Single-cell RNA sequencing reveals influences of rearing environment on cellular immunity in brown trout (*Salmo trutta*)

James Ord[1,2]✉, Helena Saura Martinez[1], Monica Hongroe Solbakken[3], Anastasiia Berezenko[1], Simone Oberhaensli[1], Stephanie Talker[4], Heike Schmidt-Posthaus[1]✉, Irene Adrian-Kalchhauser[1]✉

[1] Institute for Fish and Wildlife Health, Department of Infectious Diseases and Pathobiology, Vetsuisse Faculty, University of Bern, Länggassstrasse 122, 3012 Bern, Switzerland

[2] Organismal and Evolutionary Biology Research Programme, Faculty of Biological and Environmental Sciences, University of Helsinki, 00790 Helsinki, Finland

[3] Medical Genetics Unit, Faculty of Veterinary Medicine, Norwegian University of Life Sciences, Postboks 5003, 1432 Ås, Norway

[4] Institute of Virology and Immunology, Department of Infectious Diseases and Pathobiology, Vetsuisse Faculty, University of Bern, Länggassstrasse 122, 3001 Bern, Switzerland

✉ jms.ord18@gmail.com, ✉ heike.schmidt@unibe.ch, ✉ irene.adrian-kalchhauser@unibe.ch

ORCID

James Ord: 0000-0001-6870-8927

Monica Hongroe Solbakken: 0000-0002-9677-403X

Simone Oberhaensli: 0000-0002-2116-1685

Stephanie Talker: 0000-0002-3966-6251

Heike Schmidt-Posthaus: 0000-0003-2628-4454

Irene Adrian-Kalchhauser: 0000-0002-7563-2577

Abstract

The ability to profile gene expression at the single cell level can offer unprecedented insight into the vertebrate immune system. While such profiling of non-model organisms can present unique challenges (e.g. evolutionary idiosyncrasies), high resolution data on immune function could yield valuable insights into ecologically and economically important species. Here, we used single-cell RNA sequencing (scRNAseq) to profile the immune system transcriptional repertoire of brown trout (*Salmo trutta*), a threatened index species and commonly using for stocking endangered wild populations, using multiple individuals derived from three distinct rearing backgrounds (wild-caught, farm reared for one generation or farm-reared for at least three generations). We subsequently present a collection of markers that are representative of the species accounting for disparate rearing histories. This revealed a repertoire which is broadly evolutionarily conserved, although many of the genes characterising neutrophil, macrophage, T-cell, B-cell and other lineages likely reflect the specific facets of salmonid immune evolution. The fates of duplicate gene pairs putatively arising from the salmonid-specific whole genome duplication event were also investigated, revealing that approximately half of duplicate genes examined show some degree of functional divergence within the immune system in terms of loss or gain of expression in certain clusters. Finally, comparing cluster-level gene expression profiles between different rearing backgrounds revealed that single or multigenerational exposure to a farm environment induces transcriptional alterations in neutrophils, macrophages, B-cells and T-cells.

Introduction

Vertebrate immune systems comprise some of the most complex systems in nature, exhibiting extraordinary diversity between species (Milgroom, 2023). Differences range from the structural level (e.g. in teleost fish, the kidneys are the principal source of immune cells, not bone marrow as in mammals), to the diversity of cell types (distinct cell types found in some species but not others), enzymes, and gene repertoire (Flies, 2020). New sequencing technologies, notably single-cell RNA sequencing now offer the opportunity to map the genetic regulation of vertebrate immune systems in unprecedented detail, including in non-model organisms. Such analyses could yield insight into the evolution of the immune system, as well as providing data of great utility for veterinary medicine. However, non-model organisms present unique analytical challenges, particularly those that are highly divergent from humans and model species (mostly mammals) from which derive most of the prior knowledge on gene function. A case in point is the salmonid lineage, in which a whole genome duplication (WGD) event approx. 80MYA resulted in duplicated genes, many of which may have diverged in function thus making cell type identification difficult when relying on established markers from mammals.

Brown trout (*Salmo trutta*) are an ecologically and economically important member of the salmonid family of fishes (Lobón-Cerviá & Sanz, 2017). They are valued by recreational anglers and contribute significantly to the economically important sport fishing industry through tourism, sale of equipment,

and access rights. They also play a crucial ecological role as apex predators in lotic alpine streams by controlling populations of smaller fish and insects and influence the overall structure and health of freshwater ecosystems through cascading effects. Based on their specific requirements regarding temperature, river structure, and water quality, they serve as sentinel species for riverine health. Brown trout populations suffer earlier than other species from pollution, habitat degradation, or changes in water temperature, and conservation efforts focusing on protecting brown trout habitats can have positive effects on the overall conservation of aquatic biodiversity. As a species of cultural and heritage value, brown trout can also serve as flagship species to promote ecosystem conservation and protective management. At the same time, brown trout are under pressure from global change through the combined effects of changes in river structure, climate change, the translocation of competitor species and pathogens, and management practices. Their ability to mount appropriate immune responses is sensitive to temperature (Bailey et al., 2017; Strepparava et al., 2018) and is additionally impacted by water quality parameters (Rehberger et al., 2020; Schmidt-Posthaus et al., 2013). To support wild populations, large stocking programs are in place, yet the stocked fish fail to thrive (Baer et al., 2023), may be less resistant to pathogens, and the conditions experienced during raising offspring in artificial environments before release affect disease incidence (Palikova et al., 2017). This renders the immune system regulation of brown trout an important area of study.

The brown trout immune system is expected to largely follow the standard vertebrate build – i.e. innate and adaptive immune systems with many conserved cell types – but remains to be described in detail. Brown trout, as salmonids, have been subject to WGD followed by rediploidisation (return to disomic inheritance), accompanied by loss or retention of the resulting gene copies (Gillard et al., 2021; Macqueen & Johnston, 2014; Robertson et al., 2017). Amongst genes retained as duplicates, fates may include pseudogenization (because loss of function mutations are less detrimental), neofunctionalization (divergence of function), and sub-functionalization (division of roles in the same function) between the retained copies. For the related salmonid rainbow trout, 6733 pairs of WGD-derived gene duplicates (termed 'ohnologs') have been identified (Berthelot et al., 2014), of which only one quarter showed evidence of retaining the same function (as indicated by highly correlated expression patterns and similar expression levels across tissues). Brown trout is phylogenetically located at an intermediate position between the well-studied whitefishes (Coregonidae) and trouts (Oncorhynchidae) (Crête-Lafrenière et al., 2012). They share a common ancestor with salmon (*Salmo Salar*) around 12 Mya, and with rainbow trout (Oncorhynchus mykiss) around 30 Mya (Crête-Lafrenière et al., 2012; Lecaudey et al., 2018). The specific taxonomic relations, history of hybridisation and isolation, and even a potential multispecies taxonomy for brown trout populations are currently being discussed (Hashemzadeh Segherloo et al., 2021; Marić et al., 2023; Veličković et al., 2023). Understanding patterns of neofunctionalization in brown trout has the potential to complement existing data and provide perspective with regard to species-specific versus universal effects of WGD.

The immune system in fish features particularly interesting evolutionary trajectories. Teleost immune gene repertoires are highly specialized for each species, and feature prevalent non-canonical features compared to well-studied model systems such as loss of MHCII, novel subfamilies within NLR and TLR pattern recognition receptor families and loss of antibodies (Malmstrøm et al., 2016; Mirete-Bachiller et al., 2021; Solbakken et al., 2017; Suurväli et al., 2022; Swann et al., 2020). Therefore, the annotation of immune genes in fish can be challenging from genome sequence alone, where gene models may be incomplete or lack gene identification due to sequence divergence, or be misannotated due to highly variable gene families. For brown trout, there is a high-quality reference genome which would provide a solid foundation for downstream analyses but detailed functional annotations are lacking. Recent attempts to characterize the immune gene repertoire of brown trout have revealed extensive gene duplications in many pathways involved in immune cell differentiation and proliferation, suggesting many duplicates of immune-related genes have indeed been retained following WGD (Colgan et al., 2021). The functional implications of these duplications remain to be determined.

