Lymphocyte subsets in Atlantic cod (Gadus morhua) interrogated by single-cell sequencing

Naomi Guslund (naomi.croft@gmail.com)
University of Oslo https://orcid.org/0000-0002-8010-0065
Anders Krabberød
University of Oslo https://orcid.org/0000-0001-9481-8396
Simen Nørstebø
Norwegian University of Life Sciences
Monica Solbakken
Norwegian University of Life Sciences,
Kjetill Jakobsen
University of Oslo https://orcid.org/0000-0002-8861-5397
Finn-Eirik Johansen
University of Oslo
Shuo-Wang Qiao
University of Oslo

Article

Keywords: Atlantic Cod1, single-cell sequencing2, CD8-negative T lymphocytes3, monocyte chemotactic protein 14, scavenger receptor m1305

Posted Date: February 18th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1342445/v1

Version of Record: A version of this preprint was published at Communications Biology on July 11th, 2022. See the published version at https://doi.org/10.1038/s42003-022-03645-w.

Lymphocyte subsets in Atlantic cod (*Gadus morhua*) interrogated by single-cell sequencing

Authors and affiliations:
Naomi Croft Guslund[1*], Anders K. Krabberød[1,2], Simen F. Nørstebø[3], Monica Hongrø Solbakken[4], Kjetill S. Jakobsen[4], Finn-Eirik Johansen[5], Shuo-Wang Qiao[6*]

[1]Centre for Ecological and Evolutionary Synthesis, Department of Biosciences and the Department of Immunology, University of Oslo, Oslo, Norway
[2]Section for Genetics and Evolutionary Biology, Department of Biosciences and the Department of Immunology, University of Oslo, Oslo, Norway
[3]Department of Paraclinical Sciences, Faculty of Veterinary Medicine, Norwegian University of Life Sciences, Ås, Norway
[5]Section for Physiology and Cell Biology, Department of Biosciences, University of Oslo, Oslo, Norway
[6]Department of Immunology, Institute of Clinical Medicine, University of Oslo, Oslo, Norway

Correspondence

[*]Naomi Croft Guslund – n.c.guslund@ibv.uio.no
[*]Shuo-Wang Qiao - s.w.qiao@medisin.uio.no

Keywords: Atlantic Cod$_1$, single-cell sequencing$_2$, CD8-negative T lymphocytes$_3$, monocyte chemotactic protein 1$_4$, scavenger receptor *m130*$_5$

Abstract

Atlantic Cod (*Gadus morhua*) has lost the *major histocompatibility complex class II* presentation pathway. We recently identified CD8-positive T cells, B cells, and plasma cells in cod, but further characterization of lymphocyte subsets are needed to elucidate immune adaptations triggered by the absence of CD4-positive T lymphocytes. Here we use single-cell RNA sequencing to examine the lymphocyte heterogeneity in Atlantic cod spleen. We describe five TCR-positive T cell subsets and eight BCR-positive B cell subsets and propose a B cell trajectory of differentiation. Notably, we identify a significant subpopulation of T cells that are CD8-negative. Most of the CD8-negative T lymphocytes highly express the homolog of *monocyte chemotactic protein 1b*, and another subset of CD8-negative T lymphocytes express the homolog of the scavenger receptor *m130*. Uncovering the multiple lymphocyte cell sub-clusters reveals the different immune states present within the B and T cell populations, building a foundation for further work.

1 Introduction

Comparative studies of a limited number of model organisms have led to a commonly held view that the adaptive immune system common to jawed vertebrates has remained largely unchanged since its origin some 500 million years ago [1, 2]. More recent studies of non-model organisms have begun to challenge this view, as exemplified by the Atlantic cod (*Gadus morhua*). The Atlantic cod belongs to a growing number of identified teleost species that have lost substantial parts of the adaptive immune system. The Gadiformes order, of which the Atlantic cod is one of the most prevalent species, lost the entire set of *major histocompatibility complex (MHC) class II* genes around 80-100 million years ago [3, 4]. The *cd4* gene is also non-functional, and thus cod does not have CD4-positive T cells, which are known to play a pivotal role in raising a humoral immune response in mammalian immune systems. Despite its 'unusual' immune system, cod is a prolific species and in its natural environment is not more prone to diseases than other teleosts with more 'conventional' immune systems [5]. Although it has proven difficult to elicit specific antibody responses by immunisation,

Atlantic cod does display specific and long-term protective memory in vaccination studies [6-10]. Thus, Atlantic cod has a well-functioning immune system in the absence of CD4-positive T cells, but the mechanisms for generating specific immune memory, if it exists in cod, must differ from the well-known model based on collaboration between B cells and CD4-positive T cells, although this model is in general not validated in teleosts.

As a first step towards understanding the workings of the immune system of Atlantic cod, we have used single-cell transcriptomics to characterise the immune cell composition. In a previously published study, we analysed the transcriptomic profile of some seven thousand immune cells sampled from the spleen and blood of two Atlantic cod [11]. We identified major immune-cell populations (T cells, B cells, plasma cells, macrophages, and granulocytes) and non-immune cells (erythrocytes, thrombocytes, and spleen stromal cells), as well as minor populations of cells such as dendritic cells and $GATA3^{+}$ cytotoxic cells, a name coined by us as there was no clear classification. We hypothesized that a subset of T cells might functionally resemble CD4-positive T cells, expressing markers that may indicate their interaction with B cells. However, we did not have sufficient cell numbers for further sub-clustering of the 1,300 T cells.

In the current work, we have assembled a much larger dataset of single-cell RNA sequencing (scRNAseq) data of 57 thousand spleen cells from 34 Atlantic cod specimens in a vaccination study. The teleost spleen functions as the primordial secondary lymphoid organ where adaptive immune responses are generated [2]. By sampling from both naive and vaccinated fish before, during, and after an immune challenge, we capture cells in a broad spectrum of immune states. Sub-clustering and pseudotime trajectory analysis of the B cells revealed expression of transcription factors that may be involved in B cell differentiation. Closer inspection of the almost 14 thousand T lymphocytes revealed a large subset that does not express CD8. The majority of CD8-negative T cells express high to extremely high levels of a homolog of *monocyte chemotactic protein 1b* (*mcp1b*), whereas a smaller subset of CD8-negative T cells express a homolog of the scavenger receptor *m130*. The presence of T cell subsets in the Atlantic cod secondary lymphoid organ that express neither *cd4* nor *cd8* co-receptors poses important questions about the functional organisation of the immune system in this species.

2 Methods and materials

2.1 Atlantic cod sampling

Atlantic cod sampled for this study originate from a single breeding family (bred from 1 female and 2 males) from the NOFIMA breeding program in Tromsø, Norway. They were reared, vaccinated, and challenged at the NIVA Research Facility at Solbergstrand, outside Oslo, Norway. The rearing and sampling were performed according to the European animal welfare regulations and approved by the Norwegian authorities (FOTS ID 21758).

