

RESEARCH ARTICLE

Single-cell transcriptome analysis of the zebrafish embryonic trunk

Sanjeeva Metikala^{1,2}, Satish Casie Chetty^{1,3}, Saulius Sumanas^{1,2,4*}

1 Division of Developmental Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, United States of America, **2** Department of Pathology and Cell Biology, USF Health Heart Institute, University of South Florida, Tampa, FL, United States of America, **3** Molecular and Developmental Biology Graduate Program, University of Cincinnati, Cincinnati, OH, United States of America, **4** Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, United States of America

☞ These authors contributed equally to this work.

* ssumanas@usf.edu



Abstract

During embryonic development, cells differentiate into a variety of distinct cell types and subtypes with diverse transcriptional profiles. To date, transcriptomic signatures of different cell lineages that arise during development have been only partially characterized. Here we used single-cell RNA-seq to perform transcriptomic analysis of over 20,000 cells disaggregated from the trunk region of zebrafish embryos at the 30 hpf stage. Transcriptional signatures of 27 different cell types and subtypes were identified and annotated during this analysis. This dataset will be a useful resource for many researchers in the fields of developmental and cellular biology and facilitate the understanding of molecular mechanisms that regulate cell lineage choices during development.

OPEN ACCESS

Citation: Metikala S, Casie Chetty S, Sumanas S (2021) Single-cell transcriptome analysis of the zebrafish embryonic trunk. PLoS ONE 16(7): e0254024. <https://doi.org/10.1371/journal.pone.0254024>

Editor: Michael Klymkowsky, University of Colorado Boulder, UNITED STATES

Received: March 8, 2021

Accepted: June 17, 2021

Published: July 7, 2021

Copyright: © 2021 Metikala et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The original sequence files for scRNA-seq have been deposited to NCBI GEO database under accession number GSE152982. All other relevant data are within the manuscript and its [Supporting Information](#) files.

Funding: This research was supported by the awards from the National Institutes of Health R21 AI128445, R01 HL134815 and R01 HL153005 to S.S. and the award from American Heart Association AHA 19POST34400016 to S.M. The funders had no role in study design, data collection

Introduction

The commitment of stem cells to distinct lineages is a fundamental process that underpins embryonic development. At a molecular level, a wide array of spatiotemporally regulated signaling molecules, morphogen gradients and other factors (such as physical forces) drive changes in gene expression, which guide cells down very specific lineage trajectories. Thus, understanding the dynamics of gene expression in cell populations over time is central to mapping the paths taken by cells during differentiation. Technologies such as quantitative PCR and high throughput sequencing technologies, which have emerged over the past couple of decades, have enabled scientists to probe some of these key questions in developmental biology. While traditional 'bulk' RNA-seq analysis can efficiently reveal transcriptional variation between different organs or organisms (e.g., wt vs. mutant), subtle changes in gene expression levels at cellular resolution cannot be achieved using this method.

In recent years, the emergence and rapid advancement of single-cell RNA sequencing (scRNA-seq) technology in combination with advances in machine learning have provided unprecedented insight into global transcriptional dynamics across different cell types [1]. The ability to capture the transcriptional information of hundreds of thousands of cells of different identities over time makes scRNA-seq an invaluable tool for dissecting cellular heterogeneity

and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

during organogenesis. Analysis of cell fate transitions at a transcriptomic level has been made possible by scRNA-seq analysis and has led to new discoveries across many fields in biomedical science. This powerful technology also has the potential to reveal transcriptomic signatures of rare and uncharacterized cell populations in disease conditions, which could revolutionize treatment strategies [2–4]. Additionally, the development of several free analytical software packages like Seurat and Monocle, which have been created to mine and analyze scRNA-seq data, has greatly facilitated research utilizing scRNA-seq [5–8].

The zebrafish (*Danio rerio*) embryo has emerged as an excellent model for studying vertebrate development due to their rapid external development, optical transparency, and high fecundity from a single mating. Furthermore, the signaling pathways that drive developmental process in zebrafish are conserved in higher vertebrates. Recent studies have used single-cell RNA-seq to characterize the transcriptomic and cellular diversity of selected tissue types [9–11]. A single cell transcriptome atlas for zebrafish embryos has been reported which encompass one to five days of zebrafish development [9]. These datasets will be undoubtedly important for further studies of cell type diversity and pathways regulating choices between different lineages. However, currently available data cover only a limited number of zebrafish embryonic stages.

Here we performed single-cell transcriptomic analysis of cells isolated from the trunk region of zebrafish embryos at 30 hpf (hours post fertilization). Many important developmental processes, including definitive hematopoiesis, angiogenesis, and organogenesis are taking place at this time, and yet scRNA-seq at this developmental stage has not been previously performed. We focused on the trunk region because it is expected to contain all major cell types of different germ layers, including different subtypes of vascular endothelial and hematopoietic cells, as well as progenitors of internal organs. At the same time, we wanted to avoid additional complexity of cell types associated with the central nervous system and craniofacial tissues which were not the focus of this analysis. 20,589 single cells were isolated from the zebrafish trunk region, resulting in a higher number of cells per cluster compared with previous analyses. We present a description of 27 transcriptionally distinct populations of cells, identities of which were confirmed using previously published *in situ* expression data. This dataset will add to the growing database of zebrafish single-cell transcriptome data that is being generated by multiple labs in the zebrafish community. Together with previously published data, this resource will provide valuable transcriptional information on different populations of cells which could be mined and interrogated by researchers.

Methods

Embryo dissociation

Zebrafish embryo experiments were performed under animal protocol IACUC2019-0022, approved by the Institutional Animal Care and Use Committee at the Cincinnati Children's Hospital Medical Center. Wild-type AB embryos at 30 hpf were anesthetized in 0.002% Tricaine (Sigma) and trunks of 30 embryos were dissected using a pair forceps and immediately placed in a 1.5 ml Eppendorf tube with embryo media on ice. Trunks were then dissociated into a single-cell suspension using a cold protease tissue dissociation protocol [12]. Approximately 20,589 cells were loaded and approximately 10,000 cells were recovered with a multiplet rate of ~7.6%. Chromium Single Cell 3' Reagent Kits v2 was used (10x Genomics, Pleasanton, CA). 12 cDNA amplification cycles were used to generate cDNA. Sequencing parameters at a minimum were as follows: Read1, 26 cycles; i7 Index, 8 cycles; i5 Index, 0 cycles and Read2, 98 cycles. The sequencing library was sequenced on the HiSeq 2500 sequencer (Illumina, San Diego, CA) using one flow cell of paired-end 75 bp reads, generating 240–300 million total reads at the CCHMC DNA Sequencing core.

Single-cell cDNA library preparation and computational analysis

Single cells were captured and processed for RNA-seq using the Chromium platform (10x Genomics) at the CCHMC Gene Expression Core facility. RNA-seq was performed at the CCHMC DNA Sequencing core on Illumina HiSeq2500 sequencer using one flow cell of paired-end 75 bp reads, generating 240–300 million total reads.

Cell Ranger version 2.2.0 was utilized for processing and de-multiplexing raw sequencing data [13]. Raw basecall files were first converted to the fastq format, and subsequently the sequences were mapped to the *Danio rerio* genome (version Zv11) to generate single-cell feature counts. Downstream analysis of the gene count matrix generated by CellRanger was performed in R version 3.6.0, using Seurat version 3.1 [5,14]. The gene counts matrix was loaded into Seurat and a Seurat object was created by filtering cells which only expressed more than 200 genes and filtering genes that were expressed in at least 3 cells. Additionally, as an extra quality-control step, cells were filtered out (excluded) based on the following criteria: <400 or >2,500 unique genes expressed, or >5% of counts mapping to the mitochondrial genome. This resulted in 20,279 cells in the dataset. Reads were normalized by the “LogNormalize” function that normalizes gene expression levels for each cell by the total expression, multiplies the value by a scale factor of 10⁴ and then log-transforms the result. The top 2000 highly variable genes were calculated these genes were used for downstream analysis. Prior to dimensionality reduction, a linear transformation was performed on the normalized data. Unwanted cell-cell variation driven by mitochondrial gene expression was “regressed out” during scaling.

Dimensionality reduction was performed on the entire dataset using principal component analysis (PCA) using the list of highly variable genes generated above. The top 40 principal components which explained more variability (than expected by chance) were identified based on PC heatmaps, the JackStrawPlot and PCellbowPlot. 22 cell clusters were generated (by the default Louvain algorithm) using 40 PCs and a resolution of 0.5. UMAP dimensional reduction was utilized to visualize clusters [15]. Following clustering, genes differentially expressed in each of the clusters were determined using a method of differential expression analysis based on the non-parametric Wilcoxon rank sum test. Genes were then filtered based on being detected in $\geq 25\%$ of cells within a cluster and a Bonferroni adjusted p-value <0.05. Based on the lists of differentially expressed genes ordered by average log fold change, clusters were assigned specific cell identities. Visualization of specific gene expression patterns across groups on UMAP and violin plots was performed using functions within the Seurat package.

Sub-clustering of endothelial and endodermal/pronephric clusters

To identify heterogeneity within the endothelial and endodermal/pronephric cell populations, the two clusters were converted into separate Seurat objects and highly variable genes were calculated. A linear transformation was performed again whilst removing unwanted variation driven by mitochondrial gene expression. The top 40 significant principal components were selected for UMAP dimensionality reduction, using a resolution of 1.0 for clustering. Based on these parameters, there appeared to be 3 transcriptionally distinct sub-populations of endothelial cells, and 4 sub-populations of endodermal/pronephric cells.