This study provides baseline data for the healthy brown trout cellular immune system using single cell RNA sequencing of enriched kidney-derived mononuclear cell populations from nine *S. trutta* individuals and analyses the impact of long-term evolutionary processes (whole genome duplication) and short-term environmental exposures (recent rearing history) on cellular immunity. We use single cell RNA sequencing of brown trout from three different sources to a) describe the immune cell types of brown trout, b) identify novel markers for particular cell types, c) investigate expression patterns of immune gene ohnologues to identify neofunctionalizations, and d) describe the effects of rearing environment on cellular immunity. We identify cell types based on previously established marker genes, explore the identity of cell types with non-canonical gene expression, and explore their gene expression patterns, and thus offer a set of marker genes for brown trout to use in future RNA-based analyses of immune functions in the absence of highly specific antibodies. To explore evolutionary peculiarities in brown trout, we analyse expression patterns of salmonid-specific ohnologues and propose functional diversification patterns for some of them. To investigate the impact of short-term environmental challenges on immune function, we compare three groups of brown trout with natural vs farmed rearing histories in terms of gene expression profiles at the cluster-level.

Results

Clustering of single cell expression profiles reveals distinct immune cell populations and subpopulations based on putative cell lineage markers

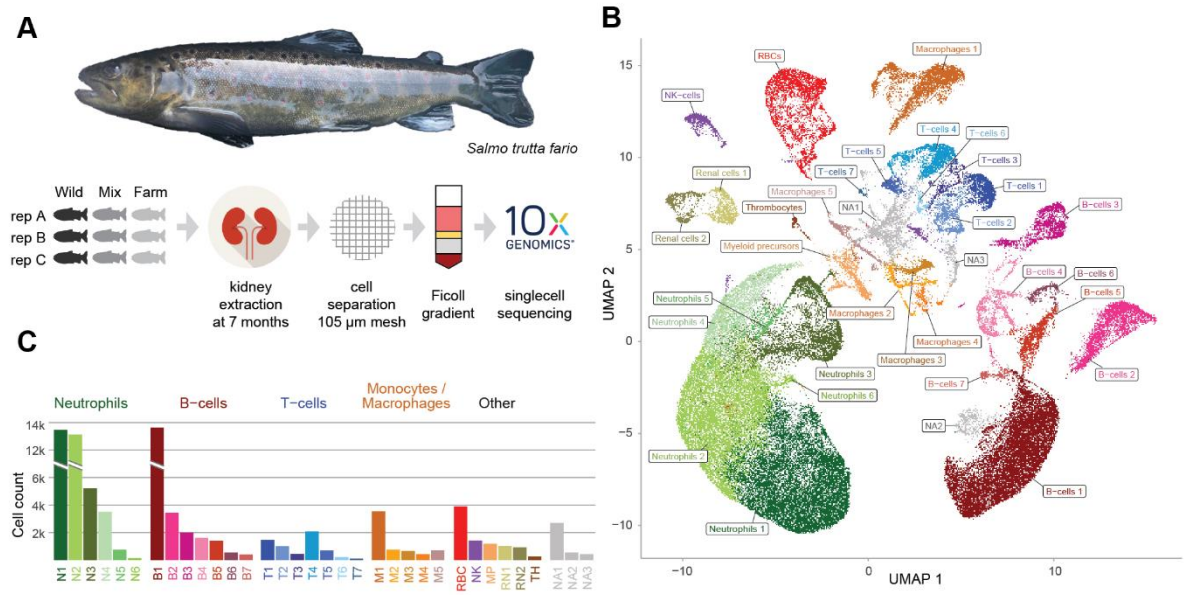

Figure 1. Brown trout immune cell clusters **(A)** Experimental setup. Kidney tissue was dissected from juvenile *Salmo trutta*, from which hematopoietic cell extracts were used as input for single-cell RNAseq (scRNAseq). **(B)** 34 immune cell clusters. UMAP projection of 83,847 cells grouped into 34 clusters according to shared gene expression variation. Of the 34 clusters, 31 were assigned putative identities (different colours and labels on the plot). **(C)** Cluster sizes. Total counts of cells contained in each of the 34 clusters, grouped according to cell lineage.

In a first step, we used single cell RNA sequencing from hematopoietic cells enriched from nine independent brown trout head and trunk kidney tissue samples (the principal source of immune cells) to identify putative immune cell types amongst 83,847 cells. The sourced individuals derived from three different backgrounds: (1) wild-caught, (2) a single generation of hatchery rearing or (3) multiple generations of hatchery rearing. An initial set of 29 clusters (**Fig. S1**) was obtained from the top 2000 variably expressed genes, and could be refined to 34 final clusters based on the expression of lineage markers (**Fig. 1B**). To assign putative cluster identities, we compiled a set of marker genes of major immune cell lineages based on orthology (namely to humans and fish, hereafter 'prior markers') and identified fifty such prior markers amongst variably expressed genes (**Table 1**; see also **Table S1** for full list of prior markers), some of which comprised multiple gene copies (e.g. *mpx*, *mmp9*, *CD74*). Prior markers were grouped into 'modules' according to the lineages they associate with, and summarised expression values were derived for each module for comparison across clusters. This enabled identification of broad cell lineages, namely neutrophils, monocytes / macrophages, B-cells, T-cells, and others (**Fig. 2**). The largest clusters comprised subsets of putative neutrophils (N1 and N2) and B-cells (B1), respectively (**Fig. 1C**).

Table 1. Prior markers of four major immune cell lineages whose S. trutta homologues were expressed in the scRNAseq dataset.

Cell lineage	Protein name (alternate name)	*S.trutta* homologue gene name (ENSEMBL)	*S. trutta* homologue ENSEMBL gene ID
Neutrophils	CD182 (CXCR2)		ENSSTUG00000013662
	CPA5	*CPA1*	ENSSTUG00000021190
	HCE*		ENSSTUG00000011854
	LCE*	*c6ast1*	ENSSTUG00000007607
	LYZ*	*LYZ*	ENSSTUG00000025205
	MMP9	*mmp9*	ENSSTUG00000015832
	MMP9	*mmp9*	ENSSTUG00000038105
Macrophages / Monocytes	CD206 (MRC1)	*MRC1*	ENSSTUG00000000974
	CD74		ENSSTUG00000006024
	CD74		ENSSTUG00000008754
	COX2 (PTGS2A)	*ptgs2a*	ENSSTUG00000026620
	IRF8	*irf8*	ENSSTUG00000021023
	MARCO	*MARCO*	ENSSTUG00000016758
	MPEG1.1	*MPEG1*	ENSSTUG00000031848
	MPO (MPX)	*mpx*	ENSSTUG00000006763
	MPO (MPX)	*mpx*	ENSSTUG00000007163
T-cells	BCL11B		ENSSTUG00000025217
	CD28		ENSSTUG00000013913
	CD28		ENSSTUG00000041050
	CD3D		ENSSTUG00000034782
	CD3E		ENSSTUG00000020148
	CD4	*CD4-2a*	ENSSTUG00000038814
	CD4	*CD4-1*	ENSSTUG00000038833
	CD8A		ENSSTUG00000018765
	CD8B		ENSSTUG00000037440
	FOXP3		ENSSTUG00000024213
	GATA3	*gata3*	ENSSTUG00000023944
	IL17A (a/f in fish)*	*il17a/f1*	ENSSTUG00000016782
	IL21	*il21*	ENSSTUG00000042553
	IL7R		ENSSTUG00000021260
	TBX21 (T-bet)	*tbx21*	ENSSTUG00000037670
	TCF7		ENSSTUG00000002790
	TCF7	*tcf7*	ENSSTUG00000023812
	TRBC1 TCR beta constant 1		ENSSTUG00000037152
	TRGC1 TCR gamma constant 1		ENSSTUG00000007088
	TRGC1 TCR gamma constant 1		ENSSTUG00000007103
	TRGC1 TCR gamma constant 1		ENSSTUG00000007109
	ZAP70	*zap70*	ENSSTUG00000023942
B-cells	AICDA	*aicda*	ENSSTUG00000016342
	BCR heavy IgD		ENSSTUG00000016267
	BCR heavy IgM		ENSSTUG00000010446

