal population suggesting the
616 serotonin might regulate the neurotransmitter/neuropeptide secretion in a similar way as
617 for insulin. Similarly, transcripts of the Thyroid hormone receptor B are also found in this
618 population, suggesting that thyroid hormone signaling may play a similar role to its
619 differentiation similar to the one in murine endocrine pancreas differentiation (Aiello et al.,
620 2014). The hormone secretion mediated by pancreatic endocrine cells depends on their
621 ability to detect changes in extracellular glucose levels. To this end, they are equipped
622 with Glucose transporters and co-transporters (Navale and Paranjape, 2016, Berger and
623 Zdzieblo, 2020). Our analysis revealed that the *Sp-Pdx1*/*Sp-Brn1/2/4* neurons produce
624 transcripts of three glucose co-transporter genes (*Sp-Slc2a1*, *Sp-Slc5a9* and *Sp-*
625 *Slc5a9L*) proposing that they are able to detect such changes in glucose levels similarly
626 to the endocrine pancreas cells.

627 **Evolutionary considerations**

628 Taken together, our data show that the sea urchin *Sp-Pdx1/Sp-Brn1/2/4* neurons uniquely
629 express a combination of key genes that are necessary for the endocrine pancreas
630 differentiation and function in the vertebrates (Figure7 & Figure 7- figure supplement 1).
631 These shared molecular features of *Sp-Pdx1/Sp-Brn1/2/4* neurons and pancreatic cells
632 suggest these represent features of a cell type present in the deuterostome ancestor.
633 What was the nature of these cells and what happened in the evolution of the descendant
634 lineages so that the molecular characteristics of these cells are found in extant cell types
635 as divergent as ectodermal neurons and pancreatic cells?

636 One possibility is that the ancestral deuterostomes already possessed a *Sp-Pdx1/Sp-*
637 *Brn1/2/4* neuronal cell type similar to the one found in today's sea urchin, and that the
638 regulatory program of this cell type was then co-opted into the endodermal lineage in the
639 vertebrates. The co-option and incorporation of gene regulatory elements to different
640 developmental or morphogenetic context has been postulated before and has the

potential to give rise to diverse cell types (Monteiro, 2012, Preger-Ben Noon and Frankel, 2015, Martik and McClay, 2015, Hu et al., 2018, Morgulis et al., 2019, McQueen and Rebeiz, 2020, Cary et al., 2020). In particular, it has been hypothesized that β pancreatic cells arose during evolution by co-option of a preexisting neuronal cell type program into the pancreas developmental lineage (Arntfield and van der Kooy, 2011) based on the many physiological, morphological and molecular features endocrine pancreatic cells share with neurons (Alpert et al., 1988, Eberhard, 2013). However, given the long list of genes shared between neurons and pancreatic cells, this would have necessarily represented a massive co-option event profoundly altering the identity and function of the receiving cells. Notably, a flexibility towards the co-option of a neuroendocrine phenotype has also been reported in the vertebrate lineage in the case of the adrenal medulla derived PC12 cell line, that upon stimulation with NGF, adopts a neurosecretory fate (Fujita et., 1989).

A second possibility is that both ectodermal neurons and endodermal pancreatic cells are direct evolutionary descendants of the same precursor cell, i.e. represent sister cell types (Arendt, 2008). This explanation would avoid co-option; instead, it postulates a direct evolutionary link between cell types derived from different germ layers - ectoderm versus endoderm. Surprising at first, a similar link between digestive and neuronal-type expression profiles has been reported repeatedly. For example, it appears to hold true for choanocytes in the early-branching sponges (Musser et al., 2019), indicating that the family of sponge choanocytes might be related to the hypothesized digestive-neuronal precursors. In the sea anemone *Nematostella vectensis*, pharyngeal cells give rise to both digestive cells and neurons (Steinmetz et al., 2017), and neuronal cell types and secretory gland cells exhibit related molecular profiles (Sebe-Pedros et al., 2018). In fact, based on these findings a close evolutionary link between endoderm and ectoderm has been postulated (Steinmetz et al., 2017). Finally, motor neurons in the vertebrate ventral neural tube and pancreatic islet cells share a specific combination of transcription factors including *pax6, nkx6*, and *hb9* (reviewed in Arendt, 2021). Overall, this second possibility is also in line with the hypothesis that at least subsets of neuron types in animal nervous systems may have originated from a mucociliary sole that initially harbored a neurosecretory network of cells with digestive and communicative functions (Arendt et al., 2015; Arendt, 2021). Via division of labor, these precursor cell types may have then given rise to both digestive as well as neuronal cell types in different animal phyla.

Materials and Methods

1. Animal husbandry and culture of embryos

Adult *Strongylocentrotus purpuratus* individuals were obtained from Patrick Leahy (Kerckhoff Marine Laboratory, California Institute of Technology, Pasadena, CA, USA) and maintained in circulating seawater aquaria at Stazione Zoologica Anton Dohrn in Naples. Gametes were obtained by vigorous shaking of the animals. Embryos and larvae were cultured at 15°C in filtered Mediterranean Sea water diluted 9:1 with deionized water.

684 **2. Larvae dissociation**

685 Dissociation of the 3 dpf *Strongylocentrotus purpuratus* plutei into single cells was
686 performed according to adaptation of several protocols (McClay, 1986, McClay, 2004,
687 Juliano et al., 2014). Larvae were collected, concentrated using a 40 µm Nitex mesh filter
688 and spun down at 500 g for 5 min. Sea water was removed and larvae were resuspended
689 in Ca^{2+} Mg^{2+}free artificial sea water. Larvae were spun down at 500 g for 5 min and
690 resuspended in dissociation buffer containing 1M glycine and 0.02 M EDTA in Ca^{2+}
691 Mg^{2+}Free artificial sea water. Larvae were incubated for 10 min on ice and mixed gently
692 via pipette aspiration every 2 min. From that point and onwards the progress of
693 dissociation was monitored. Dissociated cells were spun down at 700 g for 5 min and
694 washed several times with Ca^{2+} Mg^{2+}Free artificial sea water. Cell viability was assessed
695 via using Propidium Iodide and Fluorescein diacetate and only specimens with cell
696 viability ≥ 90% were further processed. Single cells were counted using a hemocytometer
697 and diluted according to the manufacturer's protocol (10x Genomics). Throughout this
698 procedure samples were kept at 4°C.

699

700 **3. Single cell RNA sequencing and data analysis**

701 Single cell RNA sequencing was performed using the 10x Genomics single cell capturing
702 system. Specimens from four independent biological replicates, ranging from 6000-
703 20.000 cells, were loaded on the 10X Genomics Chromium Controller. Single cell cDNA
704 libraries were prepared using the Chromium Single Cell 3' Reagent Kit (Chemistries v2
705 and v3). Libraries were sequenced by GeneCore (EMBL, Heidelberg, Germany) for 75
706 bp paired-end reads (Illumina NextSeq 500), resulting in a mean of 88M reads. Cell
707 Ranger Software Suite 3.0.2 (10x Genomics) was used for the alignment of the single-
708 cell RNA-seq output reads and generation of feature, barcode and matrices. The genomic
709 index was made in Cell Ranger using the S. purpuratus genome version 3.1 (Sea Urchin
710 Genome Sequencing et al., 2006, Kudtarkar and Cameron, 2017). Cell Ranger output
711 matrices for four biological and two technical replicates were used for further analysis in
712 Seurat v3.0.2 R package (Stuart et al., 2019). The analysis was performed according to
713 the Seurat scRNA-seq R package documentation (Butler et al., 2018, Stuart et al., 2019).
714 Genes that are transcribed in less than three cells and cells that have less than a minimum
715 of 200 transcribed genes were excluded from the analysis. The cutoff number of
716 transcribed genes was determined based on feature scatter plots and varies depending
717 on the replicate. 19.699 cells out of the 29,130 cells estimated by Cell Ranger passed the
718 quality checks and were further analyzed. Datasets were normalized and variable genes
719 were found using the vst method with a maximum of 2000 variable features. Data
720 integration was performed via identification of anchors between the six different objects.
721 Next the datasets were scaled and principal component (PCA) analysis was performed.
722 Nearest Neighbor (SNN) graph was computed with 20 dimensions (resolution 1.0) to
723 identify the clusters. Uniform Manifold Approximate and Projection (UMAP) was used to
724 perform clustering dimensionality reduction. Cluster markers were found using the genes
725 that are detected in at least 0.01 fraction of min.pct cells in the two clusters. Transcripts
726 of all genes per cell type were identified by converting a Seurat DotPlot with all these

transcripts as features into a table (ggplot2 3.2.0 R package). Subclustering analysis was performed by selecting a cell type of interest and performing similar analysis as described above. All resulting tables containing the genes transcribed within different cell type families were further annotated adding PFAM terms (Trapnell et al., 2010, Finn et al., 2014) for associated proteins, gene ontology terms and descriptions from Echinobase (Kudtarkar and Cameron, 2017).