Results

To analyze the transcriptional profiles of different cell types in the zebrafish trunk region, the trunk portion of zebrafish embryos was manually dissected at 30 hpf. The cells were dissociated using previously established protocols and subjected to single-cell RNA-seq analysis using the Chromium (10x Genomics) platform. Cells were clustered using Seurat, which resulted in the identification of 22 distinct cell populations (Figs 1 and 2A). Subsequently, cell identities

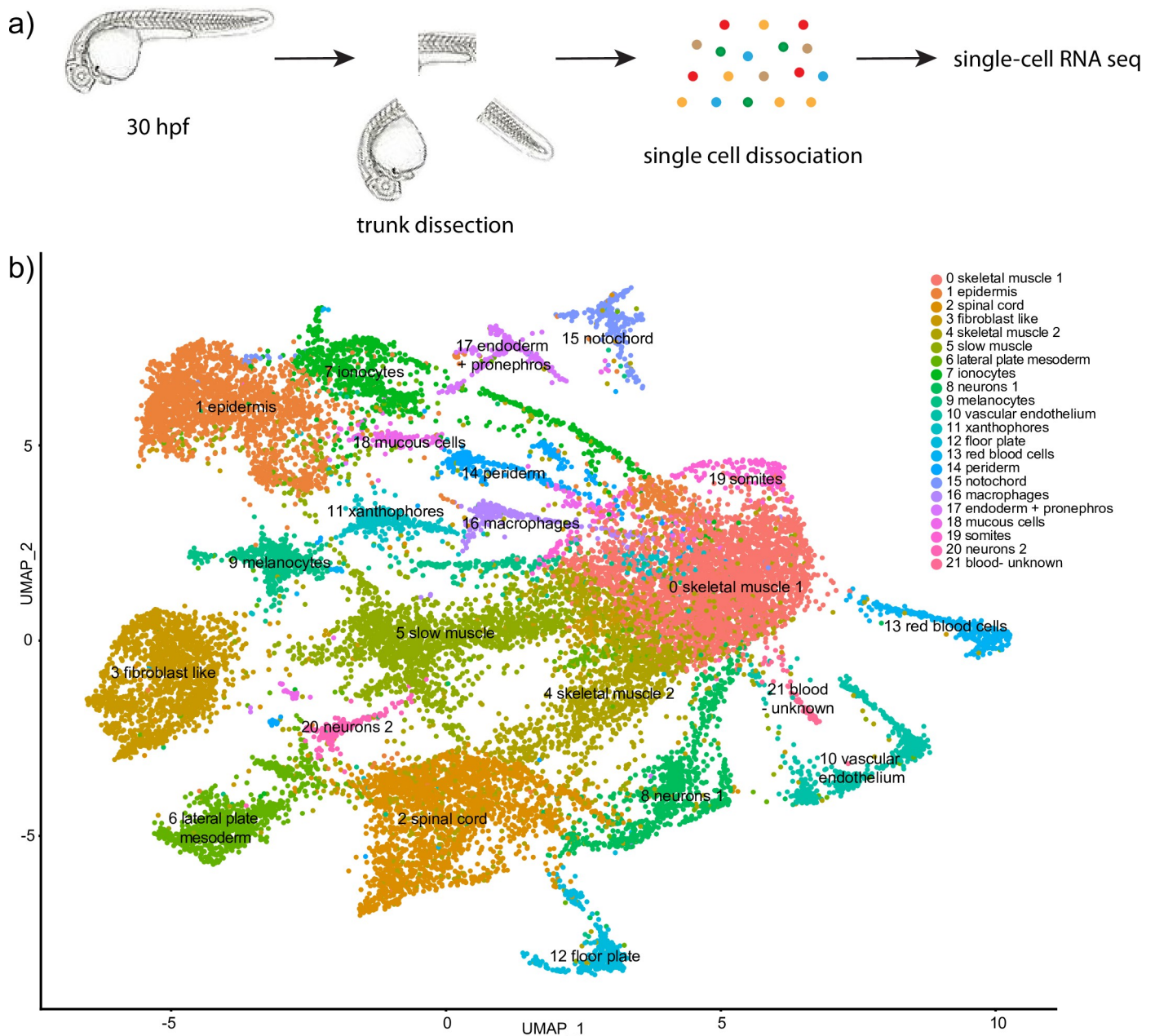


Fig 1. Single-cell RNA-seq analysis of trunk region in 30 hpf wild type embryos. (a) A diagram showing trunk dissection and single cell dissociation followed by single-cell RNA-seq analysis. (b) UMAP plot of 20,279 cells identified a total of 22 different cell clusters.

<https://doi.org/10.1371/journal.pone.0254024.g001>

were assigned based on top differentially expressed (marker) genes which were significantly enriched in each cluster (S1 Table and Fig 2). Typically, we used 5–15 top marker genes from each cluster to assign cell identities. As an additional resource, we have also generated a table showing the expression of any annotated gene in the zebrafish genome in all cell clusters (S2 Table).

Ectodermal lineages

Nine different cell types which included two different epidermal lineages (annotated as epidermis and periderm), ionocytes, mucous cells, melanocytes, xanthophores, floor plate, spinal

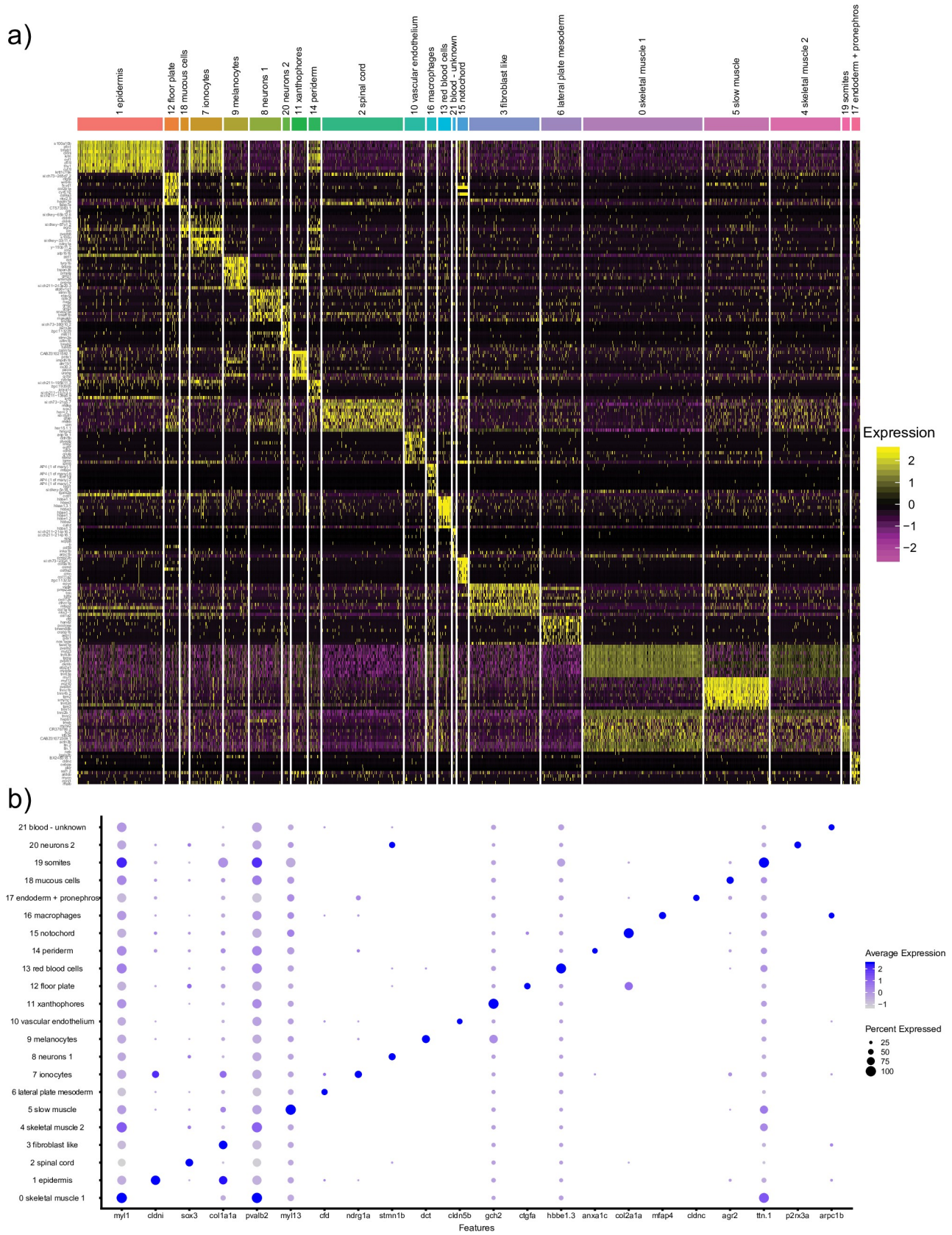


Fig 2. Expression of top marker genes in different cell clusters. (a) A heatmap showing expression of marker genes expression in different clusters. (b) A dot plot showing the expression of selected genes in different clusters.

<https://doi.org/10.1371/journal.pone.0254024.g002>

cord and two subsets of neurons were identified based on differential expression analysis (S1 Table and Fig 1).

Top marker genes for epidermis (cell cluster 1) include *profilin 1* (*pfn1*), *claudin i* (*cldni*) and *keratin 4* (*krt4*) which all share expression in the zebrafish epidermis [16–18] (Fig 3A and S1 Table). Top marker genes for periderm (cell cluster 14), include *annexin A1c* (*anxa1c*), *krt4* and *krt5*, all which have reported expression patterns in the periderm or epidermis [16,17] (Fig 3B and S1 Table). There was a significant overlap of marker genes for both populations, such as *krt4* and *type I cytokeratin, enveloping layer, like* (*cyt1l*), expressed in both cell populations. Further research will be needed to identify cell identity or functional differences between the two epidermal subpopulations.

Top marker genes for the spinal cord (cell cluster 2) included *SRY-box transcription factor 3* (*sox3*), *hairy-related 4*, *tandem duplicate 2* (*her4.2*) and *glial fibrillary acidic protein* (*gfap*), all of which share expression in the spinal cord [16] (Fig 4A and S1 Table).