BCR light IgL kappa 1 IGKC1	*IGKC*	ENSSTUG00000027283
BCR light IgL kappa 2 IGKC2	*IGKC*	ENSSTUG00000028030
BCR light IgL lambda IGLC1		ENSSTUG00000001951
BCR light IgL lambda IGLC1		ENSSTUG00000015148
CD37		ENSSTUG00000006678
CD79A	*cd79a*	ENSSTUG00000031393
CD79B		ENSSTUG00000027061
CXCR5 (CD185)	*CXCR5*	ENSSTUG00000007343
TNFRSF9	*tnfrsf9a*	ENSSTUG00000031635

*Fish-specific

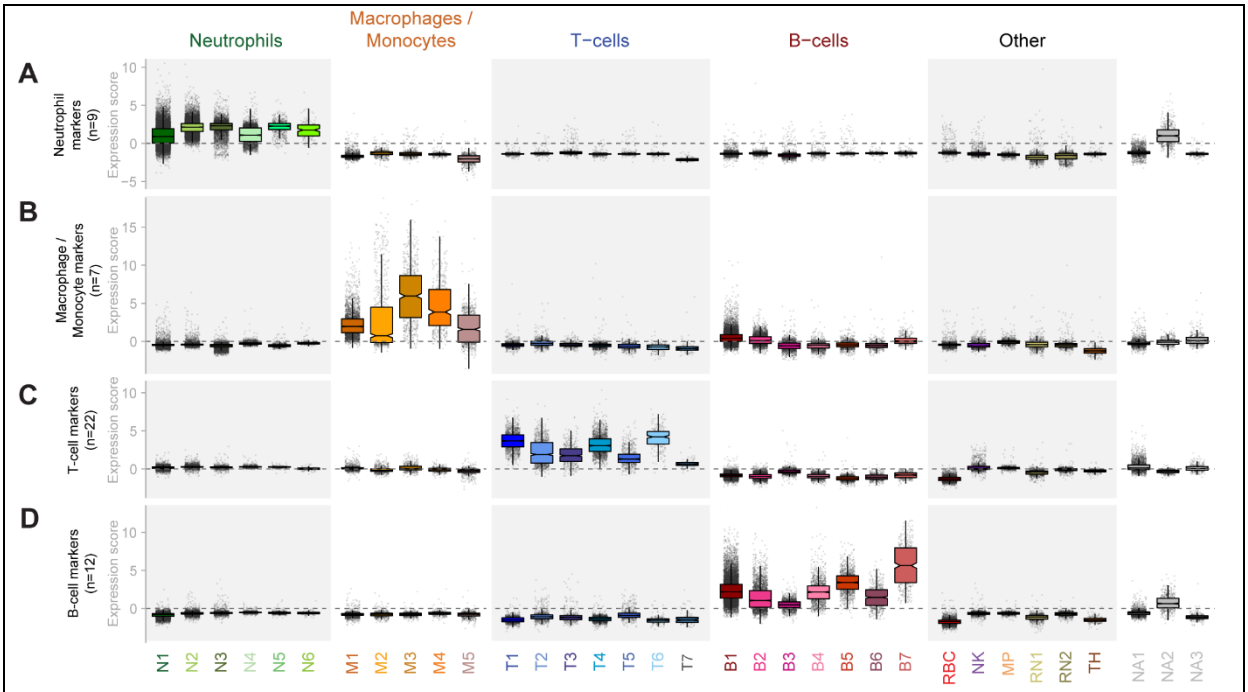

Figure 2. Expression scores per-cell of four gene modules, comprised of prior markers of cell lineages, across 34 cell clusters. Gene modules were comprised of 9 genes for the neutrophil lineage **(A)**, 7 for the monocyte / macrophage lineage **(B)**, 22 for the T-cell lineage **(C)**, and 12 for the B-cell lineage **(D)**.

At a finer scale, putative cluster identities were inferred from genes that were differentially expressed between clusters (hereafter 'cluster markers'; **Table S2**), considering principally those with a Log2FC of ≥ 1 and expression in at least 60% of cells in a cluster. These amounted to 623 total marker genes, of which 600 were ENSEMBL genes and 23 were identified from our augmented annotation. Of these, 295 genes were markers of a single cluster and 328 genes were markers of more than one cluster.

Subsequently, we were able to assign putative identities to 31 out of 34 clusters (80,166 out of 83,847 cells) (**Fig. 1B**). Three clusters were not assigned to a putative identity either due to lack of prior marker expression, lack of cluster marker expression in a high enough proportion of cells, or ambiguous expression patterns (e.g. markers of multiple cell types). For example, cluster NA2, which clustered next to B-cells, showed elevated expression of both B-cell and neutrophil markers and thus could not be

assigned to either lineage. Of the 31 clusters to which we were able to assign a putative identity, only two (RN1 and RN2, renal cells).

Expression patterns of proliferation markers reflect shared regulatory architecture across immune cell lineages

In addition to identifying cell lineages, we were also able to identify actively proliferating cell populations according to expression of homologues of prior markers of cell proliferation *pcna* (two homologues) and *mki67* (one homologue). Actively proliferating cell populations were thus identified as the neutrophil cluster N3, T-cell cluster T6 and B-cell cluster B4, as well as actively proliferating subpopulations within the major macrophage cluster M1 and the putative myeloid progenitor cluster MP (**Fig. 3**). Additional genes associated with proliferation, namely *H2AZ* and *pclaf* showed identical expression patterns across these otherwise separate clusters.

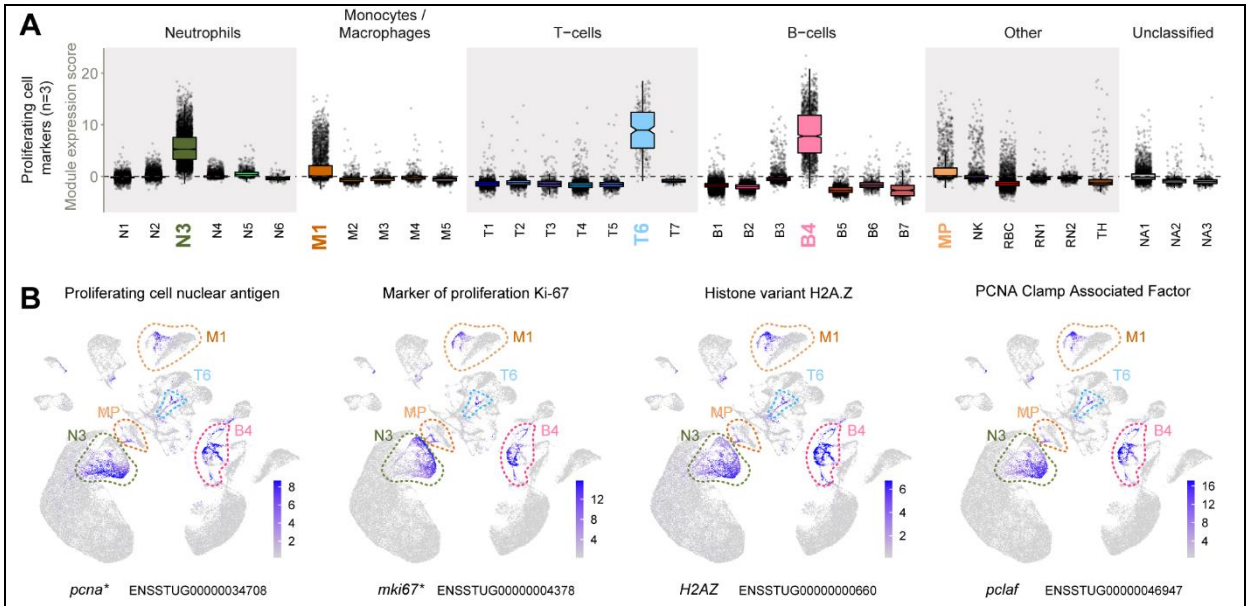

Figure 3. Proliferation marker expression. **(A)** Expression scores per-cell of a gene module comprised of three prior markers of cell proliferation, across 34 clusters. Clusters showing high proportions of inferred proliferating cells are highlighted in bold on the x-axis. **(B)** UMAP projections showing normalised expression values two prior markers of proliferation (*pcna* and *mki67*) and two further genes showing similar expression patterns. Clusters showing high proportions of inferred proliferating cells are circled. Gene names are as given in the ENSEMBL annotation (without altering case). *Denotes gene is in a list of prior markers used to infer putative cell type or state.