Sampling took place between 9 and 11 months of age (average length = 21cm, range 13-27cm). The water temperature was maintained at ~8^{o}C in keeping with the seasonal water temperature, with a water salinity of 34 PSU, light conditions were 12:12 hour light:dark, and the cod were fed with Skretting cod pellets. Spleen samples were taken from 34 Atlantic cod that were used in a vaccination study for *Vibrio anguillarum* (strain O2a ATCC 19264). Samples were taken before and after vaccination priming and boosting, before immune challenge, then day one, seven and 15 after immune challenge for naive and vaccinated fish. Vaccination priming and boost, and immune challenge were carried out by bath immersion. A schematic demonstrating sampling is shown in *Supplementary Figure 1*. Fish were killed by cranial concussion and tissue sampling was conducted within minutes of fish death. As blood was also taken for additional investigations not mentioned in this paper, gill slit could not be carried out. The spleens were removed and placed in PBS-BSA 0.01% solution and kept on ice. Spleen cell suspensions were obtained by gently forcing the tissue through a cell strainer (Falcon, 100 µm) and diluted to 200 cells/ul in fresh PBS-BSA 0.01% solution. All cells were

kept in regular microcentrifuge tubes to minimise cell loss and kept on ice for transport and laboratory work.

2.2 Single-cell cDNA library preparation and sequencing.

Single-cell RNA sequencing (scRNAseq) and the preparation of the libraries were performed according to published protocols [11-13]. A droplet sequencing generator (Dolomite) individually encapsulated cells with barcoded beads. Libraries were sequenced at the Norwegian Sequencing Centre (Oslo University Hospital), on the NextSeq500 platform with a 75 bp kit, high output mode, with paired end reads. 20 bp were sequenced in Read 1 using a custom sequencing primer (GCCTGTCCGCGGAAGCAGTGGTATCAACGCAGAGTAC) and 60 bp in Read 2 with the regular Illumina sequencing primer. The sequencing data is available at the ENA repository with Accession number PRJEB47815.

We used the Drop-seq Core Computational Protocol [14] to demultiplex the raw sequencing reads and then map them to the most recent version of the Atlantic cod genome, gadMor3 (RefSeq accession GCF_902167405.1) using STAR alignment. A gene of interest, *gata3*, was present in gadMor2 [15] but missing in gadMor3, so the *gata3* gene sequence was manually added to the gadMor3 assembly fasta file. The reads were summarised into gene expression matrices using the Drop-seq program 'DigitalExpression'.

2.3 Pre-processing workflow

The resulting count matrices for each sample were loaded into Seurat (version 4.0.2)[16-19]. The count matrices from each of the 34 cod were integrated into one Seurat object through anchor identification using the 'FindIntergationAnchors' and 'IntergrateData' Seurat functions. The percentage of mitochondrial gene expression was added by 'PercentageFeatureSet' and cell barcodes with more than 5% of mitochondrial genes, 'percent.mt' were removed. Cell barcodes with fewer than 150 genes and greater than 1500 genes, 'nFeature_RNA', and those that contained genes expressed in less than three cells, 'min.cells', were filtered out. Additionally, cells with a total number of biological molecules greater than 4000, 'nCount_RNA', were removed. These steps reduce the number of low-quality cells and reduce cell multiplets or cell barcodes which do not contain a true cell transcriptome but rather ambient RNA. Following quality control filtering, we derived a gene expression matrix of 19,279 genes across 56,994 splenic cells. An overview of the samples in each treatment group in the immune challenge protocol, sequencing library, average mapping percentages, number of cells, mapped transcripts, and genes are shown in Supplementary excel sheet 1.

After filtering, data were normalized using the 'NormalizeData' function, variable features were identified using the 'FindVariableFeatures' function, and features were scaled using the 'ScaleData' function. PCA analysis was performed, and the most significant principal components were identified by plotting the first 50 dimensions on an Elbow plot (*Supplementary Figure 2*).

2.4 Dimension reduction and assigning cluster identity

Unsupervised dimension reduction was performed using Uniform Manifold Approximation and Projection (UMAP) [20] on the first 40 principal components and the clusters were identified using the 'FindNeighbors' (dims = 1:40) and 'FindClusters' functions (res = 0.2). Cluster identities were assigned by assessing differential expression of marker genes between different clusters using the biomarkers detected from the 'FindAllMarkers' function (only.pos = TRUE, min.pct = 0.25, logfc.threshold = 0.25). The top 20 differentially expressed genes for all clusters found in the splenic samples can be seen in *Supplementary excel sheet 2*.

To examine the lymphocyte populations more closely, we selected the B cells, CD8-positive T cells, MCP1b-positive T cells, proliferating lymphocytes, and plasma cells from the global splenic UMAP, thus selecting 23,227 lymphocytic cells. These lymphocytes were then clustered on the first 40 dimensions at 0.2 resolution. As a result of this unsupervised clustering step the proliferating

lymphocytes cluster was divided into B lymphocytes or T lymphocytes. The 9,474 cells in total identified as B lymphocytes were clustered using UMAP on the first 40 dimensions with a 0.3 resolution. The 13,753 cells identified as T lymphocytes were clustered using UMAP on the first 40 dimensions with a 0.09 resolution.

2.5 Trajectory analysis

We used Slingshot (v. 1.8.0) analysis [21] to infer possible differentiating trajectories within the B lymphocyte sub-clusters. UMAP was used to determine the dimensionality and the lineages were constructed in an unsupervised way. To infer differential gene expression in the lineages predicted by the Slingshot analysis, we used the GAM package (v. 1.20) to run a generalized additive model (GAM) which uses a LOESS term for pseudotime. This identified the top 100 most variable genes across pseudotime (*Supplementary excel sheets 3*). The lineages and expression of key genes were visualised using the Slingshot tools, the ggplot2 (v. 3.3.3) [22] package, and the scater package (v. 1.18.6)[23].

2.6 Phylogeny analysis of *m130, mcp1b, and tcr* genes

All gene names reported in the scRNA analyses are derived from the automated annotation of the gadMor3 assembly. To further explore the identity of particular genes of interest, *mcp1b* (LOC115529242), *m130* (LOC115541469) and the *tcr* genes, in more detail, maximum likelihood phylogenetic trees were constructed.

Scavenger Receptor Cysteine-Rich Type 1 Protein M130 (CD163) references were collected from the groups Mammalia, Aves, Reptilia, Amphibia and Teleostei at Uniprot.org where available. The predicted gene models of the scRNA-reported annotations were then subjected to a blastp at NCBI towards mammalian and teleostei databases using default parameters. Here, 3-5 hits annotated as M130/M160/CD163 were selected to be included in the multiple sequence alignment. The scavenger receptor cysteine rich (SRCR) domain present in CD163 is an ancient and conserved domain shared by members of the group B SRCR superfamily (Herzig et al., 2010). To provide more framework to the alignment of *m130*, additional members of the group B SRCR family were included (CD5, CD5L and CD6) together with two other scavenger receptors: scavenger receptor cysteine rich family member with 4 domains (SSC4D) and Scavenger Receptor Class B Member 1 (SCARB1), all obtained in the same manner as described above.

Monocyte chemotactic protein 1 (CCL2) references were obtained as described above for CD163. Furthermore, as CCL2 is a chemokine belonging to a large gene family, a set of references for the most common CC chemokines in humans and fish were downloaded from genbank and added to the alignment. In addition, some CXC chemokines were added to function as a collective outgroup. Finally, complete CC chemokine characterizations from Channel catfish and Atlantic cod were added, together with all predicted gene models from gadMor3 reported from a tblastn search using the CCL2 references, default parameters and e-value cutoff at 1e-1 [24].