Two different cell populations with a neuronal signature were identified. Top marker genes for cell population 8 included *stathmin 1b* (*stmn1b*) and *ELAV like neuron-specific RNA binding protein 3* (*elavl3*), both of which are known neuronal markers [17,19] (Fig 4B and S1 Table). Top genes for the cell population 20 included *purinergic receptor P2X, ligand-gated ion channel, 3a* (*p2rx3a*) and *stathmin 2a* (*stmn2a*), which both label a subset of neurons (Fig 4C and S1 Table). *p2rx3a* has been reported to label primary sensory neurons (including Rohon-Beard) [17,20]. While some marker genes such as *guanine nucleotide binding protein, gamma 3* (*gng3*) labeled both neuronal populations, many other top marker genes were distinct between the two populations, suggesting that they represent different subtypes. It is likely that each of these populations corresponds to a specific neuronal subtype, and further investigation will be needed to characterize these subtypes.

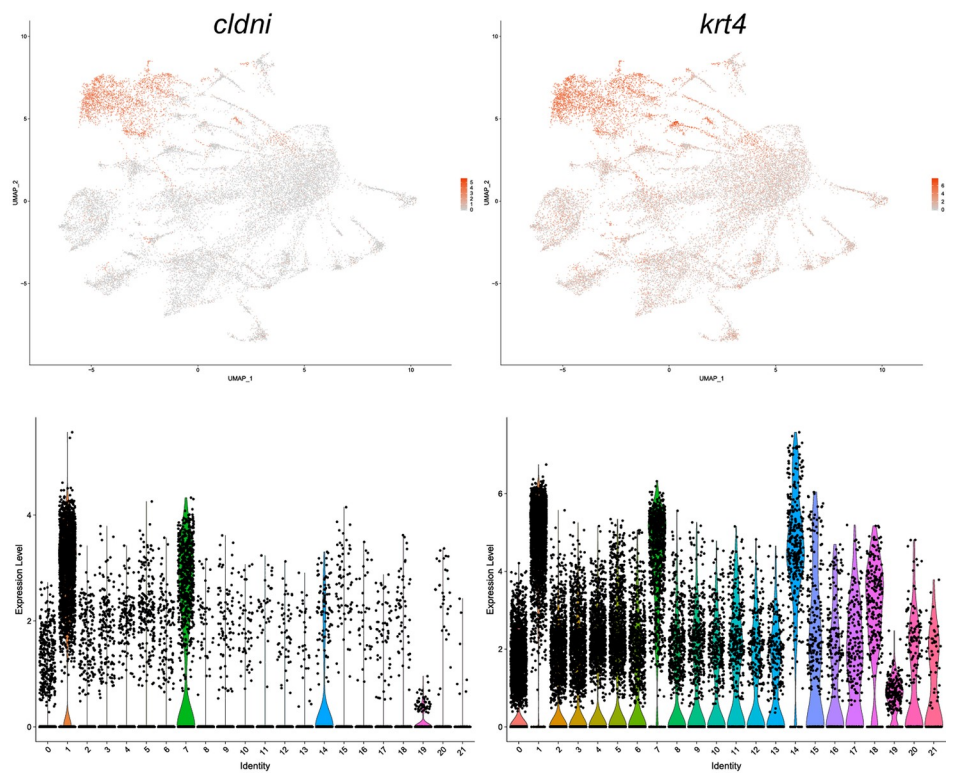
A population of cells corresponding to melanocytes was identified (cell cluster 9); genes with high expression in this group included *dopachrome tautomerase* (*dct*) and *premelanosome protein a* (*pmela*) (Fig 5A and S1 Table). Both genes have been shown to be expressed in melanocytes in zebrafish [16,21]. Top marker genes for ionocytes (cell cluster 7) included *N-myc downstream regulated 1a* (*ndrg1a*) and *ATPase Na⁺/K⁺ transporting subunit beta 1b* (*atp1b1b*) [16,22] (Fig 5B and S1 Table). The top marker genes for mucus secreting cells (cell cluster 18) included *anterior gradient 2* (*agr2*) and *pvalb8*, known markers for epidermal mucus secreting cells [17,23,24] (Fig 5C and S1 Table).

The top marker genes for cell cluster 12 were *cellular communication network factor 2a* (*ccn2a/ctgfa*) and *wingless-type MMTV integration site family, member 4b* (*wnt4b*) which are known to be expressed in the floor plate [25,26] (Fig 5D and S1 Table). Lastly, top marker genes for cell cluster 11 included *GTP cyclohydrolase 2* (*gch2*) and *peroxiredoxin 1* (*prdx1*), which are known to mark migrating neural crest cells that largely correspond to xanthophore precursors [17,27,28] (Fig 5E and S1 Table).

Endodermal lineage cell populations

We identified only a single cluster #17 that corresponds to the endodermal lineages. The top genes in this cell population included *claudin c* (*cldnc*) and *aldolase-b* (*aldob*) (S1 Table and Fig 6C). *cldnc* is a tight junction protein and *aldob* is a glycolytic enzyme involved in energy production which are expressed in the digestive tract, and liver [17,18]. Some of the marker genes (including *cldnc* and *aldob*) also have been reported to show expression in the

a) 1 epidermis



b) 14 periderm

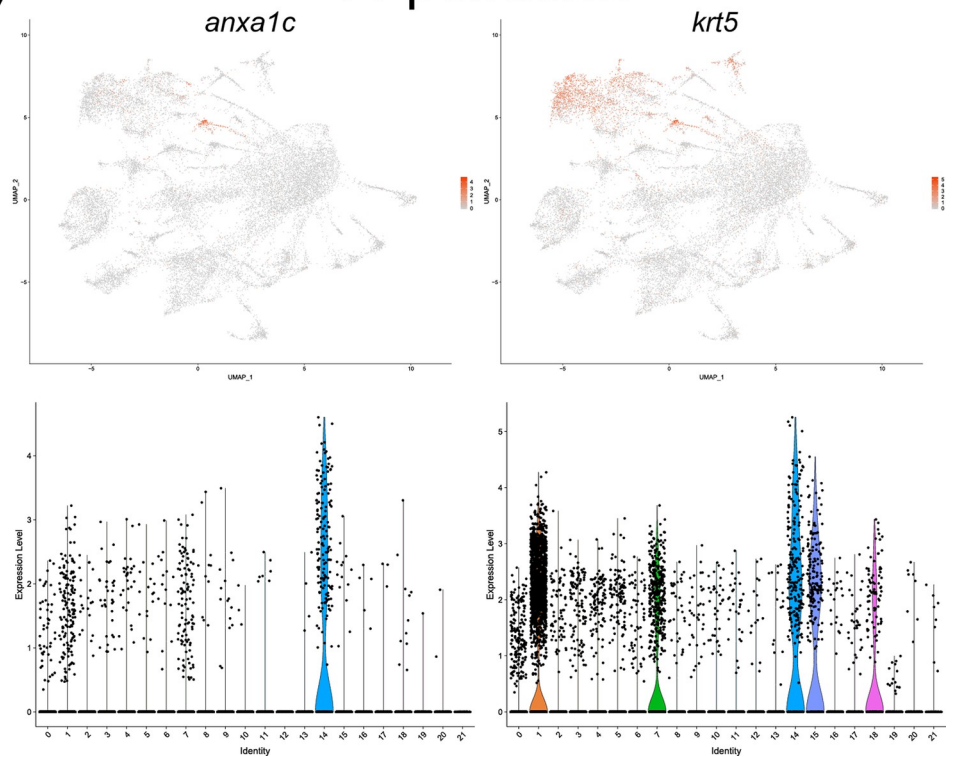


Fig 3. UMAP and violin plots showing the expression of (a) *cldni* and *krt4*, top markers for epidermis (cell cluster #1); (b) *anxa1c* and *krt5*, top markers for periderm (cluster #14).

<https://doi.org/10.1371/journal.pone.0254024.g003>

pronephric ducts [17,18]. Therefore, we cannot exclude a possibility that pronephric cells are also included in this cluster.

We hypothesized that this cluster may include cells of different identities such as pronephric, liver and intestinal progenitors, thus we attempted to subcluster this cell population further using Seurat. Four subclusters were identified (Fig 6A and 6B and S3 Table). Two of them included top marker genes enriched in pronephros, intestine and/or liver. Thus, subcluster 0 included *dap1b*, *atp1b1a* and *selenow1* (all specific to pronephros) and *gstp1*, *agr2*, *gamt* (known expression in the intestine and enteroendocrine cells). Subcluster 2 showed top