Major neutrophil subpopulations are distinguished by expression of high choriolytic enzyme

Neutrophils were identified as a set of clusters (N1 to N6, **Fig. 4A**) that formed a large agglomeration, consistent with neutrophils being the most abundant immune cell type and with a relative homogeneity compared to other immune cell types (although neutrophil heterogeneity is increasingly appreciated, see

(Qu et al., 2023)). Six prior neutrophil markers (HCE1-like, *cpa9*, *mpx*, and two copies of *mmp9*) also constituted high abundance markers in one or more of these clusters.

However, expression of cluster markers was heterogeneous across clusters. Indeed, when considering high abundance cluster markers that were exclusive to the neutrophil clusters (total of 78 markers; **Table S3**), none were common to all six (**Fig. 4B**). The largest category of markers comprised those shared between the two largest clusters, N1 and N2. Given that these two clusters comprised the majority of neutrophils we considered these shared markers to be 'pan-neutrophil' markers (i.e. markers broadly representative of neutrophils) which included *gpx1b*, *taldo1* (two copies), *mpx*, *tktb*, FBP1-like and *alox5ap* (**Fig. 4C**). Notably, human homologues of *taldo1* and *tktb* and FBP1 have been shown to play a role in initiating the oxidative burst in neutrophils in response to LPS exposure (Britt et al., 2022; Sadiku et al., 2021). The GO term 'carbohydrate metabolic processes', which includes FBP1-like, was enriched amongst markers of both N1 and N2, indicating a high energetic demand which would be consistent with pro-inflammatory cells. N6 displayed high mitochondrial gene expression (**Fig. S2**), fewer expressed genes and low overall gene expression which may imply dead or dying cells.

Second and third categories of neutrophil markers comprised those expressed either in the lower cluster N1 or in the upper clusters N2-6, respectively. Most strikingly, HCE1-like (high choriolytic enzyme 1-like) was highly expressed in all clusters except for N1 (though only partially in N4; **Fig. 4D**) and was unique in that its expression marked a pronounced border between N1 and N2 in contrast to other genes in this category which showed some residual expression in N1. HCE is responsible for breaking down the chorion during hatching of teleost fish (Nagasawa et al., 2022) but has more recently emerged as a marker of teleost granulocytes (Perdiguero et al., 2024; Wu et al., 2021). Other genes in this category included CFD1-like (complement-factor D-like), the human homologue of which has been shown to inhibit leukocyte degranulation, a process that prevents host tissue damage (Balke et al., 1995) and *cpa1*, a prior neutrophil marker and metallopeptidase. Meanwhile, genes showing elevated expression in cluster N1 were consistently enzymes including the metalloproteases *mmp9, adam28*, and *mmp13*, and glutamate decarboxylase *GADL1* (**Fig. 4E**). Indeed, 'metallopeptidase activity' was a significantly enriched GO term amongst N1 markers. Lysozyme (*lyz*), a neutrophil-specific marker in zebrafish (Harvie & Huttenlocher, 2015), was also a marker of N1 but, as a macrophage activation marker (Keshav et al., 1991), was also active in macrophage clusters (see **Fig. S3**). Pseudotime analysis of neutrophil clusters rooted at the putative myeloid progenitor cluster (MP, see **Fig. 8**) implied that N1 cells were terminally differentiated (**Fig. S4A**).

N3 was clearly a pool of actively proliferating cells (see **Fig. 3**) while N4 and N5 lacked unique markers and may comprise intermediate or immature cells. Collectively, the results highlight two major neutrophil subpopulations with specialized roles: one characterised by high expression of HCE1-like and another by abundant metallopeptidase gene expression.

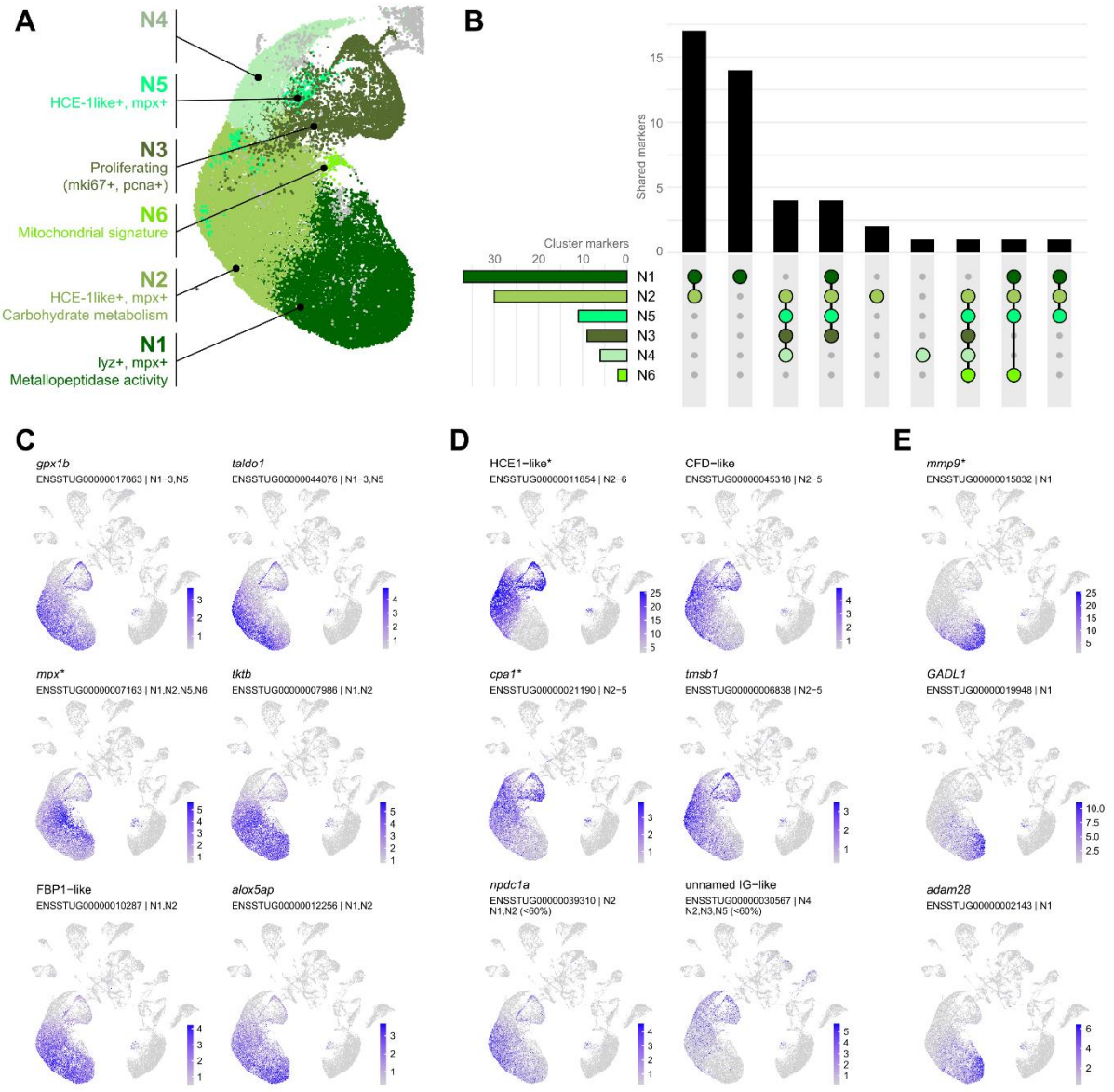

Figure 4. Neutrophil clusters and example expression patterns. **(A)**: Summary of distinguishing features of six neutrophil clusters. **(B)**: Upset plot showing numbers of cluster markers that are either distinct to certain clusters or shared between multiple clusters. Only 'high abundance' markers (expressed in at least 60% of cells in a cluster) were considered. Markers that were shared with non-neutrophil clusters were excluded. **(C – E)**: UMAP projections showing normalised expression values of markers that were **(C)** shared by between upper and lower neutrophil clusters, **(D)** expressed principally in the upper neutrophil clusters, and **(E)** expressed principally in the lower neutrophil cluster. Gene names are as given in the ENSEMBL annotation (without altering case) with the exception of entries with no gene names, in which case they are named according to protein homology (e.g. FBP1-like). The clusters for which the gene is a high abundance marker (expressed in at least 60% of cells) are listed next to the gene ID and, where relevant, clusters for which the gene is a non-high abundance marker (expressed in < 60% of cells) are listed below. *Denotes gene is in a list of prior markers used to infer putative cell type or state.