Cell type tree reconstruction was performed in R (version 4.0.3) using the neighbor-joining tree function in package 'ape' (version 5.4-1). Two final trees were constructed, one that derived pairwise expression distances between clusters using all expressed genes, and a second in the analysis was restricted to only expressed transcription factors. A gene was defined as "expressed" when it had an average expression of greater than 3 transcripts per million in at least one of the 21 major clusters. For each tree, pairwise distances were calculated as 1 – spearman correlation between each of the 20 major clusters, excluding the undefined cells, using the spearman.dist function in the bioDist package (version 1.60.0). To assess the support of each bifurcation in the tree, we performed 10,000 bootstrap replicates using the boot.phylo function in 'ape'. Although not reported in this manuscript, we also tested the effect of using different distance metrics and normalization methods on tree topology. Specifically, we reconstructed trees using distances derived from Pearson correlations of either log or square root normalized average expression values, and Euclidean distances of scaled log-normalized expression values. In general, similar topologies were found in all analyses, recovering clades of endodermal and ectodermal derivatives. Mesodermal derivatives often, but not always, grouped together, and the position of the neuron cluster moved in and out of the clade of ectodermal derivatives. This latter result may possibly arise because of conflicting signals resulting from cell type heterogeneity (i.e. different neuron types) within the cluster.

In order to assess the significance of identifying the homologs of the vertebrate genes involved in the development of the pancreas in the "post oral and lateral neurons" sub-population of neuronal cells, the normalized counts from all the transcripts present in the neuronal subcluster were extracted in the form of a matrix using GetAssayData command from Seurat v3.0.2. The counts within this matrix were then randomized with 100000 iterations using the randomizeMatrix function from the picante v1.8.2 package (Kembel et al., 2010) with the independent swap null model. A new Seurat object was created, which is the same as the neuronal subclustering object, however the actual transcript normalized counts were replaced with the randomized ones. Then the average expression of the 20 pancreatic genes was extracted from both from actual neuronal subclustering Seurat object and the one with randomized counts using AverageExpression function to visualize the expression of these pancreatic genes in the "post oral and lateral neurons" subcluster in actual data and randomized data. This process was looped 500 times for randomized data to then calculate the mean number of pancreatic genes expressed in the "post oral and lateral neurons" subcluster to obtain an estimate of the expected number of genes expressed, rounded to the closest integer, if their expression was found in this subcluster by random. The number of genes

expressed in the actual data and the randomized data was then used to generate proportions of the number of expressed and not expressed pancreatic genes in this neuronal subcluster. Then, we have performed a chi-squared test within R (Hope, 1968) to show that the proportion from the actual data proportion is significantly different to the randomized data proportion with p-value less than 0.002.

4. Whole mount RNA Fluorescent *in situ* hybridization

Fluorescent *in situ* hybridization was performed as outlined in (Perillo et al., 2021). Fluorescent signal was developed via using fluorophore conjugated tyramide technology (Perkin Elmer, Cat. #NEL752001KT). Antisense probes were transcribed from linearized DNA and labeled either during transcription via using digoxigenin-11-UTP nucleotides, or post-transcriptionally by using Fluorescein (Mirus Bio, Cat. #MIR3200) or DNP (Mirus Bio, Cat. #MIR3825) following the manufacturer's instructions. Probes for *Sp-Pdx1, Sp-Cdx, Sp-ManrC1A, Sp-Six1/2, Sp-Fgf9/16/20, Sp-Brn1/2/4, Sp-Ngn, Sp-Isl, Sp-NeuroD1, Sp-Pks1, Sp-Soxb2, Sp-An, Sp-Trh and Sp-Salmfap* were produced as previously published [*Sp-Pdx1, Sp-Cdx* (Cole et al., 2009), *Sp-ManrC1A* (Annunziata et al., 2014), *Sp-Six1/2, Sp-Fgf9/16/20* (Andrikou et al., 2015), *Sp-Brn1/2/4* (Cole and Arnone, 2009), *Sp-Ngn, Sp-Isl, Sp-NeuroD1* (Perillo et al., 2018), *Sp-Pks1* (Perillo et al, 2020), *Sp-SoxB2* (Anishchenko et al., 2018), *Sp-An, Sp-Trh, Sp-Salmfap* (Wood et al., 2018)]. Primer sequences used for cDNA isolation and probes synthesis are in Supplementary file 2. Specimens were imaged using a Zeiss LSM 700 confocal microscope.

5. Immunohistochemistry (IHC)

Immunohistochemical staining or IHC paired with FISH was performed as described in (Perillo et al., 2021). Briefly 3 dpf plutei were fixed in 4 % paraformaldehyde (PFA) in filtered sea water (FSW) for 15 min at room temperature (RT). FSW was removed and samples were incubated in 100% methanol for 1 min at RT, washed multiple times with phosphate buffer saline with 0.1% Tween 20 (PBST) and incubated blocking solution containing 1 mg/ml Bovine Serum Albumin (BSA) and 4% sheep in PBST for 1h. Primary antibodies were added in the appropriate dilution and incubated for 1h and 30 min at 37°C. Anti-acetylated alpha tubulin (Sigma-Aldrich T67930) was used to label cilia and microtubules (1:200), Anti-Msp130 (gift from Dr. David R. McClay) to label skeletogenic cells (undiluted), 1E11 (gift from Dr. Robert Burke) to mark the nervous system (1:20), 5c7 (gift from Dr. David R. McClay) to label the endoderm (undiluted), Sp-An to label the post-oral and lateral neurons (1:250), Sp-Th (Sigma-Aldrich AB152) to label catecholaminergic neurons (1:100) and Sp-Chat (GeneTex GXGTX113164S) to label cholinergic neurons (1:100). Specimens were washed multiple times with PBST and incubated for 1h with the appropriate secondary antibody (AlexaFluor) diluted 1:1000 in PBST. Larvae were washed several times with PBST and imaged using a Zeiss LSM 700 confocal microscope.

6. EdU labelling paired with immunohistochemistry

In order to understand the spatial distribution of proliferating cells across the putative broad cell types, cell proliferation assays were carried out using Click-It EdU Cell

812 Proliferation Kit for Imaging Alexa Flour™ 647 (Thermo Fisher Scientific). Larvae were
813 treated with EdU at a final concentration of 10 µM in FSW and let to grow for 2 hours.
814 Samples were fixed in 4% PFA in FSW for 15 min (RT) and washed several times with
815 PBST. PBST was removed, replaced by 100% Methanol for 1 min (RT) and followed by
816 several washes with PBST. After this step one can continue with either developing the
817 EdU signal or performing immunohistochemistry as described above. In order to develop
818 the EdU signal the Click-iT ™ reaction mix was prepared according to the manufacturer's
819 guidelines. PBST was removed and the reaction mix was added to the samples for 30
820 min (RT). Larvae were washed several times with PBST, mounted and imaged using a
821 Zeiss LSM 700 confocal microscope
822

823 **7. Gene regulatory network draft**

824 Gene regulatory modules and networks were drafted using the interactive tool for
825 building and visualizing GRNs BioTapestry (Longabaugh, 2012).

826

827 ## Competing interests

828 The authors declare no competing interests, or appearance of competing interests, in the
829 production and dissemination of this work.
830

831 ## Acknowledgements

832 The authors would like to thank Prof. Paola Oliveri (UCL) for providing critical input on the
833 analysis and structure of the manuscript. We are grateful to Drs. David McClay and Robert
834 Burke for kindly providing antibodies and Dr. Francesco Lamanna (ZMBH) for helping
835 with the computational analysis. We also would like to thank Davide Caramiello for taking
836 care of the adult sea urchins, Dr. Giovanna Benvenuto (SZN) for microscopy assistance
837 and the Arnone lab members Maria Cocurullo and Inés Fournon Berodia for their help in
838 gene cloning and preparation of several RNA probes. We also thank Drs. Vladimir Benes
839 and Bianka Baying (GeneCore, Heidelberg, Germany) for sequencing of our single cell
840 datasets. Also we are grateful to the animal technician Emily Savage (EMBL, Heidelberg,
841 Germany) for her precious assistance.
842

843 ## Funding

844 This project has been supported by the European Union's Horizon 2020 research and
845 innovation programme under the Marie Skłodowska-Curie grant agreement no. 766053
846 (EvoCELL: grant to MIA and DA, fellowship to PP) and the Advanced grant
847 'NeuralCellTypeEvo' 788921 by the European Commission (grant to DA and JM).