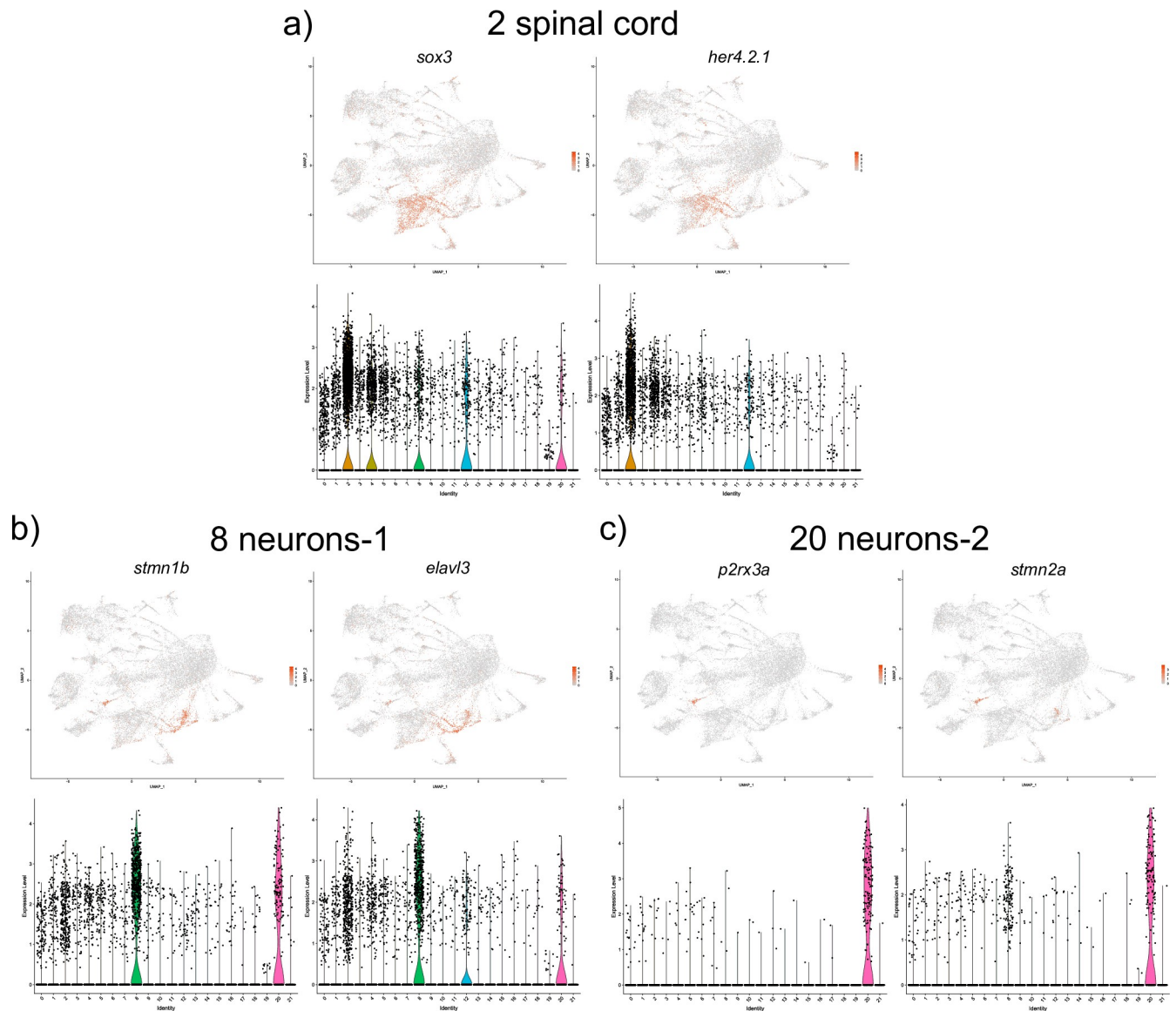


Fig 4. UMAP and violin plots showing the expression of selected marker genes for spinal cord (cluster #2, a), neurons-1 (cluster #8, b) and neurons-2 (cluster #20, c).

<https://doi.org/10.1371/journal.pone.0254024.g004>

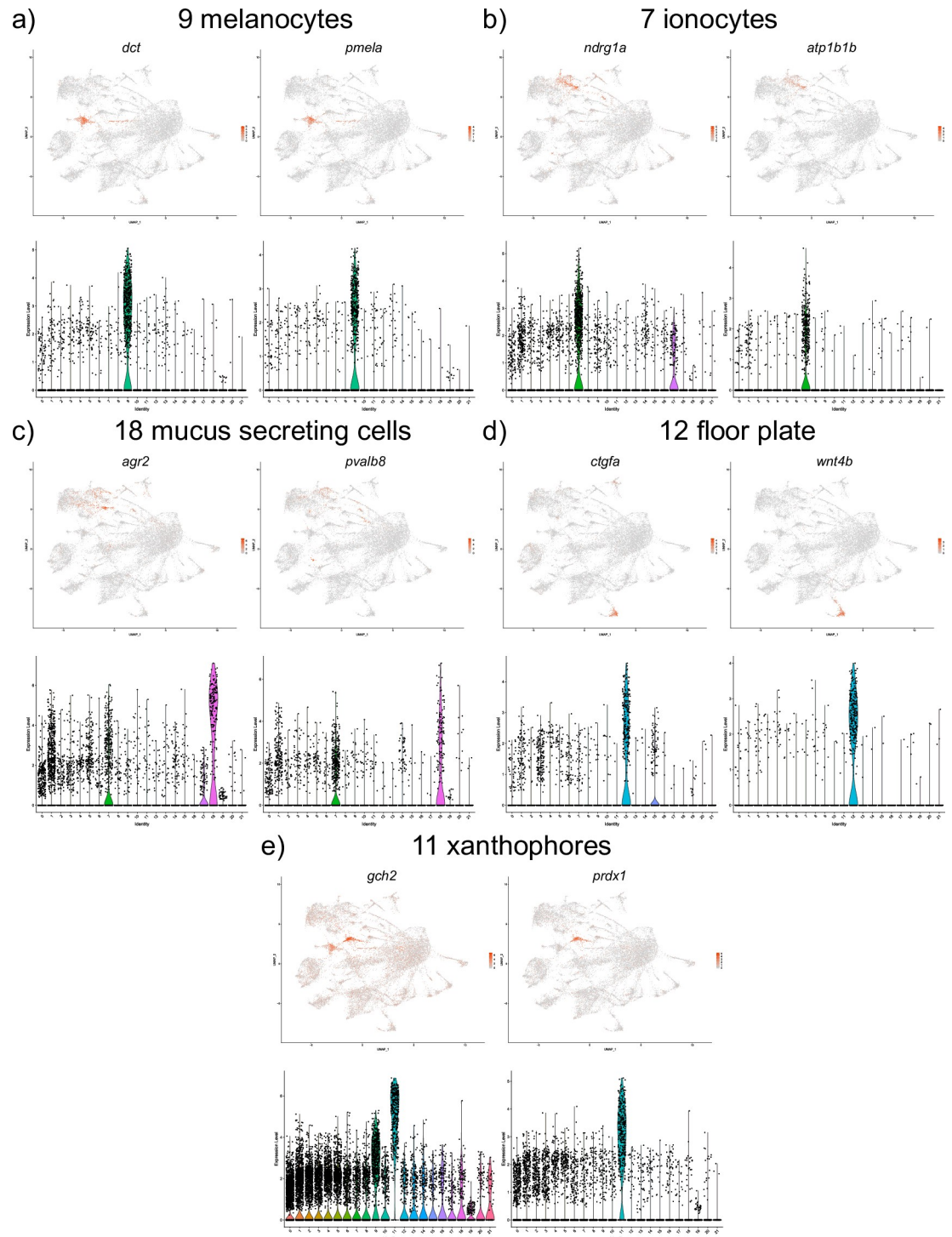


Fig 5. UMAP and violin plots showing the expression of selected marker genes for melanocytes (cluster #9, a), ionocytes (cluster #7, b), mucus secreting cells (cluster #18, c), floor plate (cluster #12, d) and xanthophores (cluster #11, e).

<https://doi.org/10.1371/journal.pone.0254024.g005>

marker genes *vdrb*, *osr2*, *sgk1*, *pabpc1a*, *col18a1a* with known expression in pronephros and digestive system. Because of the significant overlap in marker expression, it was not possible to unambiguously identify the identity of these cell populations. The remaining two cell

17 endoderm + pronephros

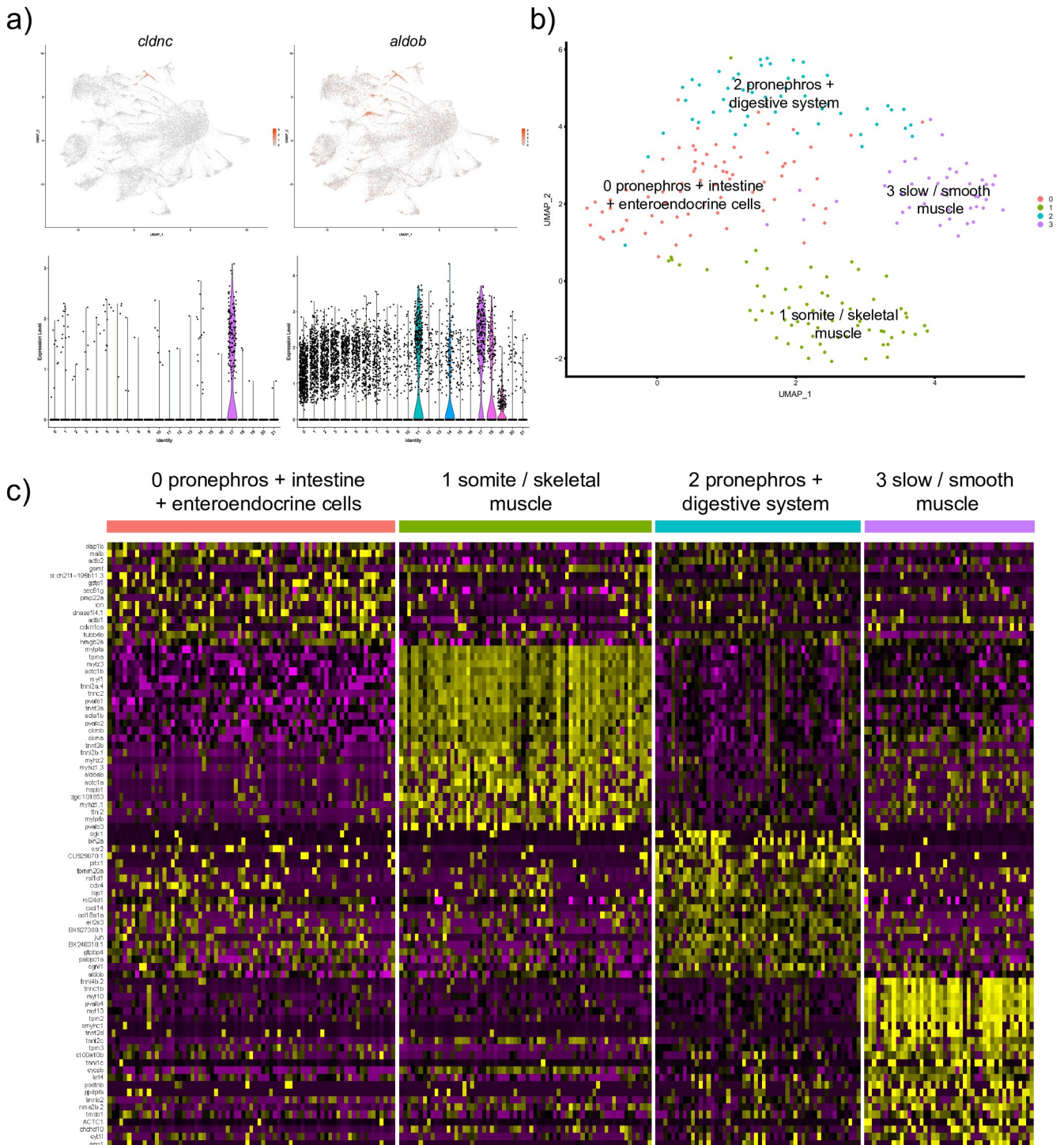


Fig 6. Analysis of the endodermal and pronephros cluster #17. (a) UMAP and violin plots showing the expression of top marker genes, *cldnc* and *aldob*. (b) UMAP plot showing further subclustering of this group that resulted in 4 cell subclusters. (c) A heatmap showing expression of marker genes in the subclusters. Note that muscle subclusters #1 and 3 are also positive for endodermal/pronephros marker expression.