Macrophages comprise a single principal cluster and smaller heterogeneous satellite clusters

We identified five clusters highly expressing prior markers of macrophages and / or monocytes, which we termed M1 to M5 (see **Fig. 2**). These comprised one large cluster (M1) and four small, fragmented clusters (**Fig. 5A**). Two prior markers of macrophages, *MPEG1* and *irf8* were present among cluster markers in addition to two copies of *CD74*, the homologue of which has been identified as a monocyte lineage marker in zebrafish (Athanasiadis et al., 2017). Among (high abundance) cluster markers, 25 were unique to the macrophage clusters (**Table S4**). The majority of these were unique to the largest cluster M1 (**Fig. 5B**). Few markers were shared between clusters indicating heterogeneity amongst macrophage populations. Four of the five clusters showed elevated expression of antigen-presenting cell markers **(Fig. 5E**) consistent with canonical macrophage and dendritic cell function.

We infer M1 to be the principal macrophage cluster, being the largest and most distinct (highest number of unique high abundance markers). Markers included CD206-like (macrophage mannose receptor 1 homologue) and FCGR1-like (**Fig. 5D**). In humans, CD206 is expressed in most macrophages and functions in phagocytosis (Azad et al., 2014), as does the mouse homologue of FCGR1 (Swanson et al., 1999). The smaller, fragmented clusters tended towards fewer expressed genes, lower overall expression and higher mitochondrial gene expression (**Fig. S2**), suggesting that they were comprised of or otherwise contained dead or dying cells. Of these smaller clusters, M3 was the most similar to M1, sharing four high-abundance markers: three paralogues of complement component 1q (*c1qa*, *c1qb*, and *c1qc*) and OLFM4-like (**Fig. 5C**). Similarly, GO term enrichment analysis showed markers of M1 and M3 to be enriched for 'immune response' and 'antigen processing and presentation' (**Table S7**). M3 had two unique markers: *FAPB7* (fatty acid binding protein 7) and *epdl1* (ependymin). Fatty acid metabolism is important in macrophages to produce inflammatory mediators (Batista-Gonzalez et al., 2020), while ependymins are a class of purportedly teleost-specific glycoproteins found in cerebrospinal fluid and involved in the central nervous system (McDougall et al., 2018). Also of note is that M3 contained cells expressing classical macrophage marker *MARCO*, although its expression was detected in only 55% of M3 cells. M4 possessed the unique marker *flt3*, the murine homologue of which is a major dendritic cell growth factor (Karsunky et al., 2003), thus implying dendritic cells (which, although thought to derive from a distinct cell lineage from monocytes, may be functionally similar; (Guilliams et al., 2014). Uniquely, M5 contained cells expressing classical macrophage marker *mrc1a* (<60% of cells), despite not showing elevated Antigen-Presenting Cell (APC) marker expression (**Fig. 5E**). Indeed, while some degree of specialisation of macrophage clusters is implied by the small number of unique high abundance markers, the relative scarcity of high abundance markers compared to genes expressed only partially within clusters (25 high-abundance markers compared to 75 non-high-abundance markers after excluding markers of other clusters) suggests substantial heterogeneity which was not delineated at this level of clustering.

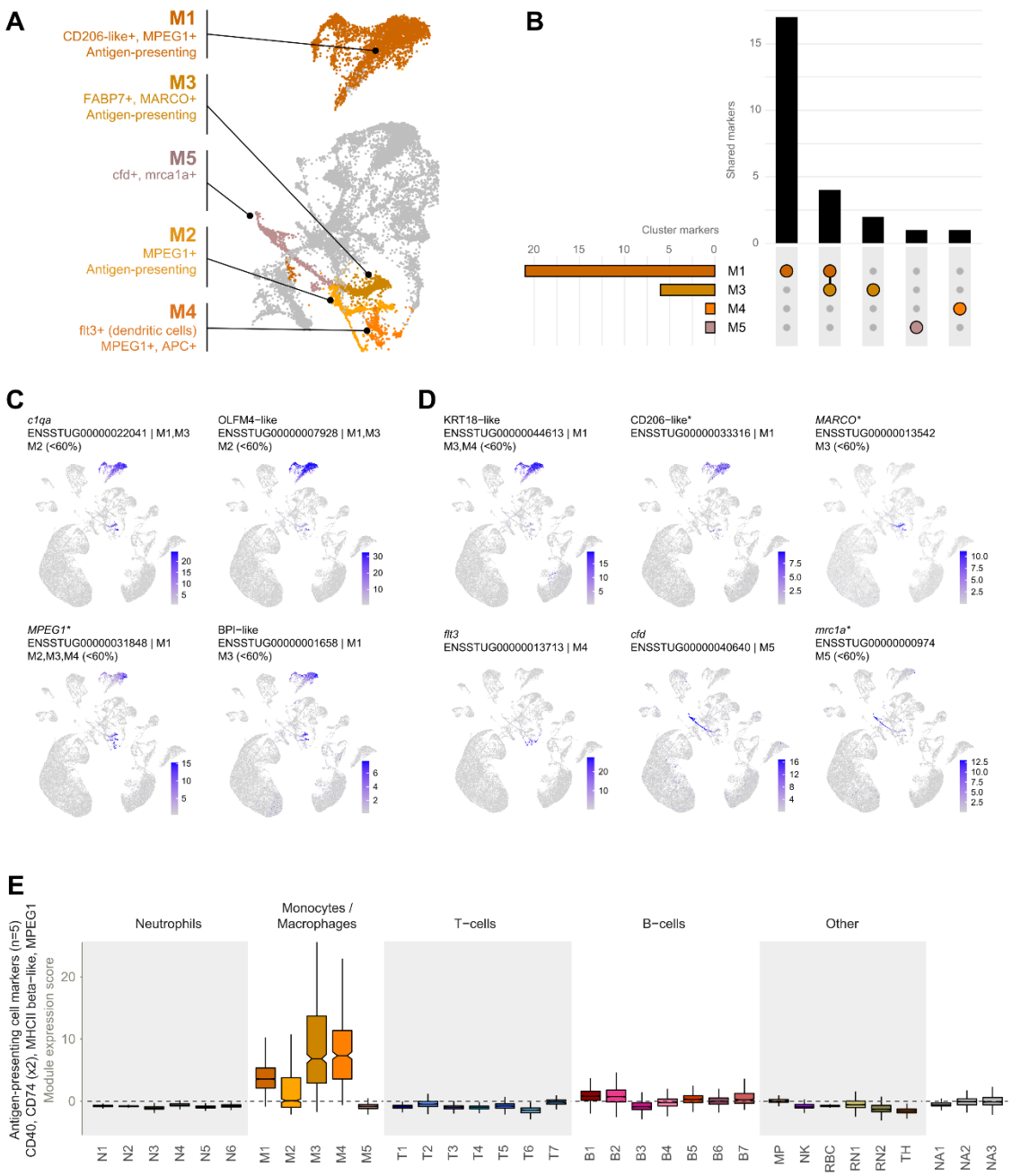

Figure 5. Monocyte / macrophage clusters and example expression patterns. **(A)**: Summary of distinguishing features of five monocyte / macrophage clusters. **(B)**: Upset plot showing numbers of cluster markers that are either distinct to certain clusters or shared between multiple clusters. Only 'high abundance' markers (expressed in at least 60% of cells in a cluster) were considered. Markers that were shared with non-macrophage clusters were excluded. **(C & D)**: UMAP projections showing normalised expression values of markers that were **(C)** shared by multiple clusters and **(D)** expressed principally in one cluster. **(E)**: Expression scores per-cell of a gene module comprised of three prior markers of antigen-presenting cells, across 34 clusters. Gene names are as given in the ENSEMBL annotation (without altering case) with the exception of entries with no gene names, in which case they are named according to protein homology (e.g. BPI-like). The clusters for which the gene is a high abundance marker (expressed in at least 60% of cells) are listed next to the gene ID and, where relevant, clusters for which the gene is a non-high abundance marker (expressed in < 60% of cells) are listed below. *Denotes gene is in a list of prior markers used to infer putative cell type or state.