The final multiple sequence alignments (protein) were aligned using MUSCLE in MEGA7 and default parameters [25]. The resulting protein alignment was subjected to a maximum likelihood run using raxml-ng [26] after model testing using modeltest-ng [27]. For CD163, the best scoring model was WAG+G4 and for CCL2 VT+I+G4. Both maximum likelihood runs were run until convergence and followed with 500 bootstrap replicates.

We also constructed a phylogenetic tree of *tcr* genes in Atlantic cod and in other selected species. The annotated *tcrb* (LOC115544273) and *tcra* (LOC115548821) genes in gadMor3, as well as putative *tcr* genes (LOC115548656 and LOC115548822) identified by blasting gadMor3 with *tcr* genes from different teleosts, reptiles, and birds, were compared. All genes with hits above 1e-100 were collected and added to the overall sequence alignment. A multiple sequence alignment was generated using MEGA7 and its MUSCLE alignment program with default setting. From here a neighbour joining tree using Poisson distribution, pairwise deletion and 200 bootstrap replicates was made again using MEGA7.

The trees are visualized and edited for presentation purposes using FigTree v 1.4.4 [28]

3 Results

3.1 Overview of cells found in Atlantic cod spleen

We analysed spleens from 34 Atlantic cod at 12 timepoints during a vaccination and immune challenge study of *V. anguillarum*, thus capturing a broad immune status. We examined these spleen samples using scRNAseq. In this study we focus on a detailed exploration of cell sub-clusters within the B and T cell lymphocytes, and thus we do not delve into the differing immune responses following vaccination and immune challenge. All the named clusters in the global UMAP in *Figure 1* as well as the larger lymphocyte sub-clusters have cells from all the sampled fish. While some of the smallest lymphocyte sub-clusters are not represented by all fish, there was no pattern of changing sub-cluster population size with respect to naive versus vaccinated fish or along the immune challenge timeline. This demonstrates that all the splenic cell populations and lymphocyte sub-cluster populations are found in both steady state and immune perturbed Atlantic cod. The numbers of cells found in each cluster or sub-cluster is shown in *Supplementary excel sheet 4.*

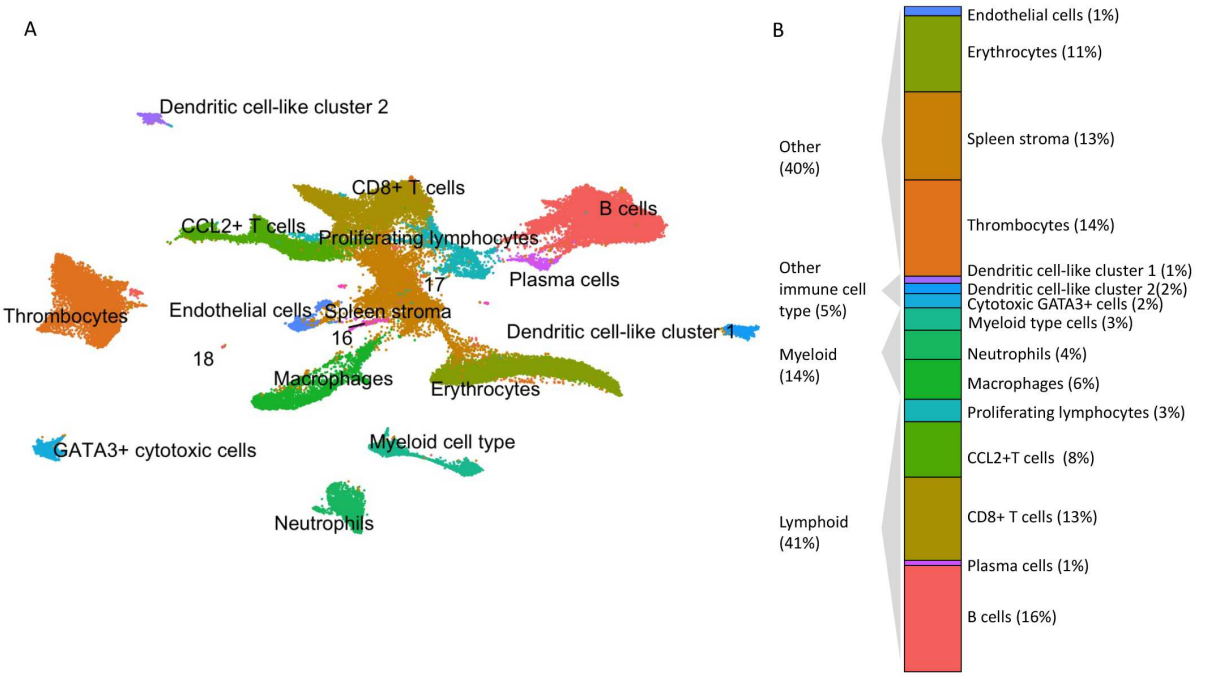

Figure 1. Unsupervised clustering of 56,994 Atlantic cod splenic cells reveals a population of CD8-negative T cells.
(A) Single-cell RNA sequencing of splenic cells is visualised in two dimensions using UMAP. Cluster identity was assigned based upon differentially expressed genes and naming from our last publication [11]. Clusters 16-18 each represent less than 0.2% of the total cells and so have not been named.
(B) Bar chart showing percentage of splenic cells as lymphoid cells, myeloid cells, other immune cells, and other splenic cells. Cells from clusters 16-18 have not been included in the percentage calculations.

Following quality control filtering, we derived a gene expression matrix across 56,994 Atlantic cod splenic cells. Visualization of cell types in two dimensions using UMAP revealed 15 cell clusters with distinct gene expression signatures (*Figure 1A,* a list of differentially expressed genes can be viewed in *Supplementary excel sheet 2)*. Three clusters, clusters 16, 17, and 18, each represented less than 0.2% of the total cells and so remain unnamed. The splenic cells contained the major types of lymphoid and myeloid cells, minor immune cell types (myeloid cell type, dendritic cell cluster 1 and 2, and GATA3^{+} cytotoxic cells), as well as other cell types (the thrombocytes, erythrocytes, spleen stromal cells, and endothelial cells). The cell cluster 'myeloid cell type' was previously named 'natural killer

cells' by us [11]. Despite a larger sample size in the present study, a lack of clear marker genes has resulted in a less specific name choice. Two clusters demonstrated gene expression profiles matching dendritic cells, both expressing the receptor tyrosine kinase *flt3*, allograft inflammatory factor 1 (*aif1*), and various cathepsin genes, and have thus been named dendritic cell cluster 1 and 2. The naming of the cell types is otherwise consistent with our first classification of cod immune cells [11].

The lymphocytes represented 41% of the splenic cells present, and were made up of B cells, plasma cells, proliferating lymphocytes, and T cells (*Figure 1B*). The T cells were clustered into two distinct groups; one group contains CD8-positive T cells, while the other did not show CD8 expression but highly expressed a homolog of *mcp1b*, thus given the name MCP1b-positive T cells (*Figure 1A*). The myeloid cells (containing the macrophages, neutrophils, and myeloid type cells) represented 14% of splenic cells, while the other immune cells (Dendritic cell cluster 1, Dendritic cell cluster 2, and GATA3^{+} cytotoxic cells) represented 5% of splenic cells (*Figure 1B*).