<https://doi.org/10.1371/journal.pone.0254024.g006>

populations expressed genes of somite/skeletal muscle (marker genes *mylz3*, *myl1*) and slow/smooth muscle identities (marker genes *myl13*, *myl10*), in addition to the endodermal markers. This observation is intriguing and suggests a possibility that these cells are of mixed endodermal/muscle identity. It is unlikely that these cells are doublets because all subclusters show similar gene counts (number of genes expressed) and unique molecular identifier (UMI) counts (i.e., number of unique transcripts) per cell (data not shown). These clusters could potentially correspond to differentiating intestinal smooth muscle cells, although further studies are needed to confirm their identities.

Mesodermal lineage cell populations

Lateral plate mesoderm. 13 different cell clusters of mesodermal origin were identified. Several clusters corresponded to the lateral plate mesoderm derived lineages. The top genes from cell cluster 10 included *aquaporin 1a.1* (*aqp1a.1*) and *claudin 5b* (*cldn5b*), both expressed in vascular endothelial cells [8,29] (Fig 7A).

Vascular endothelial cells are known to exhibit substantial diversity. Arterial, venous, lymphatic progenitors can be distinguished as early as 24 hpf or even earlier stages [30]. Because only a single vascular endothelial cluster was identified during the initial clustering, we performed further sub clustering of vascular endothelial cells using Seurat. Three different subpopulations were identified, which included arterial and venous cells, based on the signature of marker genes (*aqp1.a1*, *aqp8a.1*, *cldn5b*, *hey2* and others for arterial cells; *stab2*, *dab2*, *lyve1b* for venous cells) (Fig 7B–7D). Multiple other marker genes were identified in each population, some of which are likely to be novel arterial and venous marker genes (S4 Table). The third population was enriched in marker genes specific for muscle cells, including *pvalb2*, *mylz3*, *myl1* and others. Intriguingly, expression of vascular endothelial specific genes such as *cdh5* was also observed within this population. Previous studies have argued that somites contribute to vascular endothelial cells [31–34]. It is possible that this cell population corresponds to these transitional somite-derived endothelial cells. A plausible alternative explanation would be that some perivascular cells were tightly attached and failed to separate during cell dissociation, resulting in doublets of mixed identity. This possibility is unlikely, though, because the scRNA-seq data were filtered to exclude potential doublet cells. There was no significant difference in average number of expressed genes or UMIs between all four endothelial subclusters, further suggesting that the endothelial cells with muscle marker expression (EC- muscle cells subpopulation) are unlikely to be doublets.

We identified three different blood cell cluster groups from our analysis. Cell cluster 13 corresponds to red blood cells (RBCs), based on the marker gene hemoglobin, beta embryonic 1.3 (*hbbe1.3*) and hemoglobin alpha embryonic-3 (*hbae3*) expression [18,35–37] (S1 Table and Fig 8A). The top marker genes for cell cluster 16 included *microfibril associated protein 4* (*mfap4*) and *lysozyme g-like 1* (*lygl1*), both known to label macrophages [17,38] (S1 Table and Fig 8B). Lastly, cell cluster 21 appears to correspond to a distinct hematopoietic population. Expression of top marker genes including *cd59* and *actin related protein 2/3 complex, subunit 1B* (*arpc1b*) (S1 Table and Fig 8C) has been reported in blood cells [39,40]. Further studies will be needed to confirm the identity of this population.

Cell cluster #6 corresponds to a poorly understood cell population that is likely derived from the lateral plate mesoderm. Its top marker genes include *complement factor D* (*cfD*), *heart and neural crest derivatives expressed 2* (*hand2*) (S1 Table and Fig 8D), which are known to have a bilateral expression along the yolk sac extension at this stage (30 hpf) [17,41]. At earlier somitogenesis stages, *hand2* has a prominent expression in the LPM [41]. Some additional top markers genes in this population such as *twist1a* also share LPM expression during somitogenesis stages [42].

10 vascular endothelium

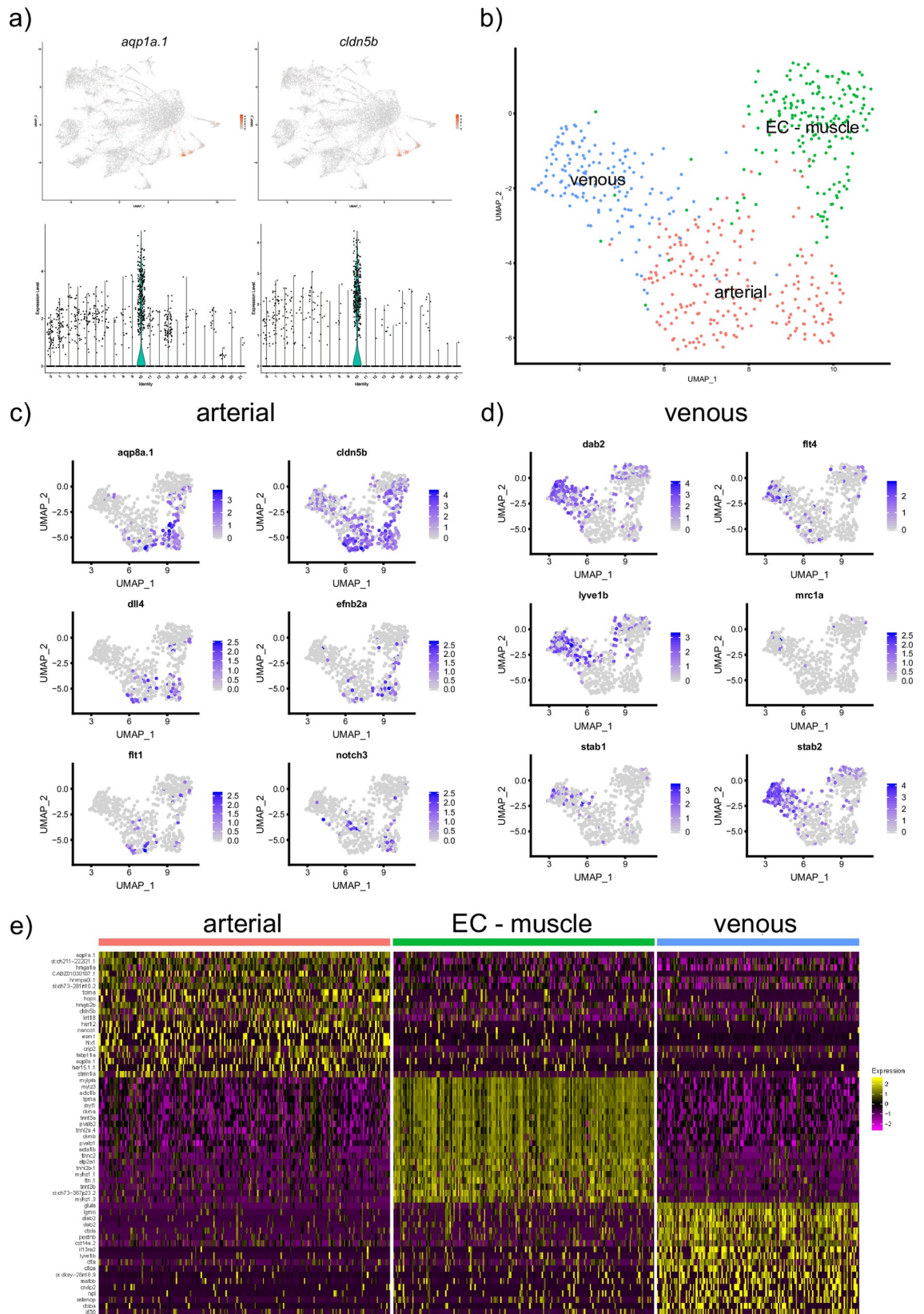


Fig 7. Analysis of the vascular endothelium cluster #10. (a) UMAP and violin plots showing the expression of top marker genes, *aqp1a.1* and *cldn5b*. (b) UMAP plot showing further subclustering of this group. 3 subclusters which show transcriptional signature of arterial, venous and EC—muscle were identified. (c,d) UMAP plot showing the expression of selected arterial and venous specific markers within the vascular endothelial cluster. (e) A heatmap showing expression of marker genes in the subclusters.