Major T-cell subpopulations can be largely distinguished by canonical marker gene expression

Seven clusters were identified as T-cells based on prior markers and designated T1-7 (**Fig. 6A**). From these however we excluded cluster T7 which was represented almost exclusively by fish from a particular origin (wild-caught group, see **Fig. S5**), and therefore markers of cell type or state would not be distinguishable from markers of genetic, environmental, or genotype-by-environment effects. We therefore considered only clusters T1-T6 in our characterisation of T-cell subpopulations.

After excluding markers of non-T-cell clusters, 53 genes were identified as T-cell markers (**Fig. 6B; Table S5**). Of these, only two were shared between clusters T1-T6 (i.e. 'pan' T-cell markers, **Fig. 6C**): CD3E-like and TRBC1−like (TCR beta constant 1-like), both of which were also prior markers and homologues of genes encoding well-characterised T-cell surface receptors. A further five markers were shared by five clusters: STK10-like, *def6b*, *il7r* (interleukin 7 receptor), *sh2d1ab*, and *fynb*. STK10 is a putative immunoregulator expressed in human T-cells (but also in dentritic cells and NK cells; (Bi et al., 2022)). Human DEF6 is important for T-cell homeostasis (Serwas et al., 2019), while human SH2D1A is involved with SLAM-signaling (important for T-cell proliferation) and indeed has known interactions with FYN (Chan et al., 2003).

Turning to individual clusters, we found that major subpopulations (namely T1, T4, and T5) could be largely distinguished by expressed homologues of canonical T-cell markers, namely CD2, CD4, CD5, and two homologues of each of CD8 and CD28 (**Fig. 6D**; only one CD8 homologue shown). T1 uniquely expressed two homologues of CD8 in high abundance (*cd8a* and *cd8b*; latter not shown in Fig. 6B, but exhibited identical expression pattern to *cd8a* indicating its expression as a heterodimer), a surface protein characteristic of memory and cytotoxic T-cells (Macintyre et al., 2011), suggesting functional conservation of these two homologues. However, it expressed only one of two CD28 homologues (suggestive of some degree of functional divergence). T1 markers were enriched for the GO term 'integrin complex and indeed integrins are important for T-cell activation by promoting stable contact with antigen-presenting cells (Burbach et al., 2007). Meanwhile, T4 was distinguished by expression of both CD28 homologues along with *cd4-2b*, the latter indicative of T-helper cells ($CD4^+$ in humans). T5 was the most distinct (10 high abundance markers not shared by other clusters), characterised by expression of a homologue of canonical T-cell marker CD2 along with SLAM5-like (human homologue also known as CD84) and *il2rb*, the latter two indicative of regulatory T-cells (Chinen et al., 2016; Cuenca et al., 2019). T2 was heterogeneous with only one unique marker expressed in high abundance (likely reflecting a cluster composed of more than one distinct subpopulation which were not completely resolved) but expressed *plk3* in combination with *ilr2b*, the latter suggesting the cluster contained regulatory T-cells. T2 also uniquely contained cells expressing two unnamed homologues of TRGC1 (TCR gamma constant 1; **Fig. S6**), suggesting that it contained the less common gamma-delta T-cells and in line with prior genome sequence evidence that these cells exist in salmonids (Yazawa et al., 2008).

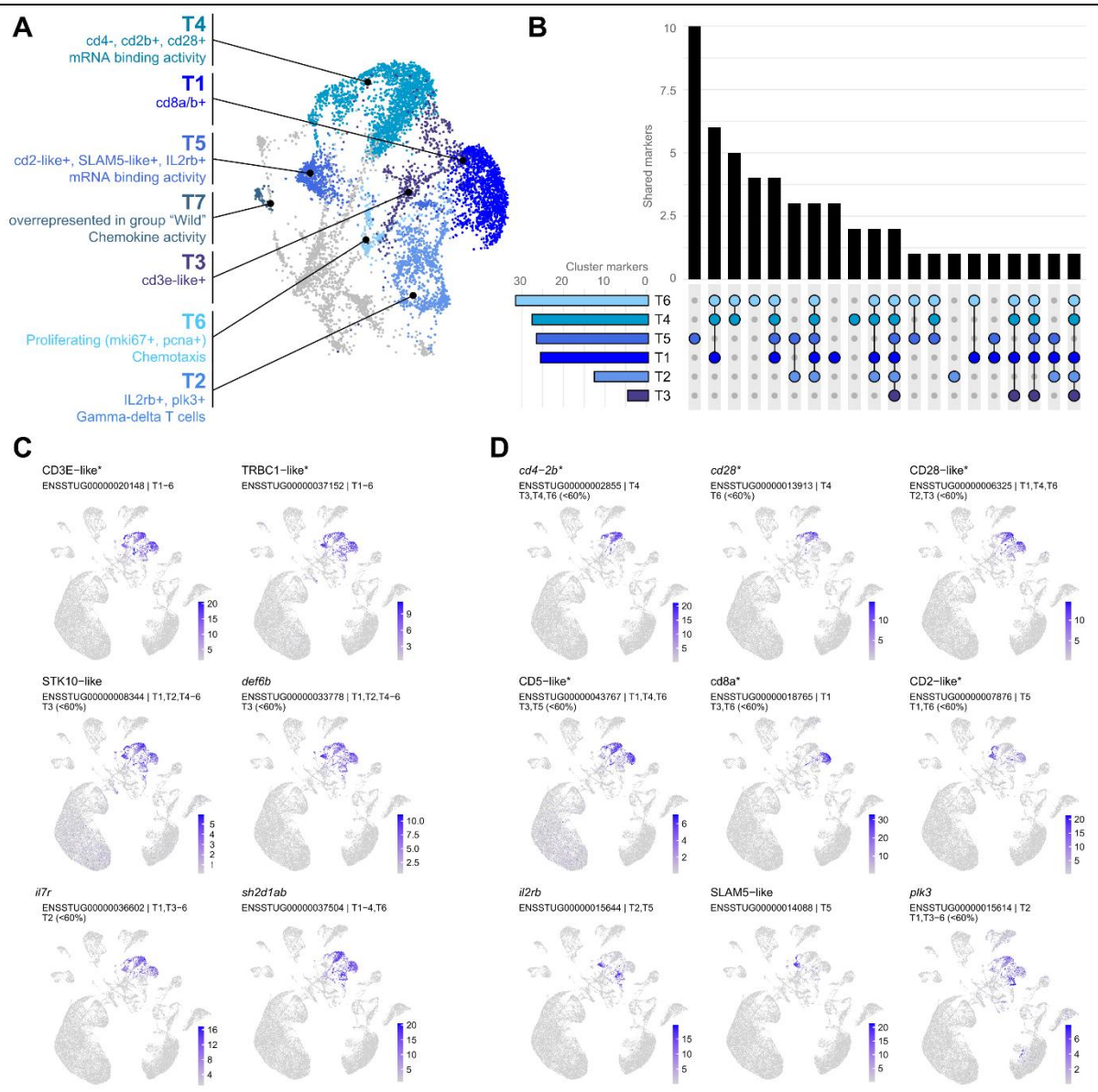

Figure 6. T-cell clusters and example expression patterns. **(A)**: Summary of distinguishing features of seven T-cell clusters. **(B)**: Upset plot showing numbers of cluster markers that are either distinct to certain clusters or shared between multiple clusters. Only 'high abundance' markers (expressed in at least 60% of cells in a cluster) were considered. Markers that were shared with non-macrophage clusters were excluded. **(C & D)**: UMAP projections showing normalised expression values of markers that were **(C)** shared by multiple clusters and **(D)** expressed principally in one cluster. Gene names are as given in the ENSEMBL annotation (without altering case) with the exception of entries with no gene names, in which case they are named according to protein homology (e.g. CD3E-like). The clusters for which the gene is a high abundance marker (expressed in at least 60% of cells) are listed next to the gene ID and, where relevant, clusters for which the gene is a non-high abundance marker (expressed in < 60% of cells) are listed below. *Denotes gene is in a list of prior markers used to infer putative cell type or state.