848

References

ADOMAKO-ANKOMAH, A. & ETTENSOHN, C. A. 2014. Growth factors and early mesoderm morphogenesis: insights from the sea urchin embryo. *Genesis,* 52**,** 158-72.

AIELLO, V., MORENO-ASSO, A., SERVITJA, J. M. & MARTIN, M. 2014. Thyroid hormones promote endocrine differentiation at expenses of exocrine tissue. *Exp Cell Res,* 322**,** 236-48.

AMISTEN, S., NEVILLE, M., HAWKES, R., PERSAUD, S. J., KARPE, F. & SALEHI, A. 2015. An atlas of G-protein coupled receptor expression and function in human subcutaneous adipose tissue. *Pharmacol Ther,* 146**,** 61-93.

ANDRIKOU, C., IOVENE, E., RIZZO, F., OLIVERI, P. & ARNONE, M. I. 2013. Myogenesis in the sea urchin embryo: the molecular fingerprint of the myoblast precursors. *Evodevo,* 4**,** 33.

ANDRIKOU, C., PAI, C. Y., SU, Y. H. & ARNONE, M. I. 2015. Logics and properties of a genetic regulatory program that drives embryonic muscle development in an echinoderm. *Elife,* 4.

ANGERER, R. C. & DAVIDSON, E. H. 1984. Molecular indices of cell lineage specification in sea urchin embryos. *Science,* 226**,** 1153-60.

ANISHCHENKO, E., ARNONE, M. I. & D'ANIELLO, S. 2018. SoxB2 in sea urchin development: implications in neurogenesis, ciliogenesis and skeletal patterning. *Evodevo,* 9**,** 5.

ANNUNZIATA, R., ANDRIKOU, C., PERILLO, M., CUOMO, C. & ARNONE, M. I. 2019. Development and evolution of gut structures: from molecules to function. *Cell Tissue Res,* 377**,** 445-458.

ANNUNZIATA, R. & ARNONE, M. I. 2014. A dynamic regulatory network explains ParaHox gene control of gut patterning in the sea urchin. *Development,* 141**,** 2462-72.

ANNUNZIATA, R., PERILLO, M., ANDRIKOU, C., COLE, A. G., MARTINEZ, P. & ARNONE, M. I. 2014. Pattern and process during sea urchin gut morphogenesis: the regulatory landscape. *Genesis,* 52**,** 251-68.

ARENDT, D. 2008. The evolution of cell types in animals: emerging principles from molecular studies. *Nat Rev Genet,* 9**,** 868-82.

ARENDT, D, BENITO-GUTIERREZ, E., BRUNET, T., MARLOW H. 2015. Gastric pouches and the mucociliary sole: setting the stage for nervous system evolution. Phil. Trans. R. Soc. B 370: 20150286

ARENDT, D. 2021.Elementary nervous systems. Phil. Trans. R. Soc. B 376: 20200347.

ARKHIPOVA, V., WENDIK, B., DEVOS, N., EK, O., PEERS, B. & MEYER, D. 2012. Characterization and regulation of the hb9/mnx1 beta-cell progenitor specific enhancer in zebrafish. *Dev Biol,* 365**,** 290-302.

ARNONE, M.I., BYRNE, M., MARTINEZ, P. 2015. Echinodermata. In: Wanninger A. (eds) Evolutionary Developmental Biology of Invertebrates 6. Springer, Vienna.

ARNONE, M. I., ANDRIKOU, C. & ANNUNZIATA, R. 2016. Echinoderm systems for gene regulatory studies in evolution and development. *Curr Opin Genet Dev,* 39**,** 129-137.

894 ARNTFIELD, M. E. & VAN DER KOOY, D. 2011. beta-Cell evolution: How the pancreas
895 borrowed from the brain: The shared toolbox of genes expressed by neural and
896 pancreatic endocrine cells may reflect their evolutionary relationship. *Bioessays,*
897 33**,** 582-7.
898 BEN-TABOU DE-LEON, S., SU, Y. H., LIN, K. T., LI, E. & DAVIDSON, E. H. 2013.
899 Gene regulatory control in the sea urchin aboral ectoderm: spatial initiation,
900 signaling inputs, and cell fate lockdown. *Dev Biol,* 374**,** 245-54.
901 BERGER, C. & ZDZIEBLO, D. 2020. Glucose transporters in pancreatic islets. *Pflugers*
902 *Arch,* 472**,** 1249-1272.
903 BISGROVE, B. W. & BURKE, R. D. 1987. Development of the Nervous-System of the
904 Pluteus Larva of Strongylocentrotus-Droebachiensis. *Cell and Tissue Research,*
905 248**,** 335-343.
906 BUCKLE, A., NOZAWA, R. S., KLEINJAN, D. A. & GILBERT, N. 2018. Functional
907 characteristics of novel pancreatic Pax6 regulatory elements. *Hum Mol Genet,*
908 27**,** 3434-3448.
909 BURKE, R. D., ANGERER, L. M., ELPHICK, M. R., HUMPHREY, G. W., YAGUCHI, S.,
910 KIYAMA, T., LIANG, S., MU, X., AGCA, C., KLEIN, W. H., BRANDHORST, B. P.,
911 ROWE, M., WILSON, K., CHURCHER, A. M., TAYLOR, J. S., CHEN, N.,
912 MURRAY, G., WANG, D., MELLOTT, D., OLINSKI, R., HALLBOOK, F. &
913 THORNDYKE, M. C. 2006a. A genomic view of the sea urchin nervous system.
914 *Dev Biol,* 300**,** 434-60.
915 BURKE, R. D., MOLLER, D. J., KRUPKE, O. A. & TAYLOR, V. J. 2014. Sea urchin
916 neural development and the metazoan paradigm of neurogenesis. *Genesis,* 52**,**
917 208-21.
918 BURKE, R. D., OSBORNE, L., WANG, D., MURABE, N., YAGUCHI, S. & NAKAJIMA,
919 Y. 2006b. Neuron-specific expression of a synaptotagmin gene in the sea urchin
920 Strongylocentrotus purpuratus. *J Comp Neurol,* 496**,** 244-51.
921 BUTLER, A., HOFFMAN, P., SMIBERT, P., PAPALEXI, E. & SATIJA, R. 2018.
922 Integrating single-cell transcriptomic data across different conditions,
923 technologies, and species. *Nat Biotechnol,* 36**,** 411-420.
924 CALESTANI, C. & ROGERS, D. J. 2010. Cis-regulatory analysis of the sea urchin
925 pigment cell gene polyketide synthase. *Dev Biol,* 340**,** 249-55.
926 CAMERON, R. A. & DAVIDSON, E. H. 1991. Cell type specification during sea urchin
927 development. *Trends Genet,* 7**,** 212-8.
928 CAMERON, R. A., HOUGH-EVANS, B. R., BRITTEN, R. J. & DAVIDSON, E. H. 1987.
929 Lineage and fate of each blastomere of the eight-cell sea urchin embryo. *Genes*
930 *Dev,* 1**,** 75-85.
931 CAO, C., LEMAIRE, L. A., WANG, W., YOON, P. H., CHOI, Y. A., PARSONS, L. R.,
932 MATESE, J. C., WANG, W., LEVINE, M. & CHEN, K. 2019. Comprehensive
933 single-cell transcriptome lineages of a proto-vertebrate. *Nature,* 571**,** 349-354.
934 CARY, G. A., MCCAULEY, B. S., ZUEVA, O., PATTINATO, J., LONGABAUGH, W. &
935 HINMAN, V. F. 2020. Systematic comparison of sea urchin and sea star
936 developmental gene regulatory networks explains how novelty is incorporated in
937 early development. *Nat Commun,* 11**,** 6235.