<https://doi.org/10.1371/journal.pone.0254024.g007>

Paraxial and axial mesoderm. Two different subsets of fast skeletal muscle were identified. The top markers in cell cluster 4 included *myosin*, *light chain 1*, *alkali; skeletal*, *fast (myl1)* and *heat shock protein, alpha-crystallin-related, 1 (hspb1)*, both known to in skeletal muscle cells [43,44] (S1 Table and Fig 9A). Cluster 0 also had multiple genes expressed in the myotome and skeletal muscle cells including parvalbumin 2 (*pvalb2*) and *myosin, light polypeptide 3, skeletal muscle (myl3)* [24,45] (S1 Table and Fig 9B). There was a significant gene expression overlap between the two groups of muscle cells, and the marker genes *myl1*, *myl3*, *pvalb2* were among the top ten marker genes for both cell groups. A cell cluster 5 was characterized

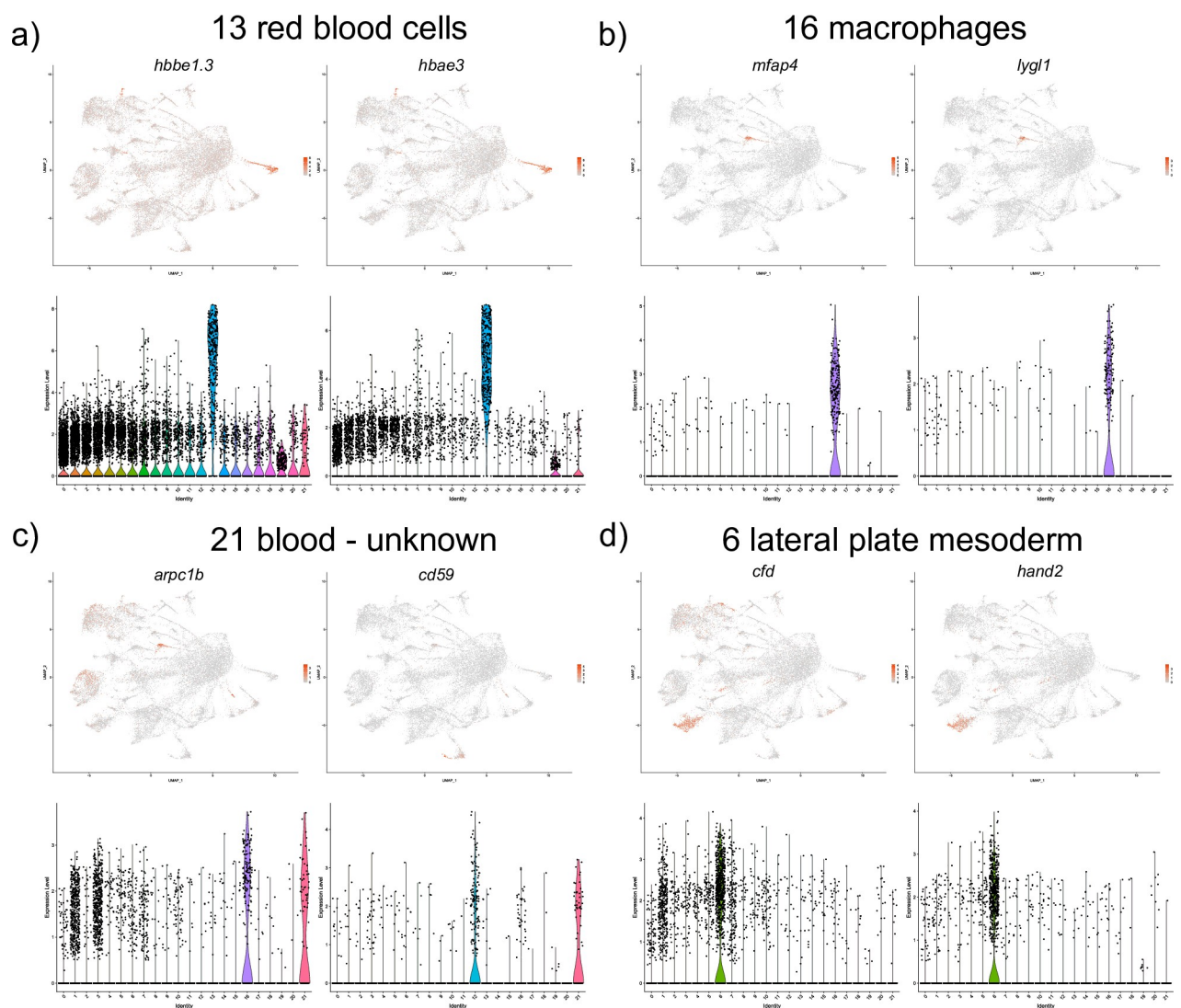


Fig 8. UMAP and violin plots showing the expression of selected marker genes for red blood cells, (cluster #13, a), macrophages (cluster #16, b), blood cells of unknown identity (cluster #21, c) and lateral plate mesoderm (cluster #6, d).

<https://doi.org/10.1371/journal.pone.0254024.g008>

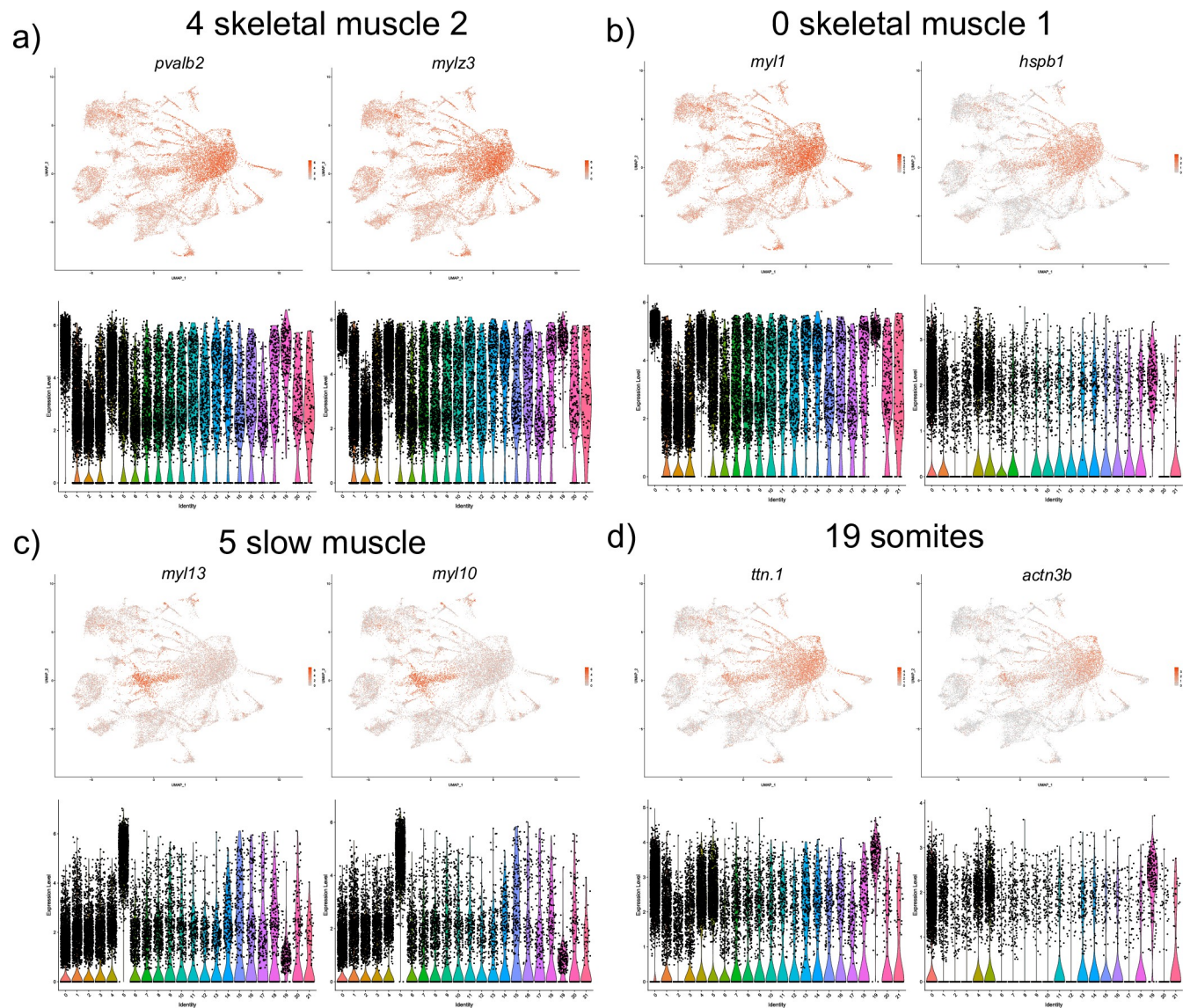


Fig 9. UMAP and violin plots showing the expression of selected marker genes for two skeletal muscle groups (clusters #4 and #0, a,b), slow muscle (cluster #5, c) and somites (cluster #19, d).

<https://doi.org/10.1371/journal.pone.0254024.g009>

by the expression of top marker genes *myosin*, *light chain 13* (*myl13*) and *myosin*, *light chain 10* (*myl10*), known markers of slow muscle cells [45] (S1 Table and Fig 9C). Cell cluster 19 included marker genes *titin* (*ttn.1* and *ttn.2*), and *actinin alpha 3b* (*actn3b*), known to be expressed in the somites and skeletal muscle cells [46,47] (S1 Table and Fig 9D). Cell cluster 3 included marker genes with a known expression in fibroblast-like cells present at the myotendinous junction including the *transforming growth factor*, *beta-induced* (*tgfb1*) and *collagen col1a1b* (S1 Table and Fig 10A). And lastly, a cell cluster #15 corresponded to the notochordal cells based on the expression of top marker genes *col2a1a* and *col9a1b* [16,48] (S1 Table and Fig 10B).