3.2 B lymphocyte sub-clusters

By selecting 9,474 B lymphocytic cells (the B cells, B cells within the proliferating lymphocytes, and plasma cells) from the global UMAP containing all splenic cells and re-clustering these cells at a higher resolution, eight sub-clusters were revealed (*Figure 2A*). Shared expression of the marker genes *cd22*, *swap70*, *cd79a* (*Figure 2B*) and immunoglobulin genes (*Supplementary Figure 3*) confirmed these were all B lymphocytes. B cell sub-clusters 0, 1, 2, 3, 4, and 6 were not conspicuously different from one another, revealing a mixed gradient pattern of gene expression, such as the transcription factors ~~-~~ *spi1* and the adapter protein *sh2d1a* (*Figure 2B*). Sub-clusters 2 and 5 expressed the TNF receptor super-family gene *tnfrsf13b*. The expression of shared immunoglobulin genes dominated the most significantly expressed genes for these clusters, with few unique marker genes (*Supplementary excel sheet 5*). B cell sub-cluster 4 differentially expressed *jun-D,* a sub-unit of the AP-1 transcription factor, which has been linked to different roles in B cell activation and differentiation [29].

B cell sub-clusters 5 and 7 had very distinct gene expression profiles. Sub-cluster 5, making up 5% of the B cells, are the previously identified plasma cells and expressed the marker genes *interferon regulatory factor 4* (*irf4*) and *peptidylprolyl isomerase b* (*ppib*) (*Figure 2B*) and demonstrated the highest average expression of immunoglobulin genes (*Supplementary Figure 3*), as expected. B cell sub-cluster 7, making up only 1% of the B cells, highly differentially expressed several genes, including *thrombospondin 1* (*thbs1*), *thymosin beta 11* (*tmsb11*), and *cellular communication network factor 2* (*ccn2*) (other genes can be seen in *Supplementary excel sheet 5*).

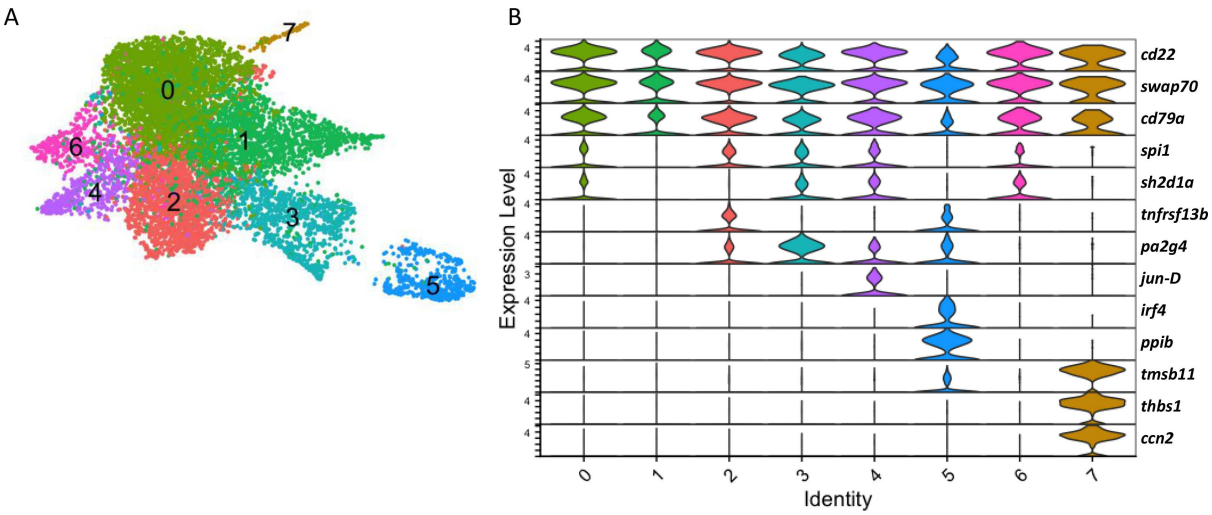

Figure 2. Closer examination of 9,474 Atlantic cod B lymphocytes reveals 8 sub-clusters with distinct transcriptome profiles.

(A) Visualization of B lymphocyte cells using UMAP.

(B) Violin plots showing expression of selected genes identified by differentially expressed gene analysis. The y-axis indicates normalized and log-transformed average expression of the selected genes and the x-axis indicates the B cell sub-cluster.

The trajectory analysis Slingshot is an inference tool which computationally orders cell clusters along a predicted developmental trajectory based on gene expression, an ordering known as pseudotime. Performing a Slingshot analysis on the B lymphocyte clusters suggested two lineages; both lineages started in B cell sub-cluster 6, with lineage 1 ending in sub-cluster 5 (the plasma cells) and lineage 2 ending in sub-cluster 7 (*Figure 3A-D*). We did not manually select any start or end points for the lineages, allowing Slingshot to compute pseudotime in an unbiased fashion. The terminal clusters of the two lineages are consistent with the discrete populations demonstrated by the differential gene analysis. B cell sub-cluster 3, which is the cluster of B cells ordered in pseudotime right before the terminally differentiated plasma cells, highly expressed *proliferation associated 2g4* (*pa2g4*), a gene involved in cell proliferation, differentiation, and survival [30].

We fit a generalised additive model (GAM) to the two predicted trajectories to explore potential differences in the progression of the B cell lineages and identified the top 100 temporally expressed genes (*Supplementary excel sheet 3*). The expression of these genes changed in a continuous manner over pseudotime. The transcription factors *pax5* and *e2a* were present among the top differentially expressed genes identified by GAM and visualised as smooth expression plots (*Figure 3E*). The gene *tnfrsf13b* is a gene of biological interest and has also been shown. All three genes are linked to B cell development: *pax5* encodes the B cell lineage-specific activator protein expressed at early stages of B cell differentiation [31], *e2a* plays a role in B and T lymphocyte development [32], and *tnfrsf13b* encodes the transmembrane activator and CAML interactor (TACI) suggested to control B cell differentiation and memory [33, 34]. At the end of lineage 1 the expression levels of *pax5* dropped in the plasma cells, while expression levels of *e2a* and *tnfrsf13b* both showed an increase. In lineage 2 the expression levels of the *pax5* and *e2a* remained steady, while expression of *tnfrsf13b* peaked in sub-cluster 2 and then had low expression through the rest of the lineage. It is unclear why there was a noticeable dip in expression level of all pax5 and e2a genes in sub-cluster 1, especially in lineage 1.