T6 markers were consistent with proliferating cells (**Fig. 3**) and, notably, while T3 did not express any unique high abundance markers and showed signatures of dead or dying cells (high mitochondrial gene expression, few expressed genes and low overall expression; **Fig. S2**).

B-cell subpopulations are highly diverse but robustly characterised by CD79A homologue

B-cells were divided into seven clusters that formed clear separation on the UMAP (**Fig. 7A**). After excluding markers of non-B-cell clusters, 113 genes were identified as B-cell markers (**Fig. 7B; Table S6**). Several shared high-abundance markers were identified and all seven clusters possessed at least one unique marker (**Fig. 7B**). Although no high abundance markers were shared by all seven clusters, several were shared by at least five, notably *cd79a* (**Fig. 7C**). We note that while *cd79b* showed similar expression patterns, it was reduced in the larger clusters B1 and B2. Other 'pan' B-cell markers included genes whose mammalian homologues have been demonstrated to be involved in B-cell proliferation or differentiation including *mef2cb* (Wilker et al., 2008), *hipk2* (Guerra et al., 2012), and one homologue of EBP41 (Liang et al., 2020), as well as canonical B-cell marker homologue CD37-like and at least six immunoglobulin-like genes (e.g. first two panels of **Fig. 7E**). Indeed, B-cell markers included several genes homologous to Ig proteins, some lacking annotation but containing predicted IG domains (see **Table S6**). These tended to show variable expression between and within clusters (**Fig. 7E**) and included an IGLL1-like gene which was uniquely overexpressed in B3.

B1 and B2 tended to show similar expression patterns, as when we examined high-abundance markers of B1 these tended to be also highly expressed in B2 (though not necessarily in high abundance), and *vice versa* (**Fig. 7D**). Both showed only weak expression of *cd79b*, and we note that in zebrafish reduced CD79 expression occurs upon differentiation to the Ig-secreting stage (Liu et al., 2017). Most specific to B1 was *zgc:194275*, homologous to a zebrafish B-cell marker (C. Hu et al., 2024), also known as PLAAT1, which has shown to be involved in regulation of antiviral responses in zebrafish (Zhao et al., 2022). Meanwhile, *arhgap24*, whose human homologue is enriched in naïve and memory B-cells (Human Protein Atlas version 24; https://www.proteinatlas.org/ENSG00000138639-ARHGAP24/single+cell (Uhlen et al., 2019)) and has been shown to inhibit cell proliferation (L. Wang et al., 2019), showed mirrored expression in B1 and B2. Meanwhile, while one homologue of EBP41 had emerged as a pan-B-cell marker (albeit weakly expressed in some clusters), a second homologue of this gene was expressed predominantly in B2, implying functional divergence. Also unique to B2 was a previously unannotated gene (mikado.19G1247, not shown), whose top BLAST hit was to NHSL2, involved in cell migration – a process critical for patrolling immune cells (Y. Wang et al., 2023). Ig-like genes were also highly expressed in B1 and B2 (**Fig. 7E**), and pseudotime analysis of B-cell clusters with B3 assigned as the starting point (**Fig. S4B**) supported the designation of B1 as terminally differentiated B-cells. B1 and B2 markers were enriched for the GO term 'actin binding'; indeed, reassembly of the B-cell actin cytoskeleton allows expansion of the cell surface to increase contact with antigen-presenting cells (Song et al., 2013).

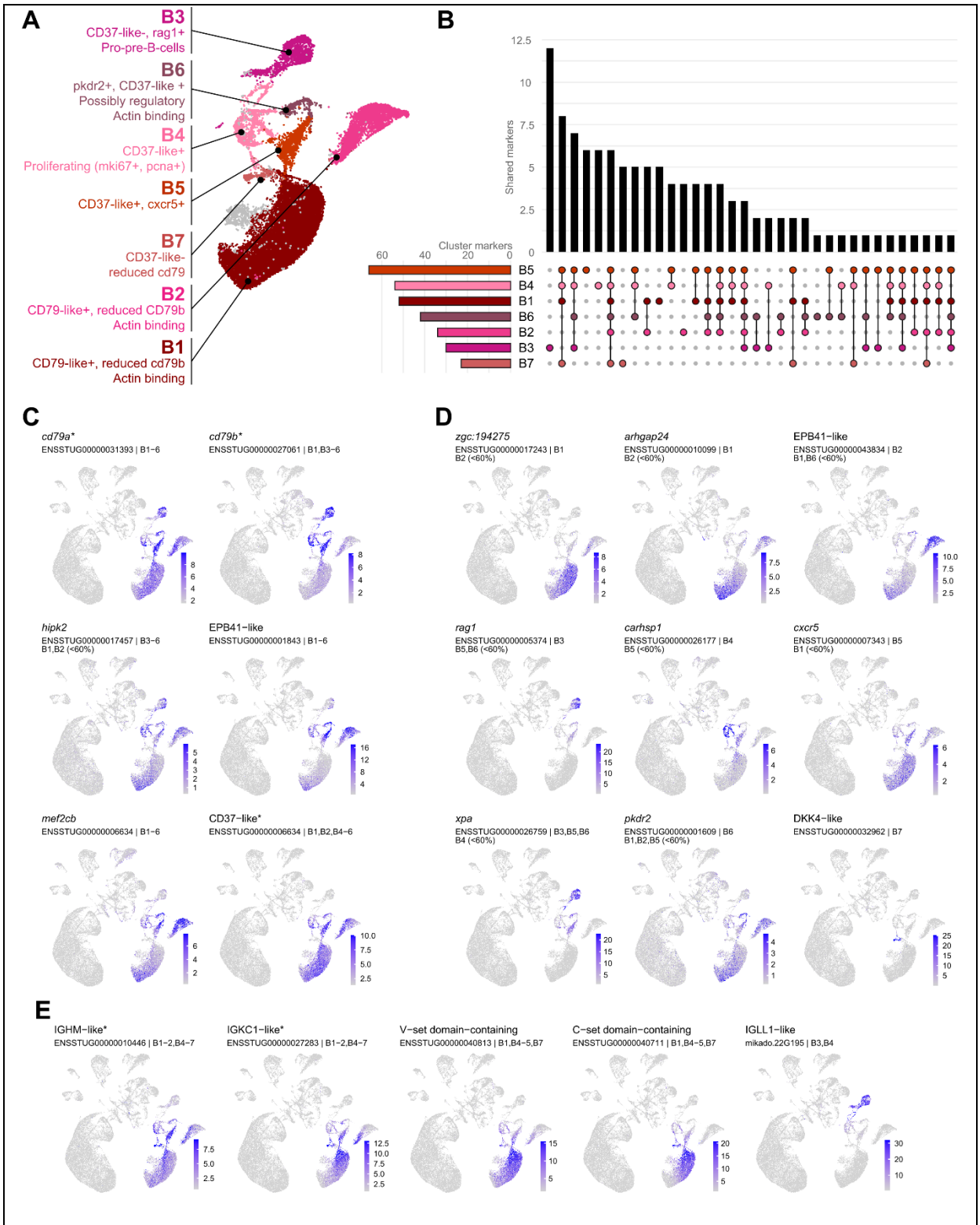

Figure 7. B-cell clusters and example expression patterns. **(A)**: Summary of distinguishing features of seven B-cell clusters. **(B)**: Upset plot showing numbers of cluster markers that are either distinct to certain clusters or shared between multiple clusters. Only 'high abundance' markers (expressed in at least 60% of cells in a cluster) were considered. Markers that were shared with non-macrophage clusters were excluded. **(C – E)**: UMAP projections showing normalised expression values of markers that **(C)** were shared by multiple clusters, **(D)** were expressed principally in one cluster, and **(E)** are homologues of immunoglobulins or which otherwise contain Ig domains. Gene names are as given in the ENSEMBL annotation (without altering case) with the exception of entries with

no gene names or genes not found in the ENSEMBL annotation, in which case they are named according to protein homology (e.g. CD37-like). The clusters for which the gene is a high abundance marker (expressed in at least 60% of cells) are listed next to the gene ID and, where relevant, clusters for which the gene is a non-high abundance marker (expressed in < 60% of cells) are listed below. *Denotes gene is in a list of prior markers used to infer putative cell type or state.