Discussion

In the current study we have identified 22 distinct cell clusters two of which were subclustered further resulting in the total of 27 cell groups with unique transcriptional signatures. Many of

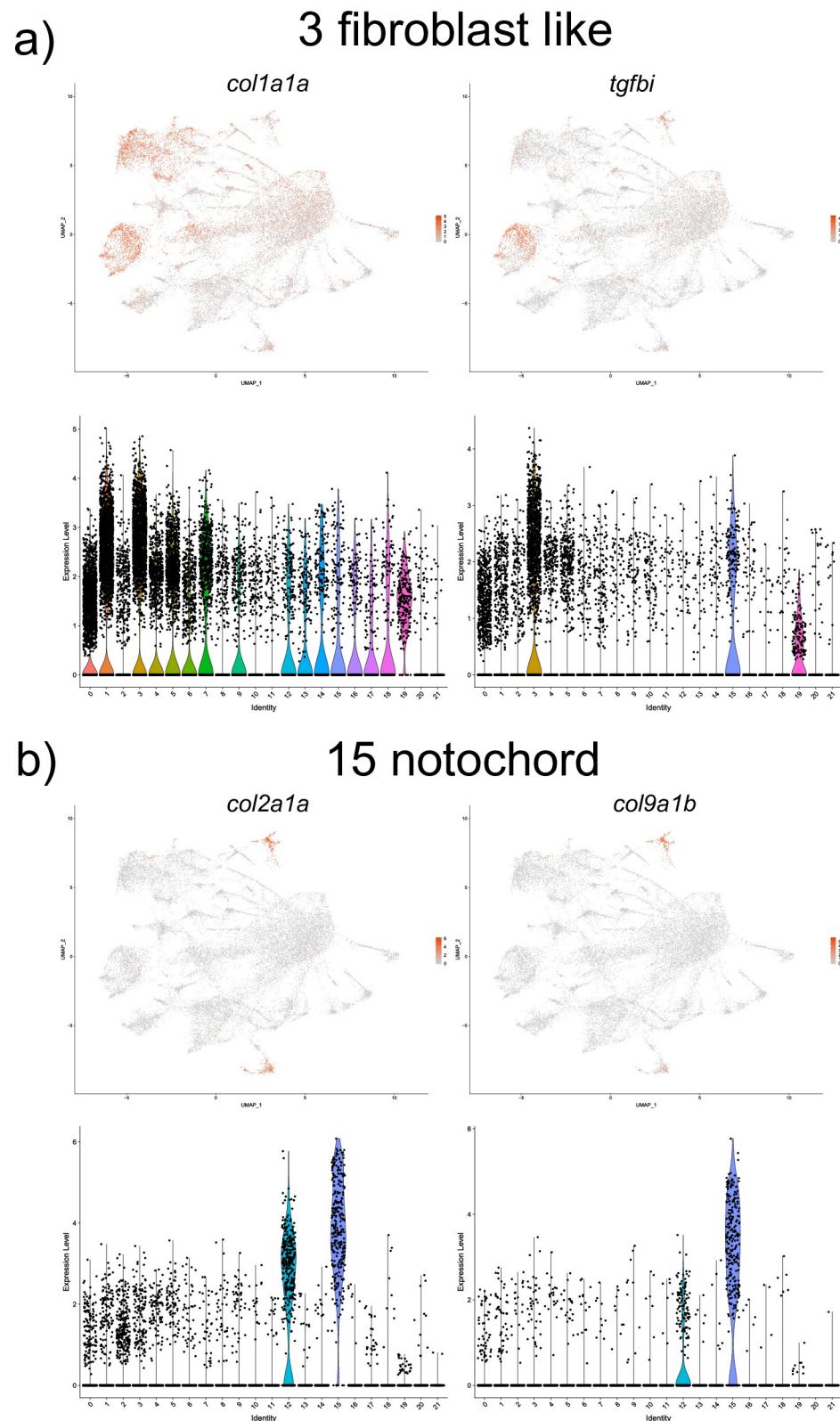


Fig 10. UMAP and violin plots showing the expression of selected marker genes for fibroblast like (cluster #3, a) and notochordal cells (cluster #15, b).

<https://doi.org/10.1371/journal.pone.0254024.g010>

the groups had unique and easily identifiable signatures (endothelial cells, melanocytes, macrophages and others). In some of the groups there was a significant overlap in marker expression. For example, skeletal muscle 1 and 2 groups (clusters 0 and 4) share expression of many top markers, including *pvalb2*, *mylz3*, *tpma* and others. At the same time, many genes also show differential expression between the two groups. Further studies will be needed to validate biological differences between these populations. And lastly, some cell clusters had poorly characterized identities, including lateral mesoderm and blood-unknown (clusters 6 and 21). Many of the top markers of cluster 21 have no or poorly characterized expression in zebrafish. Some of the marker genes, including *lmo2*, *myb* and *fli1a* are known to be expressed in hematopoietic stem cells [49]. *arpc1b* function in mouse has been implicated in T-cell and thrombocyte development [39]. Further in-depth analysis is required to analyze these cell populations.

In recent years, multiple groups have reported different single-cell databases in zebrafish. They include an atlas of neural crest lineages [50], two different studies of using whole embryo single-cell analysis during the first day of zebrafish development [10,11], a single-cell transcriptome atlas for zebrafish at 1, 2 and 5 dpf stages [9] and others. Previous studies have either focused on specific tissues or cell populations such as neural crest or vascular endothelial progenitors [31,50], or have performed analysis at embryonic stages which are different from our study. A recent study has reported a single-cell transcriptome atlas for zebrafish at 1, 2 and 5 dpf stages [9]. In this study 220 different clusters were described based on the analysis of 44,102 cells purified from whole zebrafish embryos. In contrast, our analysis was limited to the trunk region at 30 hpf stage which is a different stage than those previously analyzed. Therefore, it is difficult to make direct comparisons of these datasets. Although Farnsworth *et al* [9] have reported significantly more cell clusters, it is unclear if all of them represent truly distinct cell types. Also, our analysis was limited to a single stage at the trunk region, therefore fewer cell groups are expected. Some of the clusters identified in our study such as arterial and venous specific transcriptomes, or the unknown blood group (cluster 21) have not been reported in the previous study.

In summary, our results provide a unique resource for cell lineages located in the trunk region of a developing zebrafish embryo and will complement transcriptomic datasets generated by other groups. This information will be essential in deciphering the signaling pathways and transcriptional programs that regulate the establishment and differentiation of a variety of cell types during vertebrate development.

Supporting information

S1 Table. Differential expression of marker genes in different cell clusters.

(XLSX)

S2 Table. Average gene expression in different cell clusters.

(XLSX)

S3 Table. Differential expression of marker genes in endoderm + pronephros subcluster.

(XLSX)

S4 Table. Differential expression of marker genes in endothelial cell subcluster.

(XLSX)

Author Contributions

Conceptualization: Saulius Sumanas.

Formal analysis: Sanjeeva Metikala, Satish Casie Chetty.

Funding acquisition: Sanjeeva Metikala, Saulius Sumanas.

Investigation: Sanjeeva Metikala, Satish Casie Chetty.

Methodology: Sanjeeva Metikala, Satish Casie Chetty.

Supervision: Saulius Sumanas.

Writing – original draft: Sanjeeva Metikala, Satish Casie Chetty.

Writing – review & editing: Saulius Sumanas.

References

1. Potter SS. Single-cell RNA sequencing for the study of development, physiology and disease. *Nature Reviews Nephrology*. Nature Publishing Group; 2018. pp. 479–492. <https://doi.org/10.1038/s41581-018-0021-7> PMID: 29789704
2. Karaayvaz M, Cristea S, Gillespie SM, Patel AP, Mylvaganam R, Luo CC, et al. Unravelling subclonal heterogeneity and aggressive disease states in TNBC through single-cell RNA-seq. *Nat Commun*. 2018; 9. <https://doi.org/10.1038/s41467-018-06052-0> PMID: 30181541
3. Stubbington MJT, Rozenblatt-Rosen O, Regev A, Teichmann SA. Single-cell transcriptomics to explore the immune system in health and disease. *Science*. American Association for the Advancement of Science; 2017. pp. 58–63. <https://doi.org/10.1126/science.aan6828> PMID: 28983043
4. van Galen P, Hovestadt V, Wadsworth MH, Hughes TK, Griffin GK, Battaglia S, et al. Single-Cell RNA-Seq Reveals AML Hierarchies Relevant to Disease Progression and Immunity. *Cell*. 2019; 176: 1265–1281.e24. <https://doi.org/10.1016/j.cell.2019.01.031> PMID: 30827681
5. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol*. 2018; 36: 411–420. <https://doi.org/10.1038/nbt.4096> PMID: 29608179
6. Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, et al. Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods*. 2017; 14: 979–982. <https://doi.org/10.1038/nmeth.4402> PMID: 28825705
7. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, et al. Comprehensive Integration of Single-Cell Data. *Cell*. 2019; 177: 1888–1902.e21. <https://doi.org/10.1016/j.cell.2019.05.031> PMID: 31178118
8. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol*. 2014; 32: 381–386. <https://doi.org/10.1038/nbt.2859> PMID: 24658644
9. Farnsworth DR, Saunders LM, Miller AC. A single-cell transcriptome atlas for zebrafish development. *Dev Biol*. 2020; 459: 100–108. <https://doi.org/10.1016/j.ydbio.2019.11.008> PMID: 31782996
10. Farrell JA, Wang Y, Riesenfeld SJ, Shekhar K, Regev A, Schier AF. Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. *Science (80-)*. 2018; 360. <https://doi.org/10.1126/science.aar3131> PMID: 29700225
11. Wagner DE, Weinreb C, Collins ZM, Briggs JA, Megason SG, Klein AM. Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science (80-)*. 2018; 360: 981–987. <https://doi.org/10.1126/science.aar4362> PMID: 29700229
12. Potter AS, Steven Potter S. Dissociation of tissues for single-cell analysis. *Methods in Molecular Biology*. Humana Press Inc.; 2019. pp. 55–62. https://doi.org/10.1007/978-1-4939-9021-4_5.
13. Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al. Massively parallel digital transcriptional profiling of single cells. *Nat Commun*. 2017; 8: 1–12. <https://doi.org/10.1038/s41467-016-0009-6> PMID: 28232747
14. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell*. 2015; 161: 1202–1214. <https://doi.org/10.1016/j.cell.2015.05.002> PMID: 26000488
15. Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG, et al. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol*. 2019; 37: 38–47. <https://doi.org/10.1038/nbt.4314> PMID: 30531897