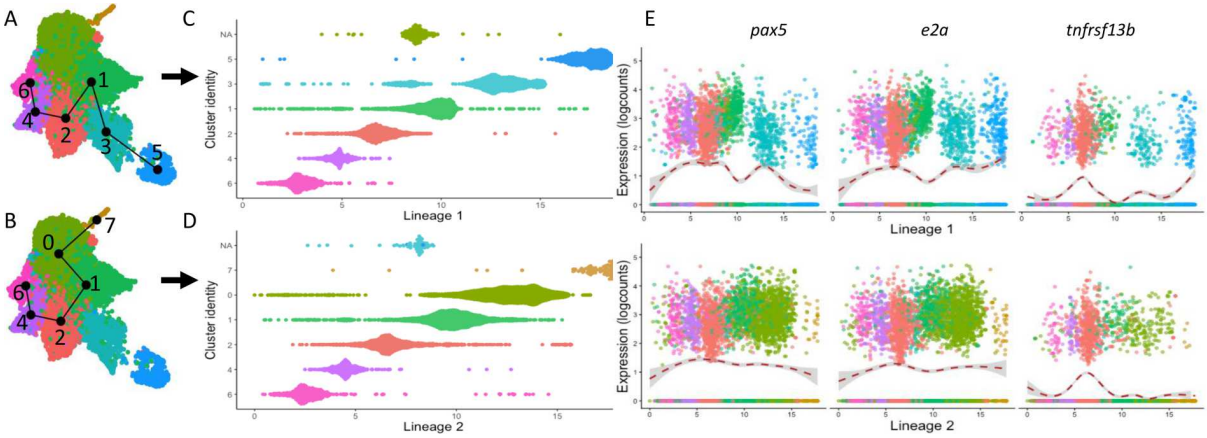

Figure 3. Pseudotime analysis of B lymphocyte cells reveal two lineages with different expression patterns of transcription factors.

(A + B) B cell sub-clusters depicting the Slingshot-predicted lineage trajectories, lineage 1 **(A)** terminating in the B cell sub-cluster 5 (the plasma cells), and lineage 2 **(B)** terminating in B cell sub-cluster 7.

(C + D) Depiction of cells within each B cell sub-cluster from the Slingshot-predicted lineage on the y-axis along pseudo-temporal ordering on the x-axis. Each dot represents a cell and its predicted temporal position in development in lineage 1 **(C)** and lineage 2 **(D)**.

(E) Smoothed expression plots along B lymphocyte lineages for the transcription factors *pax5*, *e2a*, and tnfrsf13b. Each point represents the expression level of each transcript within a single cell. The

colour is consistent with the sub-cluster identity. The red dashed line is a weighted regression to fit a smooth curve through the points (LOESS curve), and the grey area shows a 95% confidence interval.

3.3 Phylogenetic overview of *m130*, *mcp1b* and *tcr* genes

The phylogenetic tree of *m130* (*Supplementary Figure 4A*) cannot with confidence confirm that the suggested *m130* gene model from the annotation is indeed CD163 as the support within the backbone of the phylogeny was too low. However, there was support to place the gene model within the *group B scavenger receptor cysteine-rich (SRCR) superfamily* and that the gene model was closer to CD163/CD163L than the other superfamily members, making it a CD163-like candidate. Thus, we would describe the gene model *m130* as a scavenger receptor superfamily member that may have CD163-like functions. The classical CD163 property is the ability to bind and internalise low density lipoproteins [35], such as the haemoglobin-haptoglobin complex [36].

The proposed *mcp1b* gene model clustered together with several other characterized CC chemokines from Atlantic cod (*Supplementary Figure 4B*). However, the tree did not resolve beyond clustering the major chemokine family members together, the CC and CXC chemokines respectively, and overall displayed no bootstrap support in the backbone of the phylogeny. This indicated that the *mcp1b*/*CCL2* annotation of the gene model is inaccurate and further characterization of the gene model is needed. We would thus describe this gene model as a CC chemokine of unknown function.

Phylogenetic analysis of the annotated tcr genes *tcra* (LOC115548821) and *tcrb* (LOC115544273) supported the identity of these genes, while LOC115548656 was grouped with known *tcrg* genes and LOC115548822 was grouped with known *tcrd* genes, suggesting that these two genes represent Atlantic cod *tcrg* and *tcrd*, respectively (*Supplementary Figure 4C*)

3.4 T lymphocyte sub-clusters

Selecting the 13,753 T lymphocyte cells from the global UMAP and re-clustering them revealed five T cell sub-clusters with distinct gene expression profiles in each (*Figure 4A, 4B and 4C*). Expression of *tcr* genes, *cd3d*, and the *cytotoxic granzyme B* (*gzmb,* LOC115532977) genes within the 13,753 cells confirmed all these clusters are T lymphocytes (Figure 4B). All T cell clusters, except sub-cluster 2, expressed similar levels of *tcra* and *tcrb*. Sub-cluster 2 cells expressed *tcrb* but did not express *tcra*, but instead expressed putative *tcrg* and putative *tcrd*. This cluster also expressed a homolog of *m130* (*Supplementary Figure 4B*), a scavenger receptor also known as CD163 that is typically considered a marker of activated or anti-inflammatory macrophages [37]. T cell sub-cluster 0 was CD8-positive and expressed CC chemokine receptor 7 (*ccr7*) and a second granzyme B paralogue (*gzmb,* LOC115532976). Sub-clusters 1 and 4 expressed *mcp1b* in especially high levels in comparison to the other sub-clusters. Sub-cluster 1 expressed the chemokine *ck5b* (*ccl4)* and granzyme k, *gzmk*, while cluster 4 expressed interleukin 18 receptor accessory protein, *il18rap*. T cell sub-cluster 3 expressed the proliferation markers *mki67* and *pcna*, as well as *incenp*, a regulator of mitosis. The top differentially expressed genes of each T cell sub-cluster can be seen in *Supplementary excel sheet 6*. In summary, although five T cell sub-clusters were revealed by the UMAP, it was noticeable that there were three distinct T cell populations: a population which expressed *cd8* (sub-cluster 0), a population which expressed *mcp1b* (sub-clusters 1 and 4), and a population which expressed *m130* (sub-cluster 2) (*Figure 4B, 4C and 4D*). Sub-cluster 3 represented proliferating T cells, but their relationship to the other three T cell populations remains inconclusive.