938 CHESTNUT, B., CASIE CHETTY, S., KOENIG, A. L. & SUMANAS, S. 2020. Single-cell
939 transcriptomic analysis identifies the conversion of zebrafish Etv2-deficient
940 vascular progenitors into skeletal muscle. *Nat Commun,* 11**,** 2796.
941 CHO, B., YOON, S. H., LEE, D., KORANTENG, F., TATTIKOTA, S. G., CHA, N., SHIN,
942 M., DO, H., HU, Y., OH, S. Y., LEE, D., VIPIN MENON, A., MOON, S. J.,
943 PERRIMON, N., NAM, J. W. & SHIM, J. 2020. Single-cell transcriptome maps of
944 myeloid blood cell lineages in Drosophila. *Nat Commun,* 11**,** 4483.
945 COLE, A. G. & ARNONE, M. I. 2009. Fluorescent in situ hybridization reveals multiple
946 expression domains for SpBrn1/2/4 and identifies a unique ectodermal cell type
947 that co-expresses the ParaHox gene SpLox. *Gene Expr Patterns,* 9**,** 324-8.
948 COLE, A. G., RIZZO, F., MARTINEZ, P., FERNANDEZ-SERRA, M. & ARNONE, M. I.
949 2009. Two ParaHox genes, SpLox and SpCdx, interact to partition the posterior
950 endoderm in the formation of a functional gut. *Development,* 136**,** 541-9.
951 DAVIDSON, E. H., CAMERON, R. A. & RANSICK, A. 1998. Specification of cell fate in
952 the sea urchin embryo: summary and some proposed mechanisms.
953 *Development,* 125**,** 3269-90.
954 DAVIDSON, E. H. & ERWIN, D. H. 2006. Gene regulatory networks and the evolution of
955 animal body plans. *Science,* 311**,** 796-800.
956 DAVIDSON, E. H., MCCLAY, D. R. & HOOD, L. 2003. Regulatory gene networks and
957 the properties of the developmental process. *Proc Natl Acad Sci U S A,* 100**,**
958 1475-80.
959 DAVIE, K., JANSSENS, J., KOLDERE, D., DE WAEGENEER, M., PECH, U., KREFT,
960 L., AIBAR, S., MAKHZAMI, S., CHRISTIAENS, V., BRAVO GONZALEZ-BLAS,
961 C., POOVATHINGAL, S., HULSELMANS, G., SPANIER, K. I., MOERMAN, T.,
962 VANSPAUWEN, B., GEURS, S., VOET, T., LAMMERTYN, J., THIENPONT, B.,
963 LIU, S., KONSTANTINIDES, N., FIERS, M., VERSTREKEN, P. & AERTS, S.
964 2018. A Single-Cell Transcriptome Atlas of the Aging Drosophila Brain. *Cell,* 174**,**
965 982-998 e20.
966 DE LA POMPA, J. L., WAKEHAM, A., CORREIA, K. M., SAMPER, E., BROWN, S.,
967 AGUILERA, R. J., NAKANO, T., HONJO, T., MAK, T. W., ROSSANT, J. &
968 CONLON, R. A. 1997. Conservation of the Notch signalling pathway in
969 mammalian neurogenesis. *Development,* 124**,** 1139-48.
970 DULOQUIN, L., LHOMOND, G. & GACHE, C. 2007. Localized VEGF signaling from
971 ectoderm to mesenchyme cells controls morphogenesis of the sea urchin embryo
972 skeleton. *Development,* 134**,** 2293-302.
973 ESAULOVA, E., CANTONI, C., SHCHUKINA, I., ZAITSEV, K., BUCELLI, R. C., WU, G.
974 F., ARTYOMOV, M. N., CROSS, A. H. & EDELSON, B. T. 2020. Single-cell
975 RNA-seq analysis of human CSF microglia and myeloid cells in
976 neuroinflammation. *Neurol Neuroimmunol Neuroinflamm,* 7.
977 ETTENSOHN, C. A. 2020. The gene regulatory control of sea urchin gastrulation. *Mech*
978 *Dev***,** 103599.
979 FINN, R. D., BATEMAN, A., CLEMENTS, J., COGGILL, P., EBERHARDT, R. Y.,
980 EDDY, S. R., HEGER, A., HETHERINGTON, K., HOLM, L., MISTRY, J.,
981 SONNHAMMER, E. L., TATE, J. & PUNTA, M. 2014. Pfam: the protein families
982 database. *Nucleic Acids Res,* 42**,** D222-30.

983 FOSTER, S., OULHEN, N., & WESSEL, G. 2020. A single cell RNA sequencing
984 resource for early sea urchin development. Development (Cambridge, England),
985 147(17), dev191528.
986 FU, X., HE, F., LI, Y., SHAHVERANOV, A. & HUTCHINS, A. P. 2017. Genomic and
987 molecular control of cell type and cell type conversions. *Cell Regen,* 6**,** 1-7.
988 FUJITA, K., LAZAROVICI, P. & GUROFF, G. 1989. Regulation of the differentiation of
989 PC12 pheochromocytoma cells. Environ Health Perspect. 80:127-42.
990 GARNER, S., ZYSK, I., BYRNE, G., KRAMER, M., MOLLER, D., TAYLOR, V. &
991 BURKE, R. D. 2016. Neurogenesis in sea urchin embryos and the diversity of
992 deuterostome neurogenic mechanisms. *Development,* 143**,** 286-97.
993 GONOI, T., MIZUNO, N., INAGAKI, N., KUROMI, H., SEINO, Y., MIYAZAKI, J. &
994 SEINO, S. 1994. Functional neuronal ionotropic glutamate receptors are
995 expressed in the non-neuronal cell line MIN6. *J Biol Chem,* 269**,** 16989-92.
996 HARKEY, M. A., WHITELEY, H. R. & WHITELEY, A. H. 1992. Differential expression of
997 the msp130 gene among skeletal lineage cells in the sea urchin embryo: a three
998 dimensional in situ hybridization analysis. *Mech Dev,* 37**,** 173-84.
999 HARLOW, P. & NEMER, M. 1987. Coordinate and selective beta-tubulin gene
1000 expression associated with cilium formation in sea urchin embryos. *Genes Dev,*
1001 1**,** 1293-304.
1002 HART, A. W., MELLA, S., MENDRYCHOWSKI, J., VAN HEYNINGEN, V. & KLEINJAN,
1003 D. A. 2013. The developmental regulator Pax6 is essential for maintenance of
1004 islet cell function in the adult mouse pancreas. *PLoS One,* 8**,** e54173.
1005 HO, E. C., BUCKLEY, K. M., SCHRANKEL, C. S., SCHUH, N. W., HIBINO, T., SOLEK,
1006 C. M., BAE, K., WANG, G. & RAST, J. P. 2017. Perturbation of gut bacteria
1007 induces a coordinated cellular immune response in the purple sea urchin larva.
1008 *Immunol Cell Biol,* 95**,** 647.
1009 HOPE, A. C. A. 1968. 'A Simplified Monte Carlo Significance Test Procedure', *Journal*
1010 *of the Royal Statistical Society: Series B (Methodological)*, 30(3), pp. 582–598
1011 HOWARD-ASHBY, M., MATERNA, S. C., BROWN, C. T., CHEN, L., CAMERON, R. A.
1012 & DAVIDSON, E. H. 2006. Identification and characterization of homeobox
1013 transcription factor genes in Strongylocentrotus purpuratus, and their expression
1014 in embryonic development. *Dev Biol,* 300**,** 74-89.
1015 HU, Y., SCHMITT-ENGEL, C., SCHWIRZ, J., STROEHLEIN, N., RICHTER, T.,
1016 MAJUMDAR, U. & BUCHER, G. 2018. A morphological novelty evolved by co-
1017 option of a reduced gene regulatory network and gene recruitment in a beetle.
1018 *Proc Biol Sci,* 285.
1019 HUI, H. & PERFETTI, R. 2002. Pancreas duodenum homeobox-1 regulates pancreas
1020 development during embryogenesis and islet cell function in adulthood. *Eur J*
1021 *Endocrinol,* 146**,** 129-41.
1022 HUSSAIN, M. A., MILLER, C. P. & HABENER, J. F. 2002. Brn-4 transcription factor
1023 expression targeted to the early developing mouse pancreas induces ectopic
1024 glucagon gene expression in insulin-producing beta cells. *J Biol Chem,* 277**,**
1025 16028-32.
1026 ISHIBASHI, M., ANG, S. L., SHIOTA, K., NAKANISHI, S., KAGEYAMA, R. &
1027 GUILLEMOT, F. 1995. Targeted disruption of mammalian hairy and Enhancer of
1028 split homolog-1 (HES-1) leads to up-regulation of neural helix-loop-helix factors,