16. Thisse B., Pflumio S., Fürthauer M., Loppin B., Heyer V., Degraeve A., et al. Expression of the zebrafish genome during embryogenesis. (NIH R01 RR15402) ZFIN Direct Data Submission (<http://zfin.org>). 2001. Available: <https://zfin.org/ZDB-PUB-010810-1>.
17. Thisse B., Thisse C. Fast Release Clones: A High Throughput Expression Analysis. ZFIN Direct Data Submission (<http://zfin.org>) Close. 2004. Available: <https://zfin.org/ZDB-PUB-040907-1>.
18. Rauch G.J., Lyons D.A., Middendorf I., Friedlander B., Arana N., Reyes T., et al. Submission and Curation of Gene Expression Data. ZFIN Direct Data Submission (<http://zfin.org>). 2003. Available: <https://zfin.org/ZDB-PUB-031103-24>.
19. Kim CH, Ueshima E, Muraoka O, Tanaka H, Yeo SY, Huh TL, et al. Zebrafish elav/HuC homologue as a very early neuronal marker. *Neurosci Lett*. 1996; 216: 109–112. [https://doi.org/10.1016/0304-3940\(96\)13021-4](https://doi.org/10.1016/0304-3940(96)13021-4) PMID: 8904795
20. Appelbaum L, Skariah G, Mourrain P, Mignot E. Comparative expression of p2x receptors and ectonucleoside triphosphate diphosphohydrolase 3 in hypocretin and sensory neurons in zebrafish. *Brain Res*. 2007; 1174: 66–75. <https://doi.org/10.1016/j.brainres.2007.06.103> PMID: 17868657
21. Kelsh RN, Schmid B, Eisen JS. Genetic analysis melanophore development in zebrafish embryos. *Dev Biol*. 2000; 225: 277–293. <https://doi.org/10.1006/dbio.2000.9840> PMID: 10985850
22. Lin LY, Horng JL, Kunkel JG, Hwang PP. Proton pump-rich cell secretes acid in skin of zebrafish larvae. *Am J Physiol—Cell Physiol*. 2006; 290. <https://doi.org/10.1152/ajpcell.00281.2005> PMID: 16148031
23. Hsiao C Der, You MS, Guh YJ, Ma M, Jiang YJ, Hwang PP. A positive regulatory loop between foxi3a and foxi3b is essential for specification and differentiation of zebrafish epidermal ionocytes. *PLoS One*. 2007; 2. <https://doi.org/10.1371/journal.pone.0000302> PMID: 17375188
24. Hsiao C Der, Tsai WY, Tsai HJ. Isolation and expression of two zebrafish homologues of parvalbumin genes related to chicken CPV3 and mammalian oncomodulin. *Gene Expr Patterns*. 2002; 2: 163–168. [https://doi.org/10.1016/s0925-4773\(02\)00340-4](https://doi.org/10.1016/s0925-4773(02)00340-4) PMID: 12617856
25. Dickmeis T, Plessy C, Rastegar S, Aanstad P, Herwig R, Chalmei F, et al. Expression profiling and comparative genomics identify a conserved regulatory region controlling midline expression in the zebrafish embryo. *Genome Res*. 2004; 14: 228–238. <https://doi.org/10.1101/gr.1819204> PMID: 14718378
26. Duncan RN, Panahi S, Piotrowski T, Dorsky RI. Identification of Wnt Genes Expressed in Neural Progenitor Zones during Zebrafish Brain Development. Alsina B, editor. *PLoS One*. 2015; 10: e0145810. <https://doi.org/10.1371/journal.pone.0145810> PMID: 26713625
27. Thisse C., and Thisse B. High Throughput Expression Analysis of ZF-Models Consortium Clones. ZFIN Direct Data Submission (<http://zfin.org>). 2005. Available: <https://zfin.org/ZDB-PUB-051025-1>.
28. Parichy DM, Ransom DG, Paw B, Zon LI, Johnson SL. An orthologue of the kit-related gene *fms* is required for development of neural crest-derived xanthophores and a subpopulation of adult melanocytes in the zebrafish, *Danio rerio*. *Development*. 2000; 127: 3031–3044. <https://doi.org/10.1242/dev.127.14.3031> PMID: 10862741
29. Xie J, Farage E, Sugimoto M, Anand-Apte B. A novel transgenic zebrafish model for blood-brain and blood-retinal barrier development. *BMC Dev Biol*. 2010; 10: 76. <https://doi.org/10.1186/1471-213X-10-76> PMID: 20653957
30. Hogan BM, Schulte-Merker S. How to Plumb a Pisces: Understanding Vascular Development and Disease Using Zebrafish Embryos. *Dev Cell*. 2017; 42: 567–583. <https://doi.org/10.1016/j.devcel.2017.08.015> PMID: 28950100
31. Chestnut B, Casie Chetty S, Koenig AL, Sumanas S. Single-cell transcriptomic analysis identifies the conversion of zebrafish Etv2-deficient vascular progenitors into skeletal muscle. *Nat Commun*. 2020; 11. <https://doi.org/10.1038/s41467-020-16515-y> PMID: 32493965
32. Row RH, Pegg A, Kinney BA, Farr GH, Maves L, Lowell S, et al. BMP and FGF signaling interact to pattern mesoderm by controlling basic helix-loop-helix transcription factor activity. *Elife*. 2018; 7. <https://doi.org/10.7554/eLife.31018> PMID: 29877796
33. Wilting J, Brand-Saberi B, Huang R, Zhi Q, Köntges G, Ordahl CP, et al. Angiogenic potential of the avian somite. *Dev Dyn*. 1995; 202: 165–171. <https://doi.org/10.1002/aja.1002020208> PMID: 7537553
34. Wilting J, Becker J. Two endothelial cell lines derived from the somite. *Anatomy and Embryology*. 2006. <https://doi.org/10.1007/s00429-006-0120-2> PMID: 17047989
35. Brownlie A, Hersey C, Oates AC, Paw BH, Falick AM, Witkowska HE, et al. Characterization of embryonic globin genes of the zebrafish. *Dev Biol*. 2003; 255: 48–61. [https://doi.org/10.1016/s0012-1606\(02\)00041-6](https://doi.org/10.1016/s0012-1606(02)00041-6) PMID: 12618133
36. Kovina AP, Petrova N V, Gushchanskaya ES, Dolgushin K V, Gerasimov ES, Galitsyna AA, et al. Evolution of the Genome 3D Organization: Comparison of Fused and Segregated Globin Gene Clusters. [cited 21 May 2020]. <https://doi.org/10.1093/molbev/msx100> PMID: 28333290

37. Jia S, Jia W, Yu S, Hu Y, He Y. Using microarray analysis to identify genes and pathways that regulate fetal hemoglobin levels. *Ann Hum Genet.* 2020; 84: 29–36. <https://doi.org/10.1111/ahg.12346> PMID: 31396950
38. Zakrzewska A, Cui C, Stockhammer OW, Benard EL, Spaink HP, Meijer AH. Macrophage-specific gene functions in Spi1-directed innate immunity. *Blood.* 2010; 116: e1–e11. <https://doi.org/10.1182/blood-2010-01-262873> PMID: 20424185
39. Somech R, Lev A, Lee YN, Simon AJ, Barel O, Schiby G, et al. Disruption of Thrombocyte and T Lymphocyte Development by a Mutation in ARPC1B. *J Immunol.* 2017; 199: 4036–4045. <https://doi.org/10.4049/jimmunol.1700460> PMID: 29127144
40. Sun C, Wu J, Liu S, Li H, Zhang S. Zebrafish CD59 has both bacterial-binding and inhibiting activities. *Dev Comp Immunol.* 2013; 41: 178–188. <https://doi.org/10.1016/j.dci.2013.05.008> PMID: 23707788
41. Yelon D, Ticho B, Halpern ME, Ruvinsky I, Ho RK, Silver LM, et al. The bHLH transcription factor hand2 plays parallel roles in zebrafish heart and pectoral fin development. *Development.* 2000; 127: 10821756
42. Germanguz I, Lev D, Waisman T, Kim C-H, Gitelman I. Four *twist* genes in zebrafish, four expression patterns. *Dev Dyn.* 2007; 236: 2615–2626. <https://doi.org/10.1002/dvdy.21267> PMID: 17685477
43. Ravenscroft G, Zaharieva IT, Bortolotti CA, Lambroughi M, Pignataro M, Borsari M, et al. Bi-allelic mutations in MYL1 cause a severe congenital myopathy. *Hum Mol Genet.* 2018; 27: 4263–4272. <https://doi.org/10.1093/hmg/ddy320> PMID: 30215711
44. Middleton RC, Shelden EA. Small heat shock protein HSPB1 regulates growth of embryonic zebrafish craniofacial muscles. *Exp Cell Res.* 2013; 319: 860–874. <https://doi.org/10.1016/j.yexcr.2013.01.002> PMID: 23313812
45. Chen Z, Huang W, Dahme T, Rottbauer W, Ackerman MJ, Xu X. Depletion of zebrafish essential and regulatory myosin light chains reduces cardiac function through distinct mechanisms. *Cardiovasc Res.* 2008; 79: 97–108. <https://doi.org/10.1093/cvr/cvn073> PMID: 18343897
46. Holterhoff CK, Saunders RH, Brito EE, Wagner DS. Sequence and expression of the zebrafish alpha-actinin gene family reveals conservation and diversification among vertebrates. *Dev Dyn.* 2009; 238: 2936–2947. <https://doi.org/10.1002/dvdy.22123> PMID: 19842183
47. Gupta V, Discenza M, Guyon JR, Kunkel LM, Beggs AH. α -Actinin-2 deficiency results in sarcomeric defects in zebrafish that cannot be rescued by α -actinin-3 revealing functional differences between sarcomeric isoforms. *FASEB J.* 2012; 26: 1892–1908. <https://doi.org/10.1096/fj.11-194548> PMID: 22253474
48. Yan Y-L, Hatta K, Riggleman B, Postlethwait JH. Expression of a type II collagen gene in the zebrafish embryonic axis. *Dev Dyn.* 1995; 203: 363–376. <https://doi.org/10.1002/aja.1002030308> PMID: 8589433
49. Thompson MA, Ransom DG, Pratt SJ, MacLennan H, Kieran MW, Detrich HW, et al. The cloche and spadetail genes differentially affect hematopoiesis and vasculogenesis. *Dev Biol.* 1998; 197: 248–269. <https://doi.org/10.1006/dbio.1998.8887> PMID: 9630750
50. Howard AGA, Baker PA, Ibarra-García-padilla R, Moore JA, Rivas LJ, Tallman JJ, et al. An atlas of neural crest lineages along the posterior developing zebrafish at single-cell resolution. *Elife.* 2021; 10: 1–31. <https://doi.org/10.7554/eLife.60005> PMID: 33591267