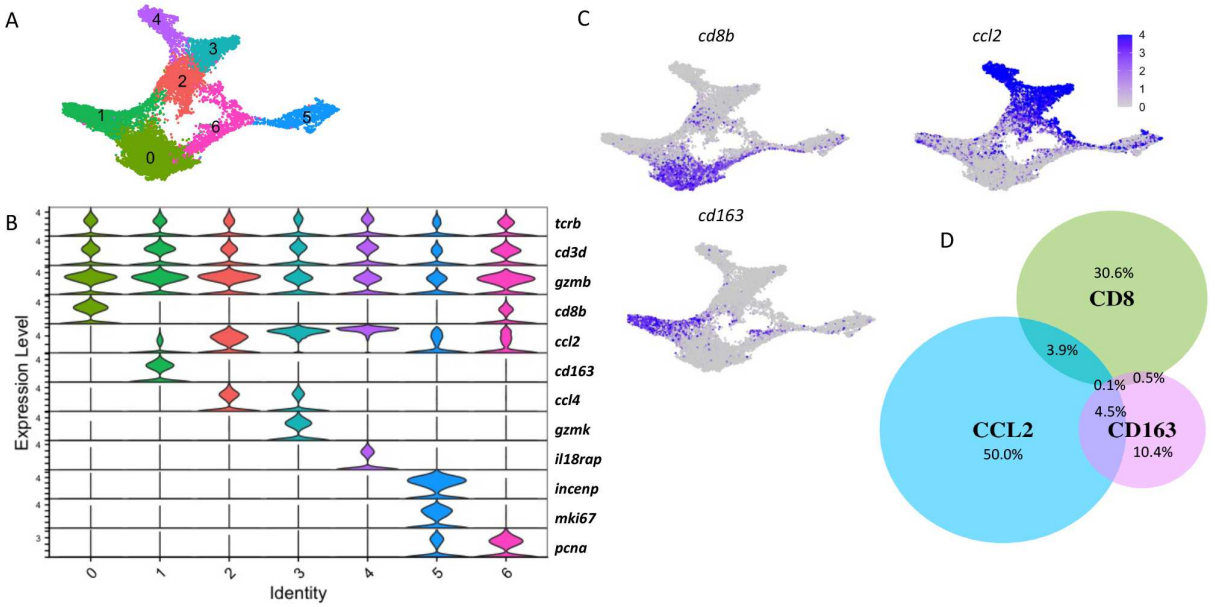

Figure 4. Closer examination of 13,753 Atlantic cod T lymphocyte cells reveals 5 sub-clusters with distinct transcriptome profiles. Expression of T cell co-receptor *cd8b*, cytokine *mcp1b*, and the scavenger receptor *m130* segregates T lymphocytes into three distinct sub-populations.
(A) Visualization of T lymphocyte cells using UMAP reveals five T cell sub-clusters.
(B) Violin plots showing expression of selected genes identified by differential expressed gene analysis. The *y-axis* indicates normalized and log-transformed average expression of the selected genes and the *x-axis* indicates the T cell sub-cluster identity.
(C) UMAP showing the expression level of *cd8b*, *mcp1b*, and *m130* in T lymphocytes. Cells are coloured by expression level.
(D) Venn diagram demonstrating that the expression of *cd8b*, *mcp1b*, and *m130* is largely non-overlapping. The percentage of T cells expressing each gene is shown.

Given the knowledge at hand, we re-analysed the T lymphocytes in our first scRNAseq dataset of cod immune cells [11] and searched for the same genes shown in Figure 4B. We again found that *cd8*, *m130*, and *mcp1b* were expressed in three mutually exclusive populations (*Supplementary Figure 5*), echoing the results shown here. This pattern was also seen in the global UMAP containing all the splenic cells (*Supplementary Figure 6*).

The only other cells in the spleen to express *mcp1b* or *m130* was the GATA3^{+} cytotoxic cells (*Supplementary Figure 6*). *gata3* is a master regulator of T helper 2 cells in mammals [38, 39] and has also been shown to play a role in the development of innate lymphoid cells [40]. This cluster was identified and named in our previous paper [11], and, consistent with our previous description, did not express other T cell markers, supporting our previous hypothesis that this cluster represents an innate lymphoid cell type.

4 Discussion

Multiple genome studies, as exemplified by several teleosts, have revealed a plasticity and adaptability of the vertebrate adaptive immune system. The Atlantic cod [3], and later the entire Gadiformes lineage [4], was shown to have lost the MHC-II/CD4 T cell pathway. Further peculiarities have been found in other teleost species, including the loss of MHC-II/CD4 in pipefish [41, 42] and monkfish [43], and the loss of MHC-II/CD4 alone or in combination with the loss of MHC-I/CD8 in several deep-sea species of anglerfish [44]. In the absence of specific antibodies, it is difficult to assess the immune cell composition as a first step towards understanding the immune system adaptations that have occurred in these newly identified adaptive immune systems. Transcriptomic analysis at the

single cell level is a powerful tool that we have used successfully in characterising immune cell populations of the Atlantic cod [11]. In this study, we have vastly increased the number of immune cells studied by scRNAseq, to 57 thousand spleen cells from 34 fish that were sampled during a vaccination and immune challenge study, compared with a total of 7 thousand blood and spleen cells from two fish in our previous study. The increased number of cells gave us the power to scrutinise in detail the B cell and T cell subsets found within 9,474 B cells and 13,753 T cells that were included in the current study.

Unsupervised sub-clustering of B lymphocytes revealed eight B cell sub-clusters, of which the sub-clusters 5 and 7 are transcriptionally most distinct. Sub-cluster 5 is identical with the plasma cell subset that was identified in the global analysis, which included all splenocytes. Trajectory analysis by Slingshot suggested a common maturation / differentiation pathway of B cells encompassing the first four common progenitor B cell sub-clusters, and then diverging into two separate trajectories each ending up in the presumed end-differentiated B cell sub-cluster 5 (the plasma cells) and sub-cluster 7. That one trajectory ends in the plasma cells, a well-established terminally differentiated cell type, gives strength to the unsupervised ordering of the clusters by Slingshot. In addition, B cell sub-cluster 3, the penultimate sub-cluster in the plasma cell lineage, highly expresses *pa2g4*, a gene involved in cell proliferation and differentiation, further strengthening the trajectory analysis. B cell sub-cluster 7, which is the end point of the other trajectory makes up only 1% of the B lymphocytes. This cluster expresses *thbs1*, *tmsb11* and *ccn2* at a high level, which leads us to speculate that this small subset of highly differentiated B cells may have effector functions that involve communication with other cells. This functionally differentiates this cluster from plasma cells, whose function involves antibody production, and so may be more in line with the regulatory functions of certain B cell subsets that have been recently suggested in mammalian immunology [45]. Functional confirmation of this sub-cluster of B cells is needed when appropriate reagents become available.

The most remarkable finding in the T lymphocyte population is the absence of T cell co-receptor *cd8* expression in several subset, accounting for 65% of all T cells. The majority of the CD8-negative T cells are positive for a CC chemokine initially annotated as *MCP1b*. However, great differences exist between fish and mammalian cytokines, and within fish species, making annotation of cytokines extremely problematic [46-48]. The sequence phylogeny of *mcp1b* gives support that this gene is a C-C motif chemokine but gives little to support this gene model is MCP-1/CCL2. Cytokines rapidly adapt in species and also experience prevalent gene losses and expansions [48]. Lacking functional data and a clear naming alternative we have kept the naming given in gadMor3: *mcp1b*. MCP-1 is a chemoattractant for monocytes and T cells in humans [49]. In humans and mice, the main producers of MCP-1 are myeloid cells (monocytes and macrophages) [50], in addition it is also produced by some non-immune cells such as epithelial and endothelial cells [51]. In our scRNAseq data of the Atlantic cod spleen, we find that *mcp1b* is primarily expressed by T cells and GATA3^{+} cytotoxic cells, and we do not find *mcp1b* expression in the myeloid or stromal cells present in the spleen, further suggesting that *mcp1b* as expressed in Atlantic cod is different from the role of MCP-1 in mammals and does not fulfil the same role. We do not know which genes encode the receptor(s) for *mcp1b* in cod, or which cells that express them. It is intriguing to note that although the expression of *gzmb* is uniformly high in all T-cell sub-clusters, the *gzmb* expression is nevertheless lower in the two sub-clusters, 3 and 4, that show extremely high levels of *mcp1b*. It is thus tempting to speculate that cells in these two sub-clusters have reduced cytotoxic functions and rather may be devoted to signalling with other not-yet identified immune cells through its secretion of the chemokine *mcp1b*.