1029 premature neurogenesis, and severe neural tube defects. *Genes Dev,* 9**,** 3136-
1030 48.
1031 ISO, T., KEDES, L. & HAMAMORI, Y. 2003. HES and HERP families: multiple effectors
1032 of the Notch signaling pathway. *J Cell Physiol,* 194**,** 237-55.
1033 ITURRIZA, F. C. & THIBAULT, J. 1993. Immunohistochemical investigation of tyrosine-
1034 hydroxylase in the islets of Langerhans of adult mice, rats and guinea pigs.
1035 *Neuroendocrinology,* 57**,** 476-80.
1036 JULIANO, C., SWARTZ, S. Z. & WESSEL, G. 2014. Isolating specific embryonic cells of
1037 the sea urchin by FACS. *Methods Mol Biol,* 1128**,** 187-96.
1038 JULIANO, C. E., VORONINA, E., STACK, C., ALDRICH, M., CAMERON, A. R. &
1039 WESSEL, G. M. 2006. Germ line determinants are not localized early in sea
1040 urchin development, but do accumulate in the small micromere lineage. *Dev Biol,*
1041 300**,** 406-15.
1042 JULIANO, C. E., YAJIMA, M. & WESSEL, G. M. 2010. Nanos functions to maintain the
1043 fate of the small micromere lineage in the sea urchin embryo. *Dev Biol,* 337**,** 220-
1044 32.
1045 JUNG, M., WELLS, D., RUSCH, J., AHMAD, S., MARCHINI, J., MYERS, S. R. &
1046 CONRAD, D. F. 2019. Unified single-cell analysis of testis gene regulation and
1047 pathology in five mouse strains. *Elife,* 8.
1048 KANETO, H., MIYATSUKA, T., KAWAMORI, D., YAMAMOTO, K., KATO, K.,
1049 SHIRAIWA, T., KATAKAMI, N., YAMASAKI, Y., MATSUHISA, M. & MATSUOKA,
1050 T. A. 2008. PDX-1 and MafA play a crucial role in pancreatic beta-cell
1051 differentiation and maintenance of mature beta-cell function. *Endocr J,* 55**,** 235-
1052 52.
1053 KATOW, H., SUYEMITSU, T., OOKA, S., YAGUCHI, J., JIN-NAI, T., KUWAHARA, I.,
1054 KATOW, T., YAGUCHI, S. & ABE, H. 2010. Development of a dopaminergic
1055 system in sea urchin embryos and larvae. *J Exp Biol,* 213**,** 2808-19.
1056 KEMBEL, S. W., COWAN, P. D., HELMUS, M. R., CORNWELL, W. K., MORTON, H.,
1057 ACKERLY, D. D., BLOMBERG, S. P. & WEBB, C. O. .2010. 'Picante: R tools for
1058 integrating phylogenies and ecology', *Bioinformatics*, 26(11), pp. 1463–1464.
1059 KHOR, J.M., GUERRERO-SANTORO, J. & ETTENSOHN C.A. 2019. Genome-wide
1060 identification of binding sites and gene targets of Alx1, a pivotal regulator of
1061 echinoderm skeletogenesis. Development. 146(16):dev180653.KLEIN, A. M.,
1062 MAZUTIS, L., AKARTUNA, I., TALLAPRAGADA, N., VERES, A., LI,
1063 V., PESHKIN, L., WEITZ, D. A. & KIRSCHNER, M. W. 2015. Droplet barcoding
1064 for single-cell transcriptomics applied to embryonic stem cells. *Cell,* 161**,** 1187-
1065 1201.
1066 KROUK, G., LINGEMAN, J., COLON, A. M., CORUZZI, G. & SHASHA, D. 2013. Gene
1067 regulatory networks in plants: learning causality from time and perturbation.
1068 *Genome Biology,* 14.
1069 KUDTARKAR, P. & CAMERON, R. A. 2017. Echinobase: an expanding resource for
1070 echinoderm genomic information. *Database (Oxford),* 2017.
1071 LI, H., ARBER, S., JESSELL, T. M. & EDLUND, H. 1999. Selective agenesis of the
1072 dorsal pancreas in mice lacking homeobox gene Hlxb9. *Nat Genet,* 23**,** 67-70.
1073 LONGABAUGH, W. J. 2012. BioTapestry: a tool to visualize the dynamic properties of
1074 gene regulatory networks. *Methods Mol Biol,* 786**,** 359-94.

1075 LUO, Y. J. & SU, Y. H. 2012. Opposing nodal and BMP signals regulate left-right
1076 asymmetry in the sea urchin larva. *PLoS Biol,* 10**,** e1001402.
1077 LYONS, D. C., KALTENBACH, S. L. & MCCLAY, D. R. 2012. Morphogenesis in sea
1078 urchin embryos: linking cellular events to gene regulatory network states. *Wiley*
1079 *Interdiscip Rev Dev Biol,* 1**,** 231-52.
1080 MAECHLER, P. & WOLLHEIM, C. B. 1999. Mitochondrial glutamate acts as a
1081 messenger in glucose-induced insulin exocytosis. *Nature,* 402**,** 685-9.
1082 MARTIK, M. L. & MCCLAY, D. R. 2015. Deployment of a retinal determination gene
1083 network drives directed cell migration in the sea urchin embryo. *Elife,* 4.
1084 MASSRI, A. J., GREENSTREET, L., AFANASSIEV, A., ESCOBAR, A. B., WRAY, G.
1085 M., SCHIEBINGER, G. & MCCLAY, D. R. 2021. Developmental Single-cell
1086 transcriptomics in the Lytechinus variegatus Sea Urchin Embryo. Development.
1087 148(19):dev198614.
1088 MATERNA, S. C., HOWARD-ASHBY, M., GRAY, R. F. & DAVIDSON, E. H. 2006. The
1089 C2H2 zinc finger genes of Strongylocentrotus purpuratus and their expression in
1090 embryonic development. *Dev Biol,* 300**,** 108-20.
1091 MATERNA, S. C., RANSICK, A., LI, E. & DAVIDSON, E. H. 2013. Diversification of oral
1092 and aboral mesodermal regulatory states in pregastrular sea urchin embryos.
1093 *Dev Biol,* 375**,** 92-104.
1094 MCCLAY, D. R. 1986. Embryo dissociation, cell isolation, and cell reassociation.
1095 *Methods Cell Biol,* 27**,** 309-23.
1096 MCCLAY, D. R. 2004. Methods for embryo dissociation and analysis of cell adhesion.
1097 *Methods Cell Biol,* 74**,** 311-29.
1098 MCCLAY, D. R. 2011. Evolutionary crossroads in developmental biology: sea urchins.
1099 *Development,* 138**,** 2639-48.
1100 MCCLAY, D. R., MIRANDA, E. & FEINBERG, S. L. 2018. Neurogenesis in the sea
1101 urchin embryo is initiated uniquely in three domains. *Development,* 145.
1102 MCCLAY, D. R., WARNER, J., MARTIK, M., MIRANDA, E. & SLOTA, L. 2020.
1103 Gastrulation in the sea urchin. *Curr Top Dev Biol,* 136**,** 195-218.
1104 MCQUEEN, E. & REBEIZ, M. 2020. On the specificity of gene regulatory networks: How
1105 does network co-option affect subsequent evolution? *Curr Top Dev Biol,* 139**,**
1106 375-405.
1107 MITCHELL, R. K., NGUYEN-TU, M. S., CHABOSSEAU, P., CALLINGHAM, R. M.,
1108 PULLEN, T. J., CHEUNG, R., LECLERC, I., HODSON, D. J. & RUTTER, G. A.
1109 2017. The transcription factor Pax6 is required for pancreatic beta cell identity,
1110 glucose-regulated ATP synthesis, and Ca(2+) dynamics in adult mice. *J Biol*
1111 *Chem,* 292**,** 8892-8906.
1112 MIZUSAWA, N., HASEGAWA, T., OHIGASHI, I., TANAKA-KOSUGI, C., HARADA, N.,
1113 ITAKURA, M. & YOSHIMOTO, K. 2004. Differentiation phenotypes of pancreatic
1114 islet beta- and alpha-cells are closely related with homeotic genes and a group of
1115 differentially expressed genes. *Gene,* 331**,** 53-63.
1116 MONTEIRO, A. 2012. Gene regulatory networks reused to build novel traits: co-option
1117 of an eye-related gene regulatory network in eye-like organs and red wing
1118 patches on insect wings is suggested by optix expression. *Bioessays,* 34**,** 181-6.
1119 MORGULIS, M., GILDOR, T., ROOPIN, M., SHER, N., MALIK, A., LALZAR, M., DINES,
1120 M., BEN-TABOU DE-LEON, S., KHALAILY, L. & BEN-TABOU DE-LEON, S.