Most of the T cells which are negative for both CD8 and MCP1b express *m130*. Phylogenetic comparisons of this gene demonstrated that *m130* clustered within the group B SRCR superfamily, and we propose it is CD163-like. In humans and rodents, CD163 is found to be expressed exclusively in cells of monocytic origin [52], which is not found here. The B SRCR superfamily is an ancient and conserved protein domain, and some orthologues genes within this family, CD163c-α-like/WC1, have been shown to play a role in the regulation of gamma delta T cells in cattle [53].

In our scRNAseq data from cod spleen, *cd8*, *mcp1b*, and *m130* are expressed in T lymphocytes in a largely exclusive manner, giving rise to three distinct and separate T cell sub-populations. The CD8-positive, MCP1b-positive, and M130-positive populations all express key T cell markers, such as *cd3*, *tcr* genes, and *gzmb* (*Figure 4A*). There are two developmental scenarios for these three distinct populations: they have either acquired the expression of their respective markers during lymphogenesis in the thymus and exit thymus as three populations of mature T cells that do not interconvert in the periphery; or that they all evolve from the same mature progenitor T cells and represent three differentiation stages. The finding that M130-positive cells uniquely express two genes putatively suggested to be *tcrg* and *tcrd*, indicates that the M130-positive cells represent a separate lineage of T cells. More detailed molecular analysis such as full-length TCR sequencing is needed to ascertain whether these cells may represent gamma delta T cells, the existence of which has not been demonstrated in cod before. Likewise, scRNAseq combined with single-cell TCR sequencing may also reveal the developmental relationship between the various T cell sub-populations, as well as whether some sub-clusters may contain unconventional T cells similar to natural killer T cells or mucosal associated invariant T cells that show restricted and semi-invariant V-gene usage in mammalian systems [54, 55]. The CD8-positive population uniquely expresses chemokine receptor ccr7 suggesting this subset may have a different trafficking pattern than the other CD8-negative T cell subsets.

The B and T cell differentiation pathways in teleosts are poorly understood, and there may be substantial differences between different lineages of teleosts. An alternative to use of classical surface markers is to identify transcription factors which are differentially expressed along predicted lineages. *pax5* and *e2a* are well-known genes participating in B cell differentiation and activation [31, 32], and have even been identified in teleost B cell lymphogenesis [56]. *pax5* and *e2a* are among the top differentially expressed genes in the predicted B cell lineages in this study, suggesting these genes also play an important role in B cell differentiation in the Atlantic cod. Expression of the gene *tnfrsf13b* was shown to be increased in the cod plasma cells, results in alignment with human studies suggesting *tnfrsf13b* has a role in the development of plasma cells and production of much of the Ig in blood [33, 34]. The TACI surface receptor encoded by this gene is suggested to be involved in T cell-independent antibody responses [57], which is interesting given the absence of CD4 T helper cells in cod.

It is interesting to note that the two novel T cell population markers we have discussed here, namely *m130* and *mcp1b*, are only expressed in the GATA3^{+} cytotoxic cell population in addition to the T cells. Despite sharing these unique markers, the GATA3^{+} cytotoxic cells do not express *tcr* and are transcriptionally distant from the other lymphocyte clusters, as visualised by the global UMAP, and so we cannot classify them as T cells. We previously hypothesised these cells represent a population of innate lymphoid cells [11], and the additional findings here make this a population of interest for future studies.

Here we present a refined characterisation of the T and B lymphocytes in the Atlantic cod spleen using single-cell transcriptomics. The discovery of novel T cell subset markers gives guidance to the future direction of studying T cell development, differentiation, and function in Atlantic cod. In addition, the trajectories analysis reveals candidate transcription factors that drive B cell differentiation. These hypotheses need to be validated in future studies.

5 Acknowledgements

Financing from the “comparative immunology of fish and humans (COMPARE)” project at the University of Oslo.
In collaboration with the “rational design of vaccination strategies for Atlantic cod (VACSACOD)” program, Research council of Norway.
Norwegian Sequencing Centre provided sequencing services.
We thank Henning Sørum (Norwegian University of Life Sciences) for discussion and planning of the immunisation experiments.

6 References

1. Schluter, S.F., et al., *'Big Bang' emergence of the combinatorial immune system.* Dev Comp Immunol, 1999. **23**(2): p. 107-11.
2. Flajnik, M.F., *A cold-blooded view of adaptive immunity.* Nat Rev Immunol, 2018. **18**(7): p. 438-453.
3. Star, B., et al., *The genome sequence of Atlantic cod reveals a unique immune system.* Nature, 2011. **477**: p. 207.
4. Malmstrom, M., et al., *Evolution of the immune system influences speciation rates in teleost fishes.* Nat Genet, 2016. **48**(10): p. 1204-1210.
5. Pilström, L., G.W. Warr, and S. Strömberg, *Why is the antibody response of Atlantic cod so poor? The search for a genetic explanation.* Fish Sci, 2005. **71**(5): p. 961-971.
6. Lund, V., S. Bordal, and M.B. Schroder, *Specificity and durability of antibody responses in Atlantic cod (Gadus morhua L.) immunised with Vibrio anguillarum O2b.* Fish Shellfish Immun, 2007. **23**(4): p. 906-910.
7. Caipang, C.M.A., M.F. Brinchmann, and V. Kiron, *Profiling gene expression in the spleen of Atlantic cod, Gadus morhua upon vaccination with Vibrio anguillarum antigen.* Comp Biochem Phys A, 2009. **153**(3): p. 261-267.
8. Gudmundsdottir, S., et al., *Specific and natural antibody response of cod juveniles vaccinated against Vibrio anguillarum.* Fish Shellfish Immun, 2009. **26**(4): p. 619-624.
9. Magnadottir, B., et al., *Natural antibodies of cod (Gadus morhua L.): Specificity, activity and affinity.* Comp Biochem Phys B 2009. **154**(3): p. 309-316.
10. Mikkelsen, H., et al., *Vibriosis vaccines based on various sero-subgroups of Vibrio anguillarum O2 induce specific protection in Atlantic cod (Gadus morhua L.) juveniles.* Fish Shellfish Immun, 2011. **30**(1): p. 330-339.
11. Guslund, N.C., et al., *Single-Cell Transcriptome Profiling of Immune Cell Repertoire of the Atlantic Cod Which Naturally Lacks the Major Histocompatibility Class II System.* Front Immunol, 2020. **11**: p. 559555.
12. Macosko, E. and M. Goldman. *Drop-seq laboratory protocol*. 2015; Available from: http://mccarrolllab.org/dropseq/Online-Dropseq-Protocol-v.-3.1-Dec-2015.pdf.
13. Macosko, Evan Z., et al., *Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets.* Cell, 2015. **161**(5): p. 1202-1214.
14. Namesh, J. *Drop-seq core computational protocol version 1.0.1*. 2015; Available from: http://mccarrolllab.org/wp-content/uploads/2016/03/Drop-seqAlignmentCookbookv1.2Jan2016.pdf.
15. Torresen, O.K., et al., *An improved genome assembly uncovers prolific tandem repeats in Atlantic cod.* BMC Genomics, 2017. **18**(1): p. 95.
16. Hao, Y., et al., *Integrated analysis of multimodal single-cell data.* Cell, 2021. **184**(13): p. 3573-3587 e29.
17. Stuart, T., et al., *Comprehensive Integration of Single-Cell Data.* Cell, 2019. **177**(7): p. 1888-1902 e21.
18. Butler, A., et al., *Integrating single-cell transcriptomic data across different conditions, technologies, and species.* Nat Biotechnol, 2018. **36**(5): p. 411-420.
19. Satija, R., et al., *Spatial reconstruction of single-cell gene expression data.* Nat Biotechnol, 2015. **33**(5): p. 495-502.