2019. Possible cooption of a VEGF-driven tubulogenesis program for biomineralization in echinoderms. *Proc Natl Acad Sci U S A,* 116**,** 12353-12362.

NAVALE, A. M. & PARANJAPE, A. N. 2016. Glucose transporters: physiological and pathological roles. *Biophys Rev,* 8**,** 5-9.

MUSSER J., SCHIPPERS K.J., NICKELI M., MIZZON G., KOHN A.B., PAPE C, HAMMELI J.U., WOLF F., LIANG C., HERNANDEZ-PLAZA A., ACHIM K., SCHIEBER N.L., FRANCIS W.R., VARGAS S. R., KLING S., RENKERT M., FEUDA R., GASPAR I., BURKHARDT P., BORK P., BECK M., KRESHUK A., WORHEIDE G., HUERTA-CEPAS J., Schwab Y., Moroz L.L., Arendt D. bioRxiv 758276

NESTOROWA, S., HAMEY, F. K., PIJUAN SALA, B., DIAMANTI, E., SHEPHERD, M., LAURENTI, E., WILSON, N. K., KENT, D. G. & GOTTGENS, B. 2016. A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation. *Blood,* 128**,** e20-31.

OISHI, Y & MANABE, I. Krüppel-Like Factors in Metabolic Homeostasis and Cardiometabolic Disease. Front Cardiovasc Med.5:69.

OKAZAKI, K. 1965. Skeleton formation of sea urchin larvae. V. Continuous observation of the process of matrix formation. *Exp Cell Res,* 40**,** 585-96.

OLIVERI, P., TU, Q. & DAVIDSON E.H. 2008. Global regulatory logic for specification of an embryonic cell lineage. Proceedings of the National Academy of Sciences, 105 (16) 5955-5962

PEHRSON, J. R. & COHEN, L. H. 1986. The fate of the small micromeres in sea urchin development. *Dev Biol,* 113**,** 522-6.

PERILLO, M. & ARNONE, M. I. 2014. Characterization of insulin-like peptides (ILPs) in the sea urchin Strongylocentrotus purpuratus: insights on the evolution of the insulin family. *Gen Comp Endocrinol,* 205**,** 68-79.

PERILLO, M., OULHEN, N., FOSTER, S., SPURRELL, M., CALESTANI, C. & WESSEL, G. 2020. Regulation of dynamic pigment cell states at single-cell resolution. *Elife,* 9.

PERILLO, M., PAGANOS, P., MATTIELLO, T., COCURULLO, M., OLIVERI, P. & ARNONE, M. I. 2018. New Neuronal Subtypes With a "Pre-Pancreatic" Signature in the Sea Urchin Stongylocentrotus purpuratus. *Front Endocrinol (Lausanne),* 9**,** 650.

PERILLO, M., PAGANOS, P., SPURRELL, M., ARNONE, M. I. & WESSEL, G. M. 2021. Methodology for Whole Mount and Fluorescent RNA In Situ Hybridization in Echinoderms: Single, Double, and Beyond. *Methods Mol Biol,* 2219**,** 195-216.

PERILLO, M., WANG, Y. J., LEACH, S. D. & ARNONE, M. I. 2016. A pancreatic exocrine-like cell regulatory circuit operating in the upper stomach of the sea urchin Strongylocentrotus purpuratus larva. *BMC Evol Biol,* 16**,** 117.

PREGER-BEN NOON, E. & FRANKEL, N. 2015. Evolving Genital Structures: A Deep Look at Network Co-option. *Dev Cell,* 34**,** 485-6.

QI, F., QIAN, S., ZHANG, S. & ZHANG, Z. 2020. Single cell RNA sequencing of 13 human tissues identify cell types and receptors of human coronaviruses. *Biochem Biophys Res Commun,* 526**,** 135-140.

RAFIQ, K., CHEERS, M. S. & ETTENSOHN, C. A. 2012. The genomic regulatory control of skeletal morphogenesis in the sea urchin. *Development,* 139**,** 579-90.

RAST, J. P., SMITH, L. C., LOZA-COLL, M., HIBINO, T. & LITMAN, G. W. 2006. Genomic insights into the immune system of the sea urchin. *Science,* 314**,** 952-6.

RIZZO, F., FERNANDEZ-SERRA, M., SQUARZONI, P., ARCHIMANDRITIS, A. & ARNONE, M. I. 2006. Identification and developmental expression of the ets gene family in the sea urchin (Strongylocentrotus purpuratus). *Dev Biol,* 300**,** 35-48.

RODRIGUEZ-DIAZ, R., DANDO, R., JACQUES-SILVA, M. C., FACHADO, A., MOLINA, J., ABDULREDA, M. H., RICORDI, C., ROPER, S. D., BERGGREN, P. O. & CAICEDO, A. 2011. Alpha cells secrete acetylcholine as a non-neuronal paracrine signal priming beta cell function in humans. *Nat Med,* 17**,** 888-92.

SEA URCHIN GENOME SEQUENCING, C., SODERGREN, E., WEINSTOCK, G. M., DAVIDSON, E. H., CAMERON, R. A., GIBBS, R. A., ANGERER, R. C., ANGERER, L. M., ARNONE, M. I., BURGESS, D. R., BURKE, R. D., COFFMAN, J. A., DEAN, M., ELPHICK, M. R., ETTENSOHN, C. A., FOLTZ, K. R., HAMDOUN, A., HYNES, R. O., KLEIN, W. H., MARZLUFF, W., MCCLAY, D. R., MORRIS, R. L., MUSHEGIAN, A., RAST, J. P., SMITH, L. C., THORNDYKE, M. C., VACQUIER, V. D., WESSEL, G. M., WRAY, G., ZHANG, L., ELSIK, C. G., ERMOLAEVA, O., HLAVINA, W., HOFMANN, G., KITTS, P., LANDRUM, M. J., MACKEY, A. J., MAGLOTT, D., PANOPOULOU, G., POUSTKA, A. J., PRUITT, K., SAPOJNIKOV, V., SONG, X., SOUVOROV, A., SOLOVYEV, V., WEI, Z., WHITTAKER, C. A., WORLEY, K., DURBIN, K. J., SHEN, Y., FEDRIGO, O., GARFIELD, D., HAYGOOD, R., PRIMUS, A., SATIJA, R., SEVERSON, T., GONZALEZ-GARAY, M. L., JACKSON, A. R., MILOSAVLJEVIC, A., TONG, M., KILLIAN, C. E., LIVINGSTON, B. T., WILT, F. H., ADAMS, N., BELLE, R., CARBONNEAU, S., CHEUNG, R., CORMIER, P., COSSON, B., CROCE, J., FERNANDEZ-GUERRA, A., GENEVIERE, A. M., GOEL, M., KELKAR, H., MORALES, J., MULNER-LORILLON, O., ROBERTSON, A. J., GOLDSTONE, J. V., COLE, B., EPEL, D., GOLD, B., HAHN, M. E., HOWARD-ASHBY, M., SCALLY, M., STEGEMAN, J. J., ALLGOOD, E. L., COOL, J., JUDKINS, K. M., MCCAFFERTY, S. S., MUSANTE, A. M., OBAR, R. A., RAWSON, A. P., ROSSETTI, B. J., GIBBONS, I. R., HOFFMAN, M. P., LEONE, A., ISTRAIL, S., MATERNA, S. C., SAMANTA, M. P., STOLC, V., et al. 2006. The genome of the sea urchin Strongylocentrotus purpuratus. *Science,* 314**,** 941-52.

SEBE-PEDROS, A., SAUDEMONT, B., CHOMSKY, E., PLESSIER, F., MAILHE, M. P., RENNO, J., LOE-MIE, Y., LIFSHITZ, A., MUKAMEL, Z., SCHMUTZ, S., NOVAULT, S., STEINMETZ, P. R. H., SPITZ, F., TANAY, A. & MARLOW, H. 2018. Cnidarian Cell Type Diversity and Regulation Revealed by Whole-Organism Single-Cell RNA-Seq. *Cell,* 173**,** 1520-1534 e20.

SEVERO, M. S., LANDRY, J. J. M., LINDQUIST, R. L., GOOSMANN, C., BRINKMANN, V., COLLIER, P., HAUSER, A. E., BENES, V., HENRIKSSON, J., TEICHMANN, S. A. & LEVASHINA, E. A. 2018. Unbiased classification of mosquito blood cells by single-cell genomics and high-content imaging. *Proc Natl Acad Sci U S A,* 115**,** E7568-E7577.

SHARMA, S., WANG, W. & STOLFI, A. 2019. Single-cell transcriptome profiling of the Ciona larval brain. *Dev Biol,* 448**,** 226-236.

1212 SHEKHAR, K. & MENON, V. 2019. Identification of Cell Types from Single-Cell
1213 Transcriptomic Data. *Methods Mol Biol,* 1935**,** 45-77.
1214 SHASHIKANT, T., & ETTENSOHN, C. A. 2019. Genome-wide analysis of chromatin
1215 accessibility using ATAC-seq. Methods in cell biology, 151, 219–235.
1216 SIABERT, S., FARRELL, J.A., CAZET, J.F., ABEYKOON, Y., PRIMACK, A.S.,
1217 SCHNITZLER, C.E., JULIANO, CE. 2019. Stem cell differentiation trajectories in
1218 Hydra resolved at single-cell resolution. Science. 365(6451)SLADITSCHEK, H.
1219 L., FIUZA, U. M., PAVLINIC, D., BENES, V., HUFNAGEL, L. & NEVEU, P. A.
1220 2020. MorphoSeq: Full Single-Cell Transcriptome Dynamics Up to Gastrulation in
1221 a Chordate. *Cell,* 181**,** 922-935 e21.
1222 SLOTA, L. A. & MCCLAY, D. R. 2018. Identification of neural transcription factors
1223 required for the differentiation of three neuronal subtypes in the sea urchin
1224 embryo. *Dev Biol,* 435**,** 138-149.
1225 SLOTA, L. A., MIRANDA, E., PESKIN, B. & MCCLAY, D. R. 2020. Developmental origin
1226 of peripheral ciliary band neurons in the sea urchin embryo. *Dev Biol,* 459**,** 72-78.
1227 SMITH, M. M., CRUZ SMITH, L., CAMERON, R. A. & URRY, L. A. 2008. The larval
1228 stages of the sea urchin, Strongylocentrotus purpuratus. *J Morphol,* 269**,** 713-33.
1229 SOLIMAN, S. 1983. Pharmacological control of ciliary activity in the young sea urchin
1230 larva. Effects of monoaminergic agents. *Comp Biochem Physiol C Comp*
1231 *Pharmacol Toxicol,* 76**,** 181-91.
1232 STEINMETZ PRH, AMAN A, KRAUS JEM, TECHNAU U. 2017. Gut-like ectodermal
1233 tissue in a sea anemone challenges germ layer homology. Nat Ecol Evol.
1234 Oct;1(10):1535-1542
1235 STRATHMANN, M. F. 1987. *Reproduction and development of marine invertebrates of*
1236 *the northern Pacific coast: data and methods for the study of eggs, embryos, and*
1237 *larvae,* Seattle, University of Washington Press.
1238 STUART, T., BUTLER, A., HOFFMAN, P., HAFEMEISTER, C., PAPALEXI, E.,
1239 MAUCK, W. M., 3RD, HAO, Y., STOECKIUS, M., SMIBERT, P. & SATIJA, R.
1240 2019. Comprehensive Integration of Single-Cell Data. *Cell,* 177**,** 1888-1902 e21.
1241 SUN, Z. & ETTENSOHN, C. A. 2014. Signal-dependent regulation of the sea urchin
1242 skeletogenic gene regulatory network. *Gene Expr Patterns,* 16**,** 93-103.
1243 SUN, Z. & ETTENSOHN, C. A. 2017. TGF-beta sensu stricto signaling regulates
1244 skeletal morphogenesis in the sea urchin embryo. *Dev Biol,* 421**,** 149-160.
1245 SWAPNA, L.S., MOLINARO, A.M., LINDSAY-MOSHER, N., PEARSON, B.J.,
1246 PARKINSON, J. 2018. Comparative transcriptomic analyses and single-cell
1247 RNA sequencing of the freshwater planarian Schmidtea mediterranea identify
1248 major cell types and pathway conservation. Genome Biol 19, 124
1249 TANG, F., BARBACIORU, C., WANG, Y., NORDMAN, E., LEE, C., XU, N., WANG, X.,
1250 BODEAU, J., TUCH, B. B., SIDDIQUI, A., LAO, K. & SURANI, M. A. 2009.
1251 mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods,* 6**,** 377-
1252 82.
1253 TEITELMAN, G., ALPERT, S., POLAK, J. M., MARTINEZ, A. & HANAHAN, D. 1993.
1254 Precursor cells of mouse endocrine pancreas coexpress insulin, glucagon and
1255 the neuronal proteins tyrosine hydroxylase and neuropeptide Y, but not
1256 pancreatic polypeptide. *Development,* 118**,** 1031-9.

1257 TRAPNELL, C., WILLIAMS, B. A., PERTEA, G., MORTAZAVI, A., KWAN, G., VAN
1258 BAREN, M. J., SALZBERG, S. L., WOLD, B. J. & PACHTER, L. 2010. Transcript
1259 assembly and quantification by RNA-Seq reveals unannotated transcripts and
1260 isoform switching during cell differentiation. *Nat Biotechnol,* 28**,** 511-5.
1261 TRITSCHLER, S., THEIS, F. J., LICKERT, H. & BOTTCHER, A. 2017. Systematic
1262 single-cell analysis provides new insights into heterogeneity and plasticity of the
1263 pancreas. *Mol Metab,* 6**,** 974-990.
1264 TU, Q., BROWN, C. T., DAVIDSON, E. H. & OLIVERI, P. 2006. Sea urchin Forkhead
1265 gene family: phylogeny and embryonic expression. *Dev Biol,* 300**,** 49-62.
1266 TU, Q., CAMERON, R. A. & DAVIDSON, E. H. 2014. Quantitative developmental
1267 transcriptomes of the sea urchin Strongylocentrotus purpuratus. *Dev Biol,* 385**,**
1268 160-7.
1269 WAGNER, D. E., WEINREB, C., COLLINS, Z. M., BRIGGS, J. A., MEGASON, S. G. &
1270 KLEIN, A. M. 2018. Single-cell mapping of gene expression landscapes and
1271 lineage in the zebrafish embryo. *Science,* 360**,** 981-987.
1272 WEI, Z., ANGERER, L. M. & ANGERER, R. C. 2016. Neurogenic gene regulatory
1273 pathways in the sea urchin embryo. *Development,* 143**,** 298-305.
1274 WEI, Z., ANGERER, R. C. & ANGERER, L. M. 2011. Direct development of neurons
1275 within foregut endoderm of sea urchin embryos. *Proc Natl Acad Sci U S A,* 108**,**
1276 9143-7.
1277 WEI, Z., RANGE, R., ANGERER, R. & ANGERER, L. 2012. Axial patterning interactions
1278 in the sea urchin embryo: suppression of nodal by Wnt1 signaling. *Development,*
1279 139**,** 1662-9.
1280 WEI, Z., YAGUCHI, J., YAGUCHI, S., ANGERER, R. C. & ANGERER, L. M. 2009. The
1281 sea urchin animal pole domain is a Six3-dependent neurogenic patterning center.
1282 *Development,* 136**,** 1179-89.
1283 WESSEL, G. M. & MCCLAY, D. R. 1987. Gastrulation in the sea urchin embryo requires
1284 the deposition of crosslinked collagen within the extracellular matrix. *Dev Biol,*
1285 121**,** 149-65.
1286 WOOD, N.J., T. M., WARD, E.T.M., ROWE, M.L., PERILLO, M., ARNONE, M.I. ,
1287 ELPHICK, M.R. & OLIVERI, P. 2018. Neuropeptidergic systems in pluteus larvae
1288 of the sea urchin Strongylocentrotus purpuratus: neurochemical complexity in a
1289 “simple” nervous system. *Fronties in Endocrinology*.
1290 XIMERAKIS, M., LIPNICK, S. L., INNES, B. T., SIMMONS, S. K., ADICONIS, X.,
1291 DIONNE, D., MAYWEATHER, B. A., NGUYEN, L., NIZIOLEK, Z., OZEK, C.,
1292 BUTTY, V. L., ISSERLIN, R., BUCHANAN, S. M., LEVINE, S. S., REGEV, A.,
1293 BADER, G. D., LEVIN, J. Z. & RUBIN, L. L. 2019. Single-cell transcriptomic
1294 profiling of the aging mouse brain. *Nat Neurosci,* 22**,** 1696-1708.
1295 YAGUCHI, J., TAKEDA, N., INABA, K. & YAGUCHI, S. 2016. Cooperative Wnt-Nodal
1296 Signals Regulate the Patterning of Anterior Neuroectoderm. *PLoS Genet,* 12**,**
1297 e1006001.
1298 YAGUCHI, S., YAGUCHI, J. & TANAKA, H. 2017. Troponin-I is present as an essential
1299 component of muscles in echinoderm larvae. *Sci Rep,* 7**,** 43563.
1300 YU, Z., LIAO, J., CHEN, Y., ZOU, C., ZHANG, H., CHENG, J., LIU, D., LI, T., ZHANG,
1301 Q., LI, J., YANG, X., YE, Y., HUANG, Z., LONG, X., YANG, R. & MO, Z. 2019.