20. McInnes, L., et al., *UMAP: Uniform Manifold Approximation and Projection.* Journal of Open Source Software, 2018. **3**: p. 861.
21. Street, K., et al., *Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics.* BMC Genomics, 2018. **19**(1): p. 477.
22. Villanueva, R.A.M. and Z.J. Chen, *ggplot2: Elegant Graphics for Data Analysis, 2nd edition.* Measurement-Interdisciplinary Research and Perspectives, 2019. **17**(3): p. 160-167.
23. McCarthy, D.J., et al., *Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R.* Bioinformatics, 2017. **33**(8): p. 1179-1186.
24. Camacho, C., et al., *BLAST+: architecture and applications.* BMC Bioinformatics, 2009. **10**: p. 421.
25. Kumar, S., G. Stecher, and K. Tamura, *MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets.* Molecular Biology and Evolution, 2016. **33**(7): p. 1870-1874.
26. Kozlov, A.M., et al., *RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference.* Bioinformatics, 2019. **35**(21): p. 4453-4455.
27. Darriba, D., et al., *ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models.* Molecular Biology and Evolution, 2019. **37**(1): p. 291-294.
28. Rambaut. 2009; Available from: https://github.com/rambaut/figtree/.
29. Shaulian, E. and M. Karin, *AP-1 as a regulator of cell life and death.* Nat Cell Biol, 2002. **4**(5): p. E131-6.
30. Liu, Z., et al., *Ebp1 isoforms distinctively regulate cell survival and differentiation.* Proc Natl Acad Sci U S A, 2006. **103**(29): p. 10917-22.
31. Cobaleda, C., et al., *Pax5: the guardian of B cell identity and function.* Nat Immunol, 2007. **8**(5): p. 463-70.
32. Kwon, K., et al., *Instructive role of the transcription factor E2A in early B lymphopoiesis and germinal center B cell development.* Immunity, 2008. **28**(6): p. 751-62.
33. Tsuji, S., et al., *TACI deficiency impairs sustained Blimp-1 expression in B cells decreasing long-lived plasma cells in the bone marrow.* Blood, The Journal of the American Society of Hematology, 2011. **118**(22): p. 5832-5839.
34. Tsuji, S., et al., *TACI deficiency enhances antibody avidity and clearance of an intestinal pathogen.* The Journal of clinical investigation, 2014. **124**(11): p. 4857-4866.
35. Brown, M.S. and J.L. Goldstein, *Receptor-mediated endocytosis: insights from the lipoprotein receptor system.* Proceedings of the National Academy of Sciences of the United States of America, 1979. **76**(7): p. 3330-3337.
36. Kristiansen, M., et al., *Identification of the haemoglobin scavenger receptor.* Nature, 2001. **409**(6817): p. 198-201.
37. Abraham, N.G. and G. Drummond, *CD163-Mediated hemoglobin-heme uptake activates macrophage HO-1, providing an antiinflammatory function.* Circ Res, 2006. **99**(9): p. 911-4.
38. Lee, G.R., P.E. Fields, and R.A. Flavell, *Regulation of IL-4 Gene Expression by Distal Regulatory Elements and GATA-3 at the Chromatin Level.* Immunity, 2001. **14**(4): p. 447-459.
39. Yagi, R., J. Zhu, and W.E. Paul, *An updated view on transcription factor GATA3-mediated regulation of Th1 and Th2 cell differentiation.* Int immunol, 2011. **23**(7): p. 415-420.

40. Tindemans, I., et al., *GATA-3 Function in Innate and Adaptive Immunity.* Immunity, 2014. **41**(2): p. 191-206.
41. Haase, D., et al., *Absence of major histocompatibility complex class II mediated immunity in pipefish, Syngnathus typhle: evidence from deep transcriptome sequencing.* Biol Lett, 2013. **9**(2): p. 20130044.
42. Roth, O., et al., *Evolution of male pregnancy associated with remodeling of canonical vertebrate immunity in seahorses and pipefishes.* Proc Natl Acad Sci U S A, 2020. **117**(17): p. 9431-9439.
43. Dubin, A., et al., *Complete loss of the MHC II pathway in an anglerfish, Lophius piscatorius.* Biol Lett, 2019. **15**(10): p. 20190594.
44. Swann, J.B., et al., *The immunogenetics of sexual parasitism.* Science, 2020. **369**(6511): p. 1608-1615.
45. Shang, J., H. Zha, and Y. Sun, *Phenotypes, Functions, and Clinical Relevance of Regulatory B Cells in Cancer.* Front Immunol, 2020. **11**: p. 582657.
46. Buchmann, K., *Evolution of Innate Immunity: Clues from Invertebrates via Fish to Mammals.* Front Immunol, 2014. **5**: p. 459.
47. Zhu, L.Y., et al., *Advances in research of fish immune-relevant genes: a comparative overview of innate and adaptive immunity in teleosts.* Dev Comp Immunol, 2013. **39**(1-2): p. 39-62.
48. Bird, S. and C. Tafalla, *Teleost Chemokines and Their Receptors.* Biology (Basel), 2015. **4**(4): p. 756-84.
49. Carr, M.W., et al., *Monocyte chemoattractant protein 1 acts as a T-lymphocyte chemoattractant.* Proc Natl Acad Sci U S A, 1994. **91**(9): p. 3652-6.
50. Yoshimura, T., et al., *Human monocyte chemoattractant protein-1 (MCP-1). Full-length cDNA cloning, expression in mitogen-stimulated blood mononuclear leukocytes, and sequence similarity to mouse competence gene JE.* FEBS Lett, 1989. **244**(2): p. 487-93.
51. Deshmane, S.L., et al., *Monocyte chemoattractant protein-1 (MCP-1): an overview.* J Interferon Cytokine Res, 2009. **29**(6): p. 313-26.
52. Barbe, E., et al., *Characterization and expression of the antigen present on resident rat macrophages recognized by monoclonal antibody ED2.* Immunobiology, 1990. **182**(1): p. 88-99.
53. Herzig, C.T.A., et al., *Evolution of the CD163 family and its relationship to the bovine gamma delta T cell co-receptor WC1.* BMC Evolutionary Biology, 2010. **10**(1): p. 181.
54. Bendelac, A., et al., *MOUSE CD1-SPECIFIC NK1 T CELLS: Development, Specificity, and Function.* Annual Review of Immunology, 1997. **15**(1): p. 535-562.
55. Treiner, E., et al., *Selection of evolutionarily conserved mucosal-associated invariant T cells by MR1.* Nature, 2003. **422**(6928): p. 164-169.
56. Zwollo, P., *Dissecting teleost B cell differentiation using transcription factors.* Dev Comp Immunol, 2011. **35**(9): p. 898-905.
57. von Bülow, G.U., J.M. van Deursen, and R.J. Bram, *Regulation of the T-independent humoral response by TACI.* Immunity, 2001. **14**(5): p. 573-82.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementaryexcelsheets.xlsx
- Supplementaryfigures.pdf