1302 Single-Cell Transcriptomic Map of the Human and Mouse Bladders. *J Am Soc*
1303 *Nephrol,* 30**,** 2159-2176.
1304 YUH, C. H., LI, X., DAVIDSON, E. H. & KLEIN, W. H. 2001. Correct Expression of
1305 spec2a in the sea urchin embryo requires both Otx and other cis-regulatory
1306 elements. *Dev Biol,* 232**,** 424-38.
1307 ZARET, K. S. & GROMPE, M. 2008. Generation and regeneration of cells of the liver
1308 and pancreas. *Science,* 322**,** 1490-4.
1309 ZENG, A., LI, H., GUO, L., GAO, X., MCKINNEY, S., WANG, Y., YU, Z., PARK, J.,
1310 SEMERAD, C., ROSS, E., CHENG, LC., DAVIES, E., LEI, K., WANG, W.,
1311 PERERA, A., HALL, K., PEAK, A., BOX, A.A, SANCHEZ ALVARADO, A. 2018.
1312 Prospectively Isolated Tetraspanin+ Neoblasts Are Adult Pluripotent Stem Cells
1313 Underlying Planaria Regeneration. Cell. 2018 Jun 14;173(7):1593-1608.e20.
1314 ZHAO, J., ZHANG, S., LIU, Y., HE, X., QU, M., XU, G., WANG, H., HUANG, M., PAN,
1315 J., LIU, Z., LI, Z., LIU, L. & ZHANG, Z. 2020. Single-cell RNA sequencing reveals
1316 the heterogeneity of liver-resident immune cells in human. *Cell Discov,* 6**,** 22.

1317

1318 ## Figure supplements

1319 **Figure 1 – Figure supplement 1. Overlap of the different replicates and characterization of**
1320 **the undefined cluster. A)** UMAPs split by original identity showing the overlap of the different
1321 biological and technical replicates. **B)** Overlay UMAP showing the overlap of the different
1322 biological and technical replicates. **C)** Distribution of cells per putative broad cell type. Color-code
1323 is the same as indicated on the UMAP in Figure 1B. **C)** Table showing the percentage of WHL
1324 gene models expressed per cell type family compared to the total number of genes we identified
1325 (15,578). **E)** Violin plot showing the number of genes (nFeature_RNA) and number of molecules
1326 (nCount_RNA) across the cell type families. **F)** Dotplot of the total marker genes (69) expressed
1327 in the undefined cluster, that shows high similarity with ectodermally derived cell type families.
1328

1329 **Figure 2 – Figure supplement 1. Expression pattern of genes used to annotate the clusters.**
1330 FISH of S. purpuratus larvae with specific probes detecting the mRNAs for *Sp-Alpi* **(A)**, *Sp-IrxA*
1331 **(B)**, *SPU_002797* **(C)**, *SPU_021898* **(D)**, *Sp-ManrC1A* **(E)**, *Sp-Nkx6.1* **(F)**, *Sp-FoxP* **(G)**, *Sp-Emx*
1332 **(H)**, *Sp-Pks1* **(I)**, *SPU_016308* **(J)**, *Sp-Six1/2* **(K)**, *Sp-MacpfA2* **(L)**, *SPU_010640* **(M)**, *Sp-FoxJ2*
1333 **(N)**, *Sp-Chordin* **(O)** and *Sp-Tob* **(P)**. Nuclei are labelled with DAPI (in blue). All images are stacks
1334 of merged confocal Z sections.

1335 **Figure 2 – Figure supplement 2. ScRNA-seq predicted expression patterns of genes used**
1336 **to annotate the clusters. A)** Dotplot showing the average expression of genes used as markers
1337 to identify the cell type families. Cell type families are grouped based on their developmental
1338 origins: ectodermally derived, in blue, mesodermally derived, in red, and endodermally derived,
1339 in yellow. **B)** FISH using the antisense probe of *Sp-Hbn* (B1) paired with immunofluorescent
1340 detection of Msp130 (B2). Merged channels of both signals show localization of Sp-Hbn
1341 transcripts in Msp130 positive PMCs (B3). Nuclei are labelled with DAPI (in blue). All images are
1342 stacks of merged confocal Z sections. PMCs: Primary mesenchyme cells.

1343 **Figure 2 – Figure supplement 3. ScRNA-seq is able to detect expression patterns,**
1344 **previously undetectable by *in situ* hybridization.** Dotplot of genes known for their involvement
1345 in germline determination and maintenance showing overall enrichment of their average
1346 expression in the coelomic pouches cell type family.

1347 **Figure 2 – Figure supplement 4. Proliferation status and dynamics of the larval cell type**
1348 **families. A)** Dotplot showing the average expression of genes encoding *cdk1*, *pcna*, DNA
1349 polymerases, DNA ligases, condensins, and centromere proteins. The cell type families are
1350 grouped based on their developmental origins: ectodermally derived, in blue, mesodermally
1351 derived, in red and endodermally derived cell types, in yellow. **B)** Proliferating cells as indicated
1352 by EdU labelling. B1) Larva labelled with EdU and the skeletogenic marker Msp130. B2) Selected
1353 confocal sections of larva in oral view labelled with EdU. B3) Selected confocal sections showing
1354 the internal part of the same larva as in B2 in oral view labelled with EdU. B4) Larva labelled with
1355 EdU and the endodermal marker 5c7. Abo, Aboral ectoderm; Ap, Apical plate; Cb, Ciliary band;
1356 Es, Esophagus; Oe, Oral ectoderm; St, Stomach.

1357 **Figure 3 – Figure supplement 1. Cell type family trees of the 3 dpf pluteus larva.** Cell type
1358 family trees reconstructed using all expressed genes **(A)** and only transcription factors **(B).**

1359 **Figure 4– Figure supplement 1. Distribution of the Zinc finger TF family members across**
1360 **the cell type families.** Dotplot showing the mRNA average expression of the zinc finger
1361 transcription family members. The cell type families are grouped based on their developmental
1362 origins: ectodermally derived, in blue, mesodermally derived, in red and endodermally derived, in
1363 yellow.

1364 **Figure 4– Figure supplement 2. Differentially expressed TFs across cell type families.**
1365 Dotplot showing the average scaled expression of differentially expressed TFs with p-value less
1366 than 0.5 among the 21 clusters. The cluster order is according to the transcription factor-based
1367 tree as depicted in Figure 3-Figure supplement 1B.

1368 **Figure 5– Figure supplement 1. Reconstruction of the molecular signature of the gut at a**
1369 **single cell resolution.** Dotplot showing the regionalized average expression of marker genes
1370 labeling specific endodermal domains. The scRNA-seq prediction recapitulates the previously
1371 described compartmentalization of gene expression along the digestive tract of the 3 dpf sea
1372 urchin larva.

1373 **Figure 6- Figure supplement 1. Subclustering of the skeletal cell type family.** Dotplot
1374 showing the mRNA average expression of 22 skeletal gene markers and their distribution among
1375 5 skeletal subtypes.

1376 **Figure 6- Figure supplement 2. Subclustering of the immune cells cluster. A)** UMAP showing
1377 8 distinct immune cells populations as revealed by our subclustering analysis. **B)** Dotplot showing
1378 gene markers expressed in the different immune cell population. Pigment cells gene markers are
1379 highlighted in purple (box), while non pigmented globular cells are evident in yellow (box).

1380 **Figure 6- Figure supplement 3. Co-localization of *Sp-Fgf9/16/20* and *Sp-FgfR1* in the**
1381 **cardiac sphincter region.** FISH using antisense probes for *Sp-Fgf9/16/20* (magenta) and *Sp-*
1382 *FgfR1* (green). Nuclei are labelled with DAPI (in blue). All images are stacks of merged confocal
1383 Z sections. Cs, cardiac sphincter; Es, esophagus; M, mouth.

1384 **Figure 7- Figure supplement 1. Pancreatic genes expression significance testing in the**
1385 **"post oral and lateral neurons" subcluster.** Barplots showing the number of genes expressed
1386 in the actual data and the randomized data.