The most distinct cluster was B3, having the highest count of unique markers. Most notably, B3 showed high expression of *rag1* (recombination activating gene 1), a gene well known for its crucial role in B-cell development (Kuo & Schlissel, 2009), see also (Novoa et al., 2019). Other markers of B3 whose homologues have been shown to have roles in B-cell or pre-B-cell development includes *kap12* (Kartal-Kaess et al., 2012), *itga5* (Y. Hu et al., 2016), and *tek* (Reth & Nielsen, 2014), thus collectively supporting a pre- or pro-B-cell identity. B4 was a highly proliferative cluster as evidenced by expression of proliferation markers (see **Fig. 3**), and both B3 and B4 markers were enriched for the GO term 'structural constituent of chromatin' suggestive of chromatin remodeling and differentiation. B5 was notably marked by *cxcr5* – associated with B-cell migration (Förster et al., 1996), and with B helper T cells (Breitfeld et al., 2000). B5 and B6 shared *xpa* (associated with DNA repair; (Kim et al., 2023) as a marker with the pro/pre-B-cell cluster B3, suggestive of an intermediate stage, while B6 was marked by expression of *pkdr2*, associated with regulatory B-cells (Michaud et al., 2022). B6 markers were enriched for actin binding as were B1 and B2, suggesting similar interactions with antigen-presenting cells. Finally, B7 was a small cluster of seemingly highly differentiated cells as indicated by lack of expression of both CD79 genes (Liu et al., 2017) and intense expression of Ig-like genes (see **Fig. 7E**). Unique high-abundance markers of B7 included DKK4-like, a similar gene to which in mammals, DKK3, is a modulator of B-cell fate and function (Ludwig et al., 2015).

Other distinct cell types are resolved by marker genes

Natural killer-like cells (cluster NK) were identified by prior markers such as *prf1* (perforin) and possessed four unique cluster markers which also included a previously unannotated homologue of NK-lysin (**Fig. 8A**). A small but highly distinct cluster (41 unique markers) appeared to correspond with thrombocytes (cluster TH) according to expression of prior markers such as *mpl* (thrombopoietin receptor-like), as well as *thbs1b* (thrombospondin 1-like) (**Fig 8B**), both of which have mammalian homologues involved in platelet formation and / or aggregation (Hitchcock et al., 2021; Leung, 1984). A large cluster was identified as red blood cells (cluster RBC) based principally on expression of hemoglobin homologues such as *hbaa2* (**Fig. 8C**). Hemoglobin homologues were also however sporadically, but highly expressed in the central unclassified cluster NA1, suggesting it contained erythroid progenitors. Eleven genes were markers uniquely of the RBC cluster, including *alas2* (5-aminolevulinate synthase, erythroid-specific, mitochondrial-like), thus comprising markers of mature RBCs. Another small cluster was identified as putatively comprising myeloid progenitor cells (cluster MP), whose unique markers included *kita* and *egr1* (**Fig. 8D**), homologues of which are heavily

implicated in early myeloid cell differentiation (Ashman et al., 1999; Kulkarni, 2022). Among other markers of MP were three *fos* genes and a homologue of transcription factor JUND which, along with EGR1 are all implicated in early myeloid function (Yuan et al., 2007). Finally, two clusters with similar expression patterns were putatively identified as renal (kidney) cells (clusters RN1 and RN2), markers of which included *ly97.3* (CD59-like; **Fig. 8E**), *si:dkey-194e6.1* (collectrin-like) and *gstr* (glutathione S-transferase A-like), mammalian homologues of which are highly expressed in cells of the kidney (Harrison et al., 1989; Nangaku et al., 1998; Zhang et al., 2001). RN1 and RN2 were also the only clusters showing appreciable expression of *gapdh* (**Fig. 8E**).

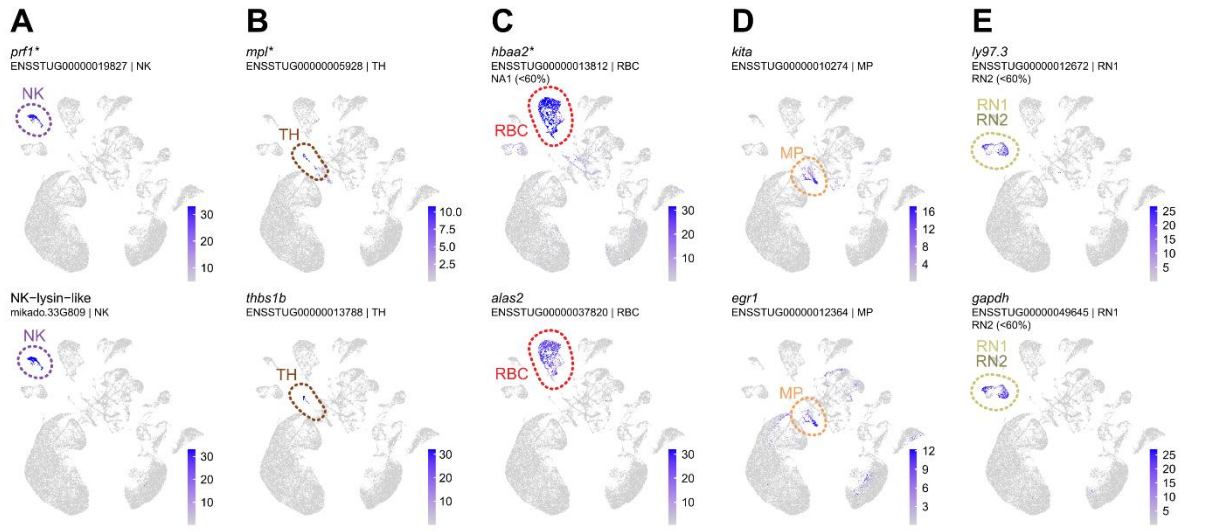

Figure 8. UMAP projections showing normalised expression values of markers corresponding with putative **(A)** Natural-killer cells (NK), **(B)** thrombocytes (TH), **(C)** red blood cells (RBC), **(D)** myeloid progenitor cells (MP), and **(E)** renal cells (RN). Relevant clusters are circled in each subplot. Gene names are as given in the ENSEMBL annotation (without altering case) with the exception of one gene which was not contained in the ENSEMBL annotation and which is named according to protein homology (NK-lysin-like). The clusters for which the gene is a high abundance marker (expressed in at least 60% of cells) are listed next to the gene ID and, where relevant, clusters for which the gene is a non-high abundance marker (expressed in < 60% of cells) are listed below. *Denotes gene is in a list of prior markers used to infer putative cell type or state.

Some gene pairs arising from SSR4 whole genome duplication show signs of functional divergence in the immune system

Examples of putative functional divergence amongst homologues were identified in the course of characterising clusters based on marker expression, such as homologues of CD28 (**Fig. 6D**) and EBP41 (**Fig. 7D**). We therefore investigated the degree to which duplicate gene pairs arising from the salmonid-specific 'SSR4' whole genome duplication event (hereafter 'ohnologues') exhibit signs of functional divergence such as subfunctionalisation (i.e. division of labour) or neofunctionalisation (i.e. one copy gains a new function) as opposed to both ohnologues showing the same expression patterns. To this end,

the 2000 variably expressed genes were queried for homologues in the northern pike (*E. Lucius*), representing a sister lineage to the salmonids that did not undergo the SSR4 WGD. For genes with pike homologues, we filtered these to retain only those with a 2:1 ratio (trout:pike) and for which each member of the pair resided on a different trout chromosome. This resulted in a list of 175 putative ohnolog pairs amongst the variably expressed genes.

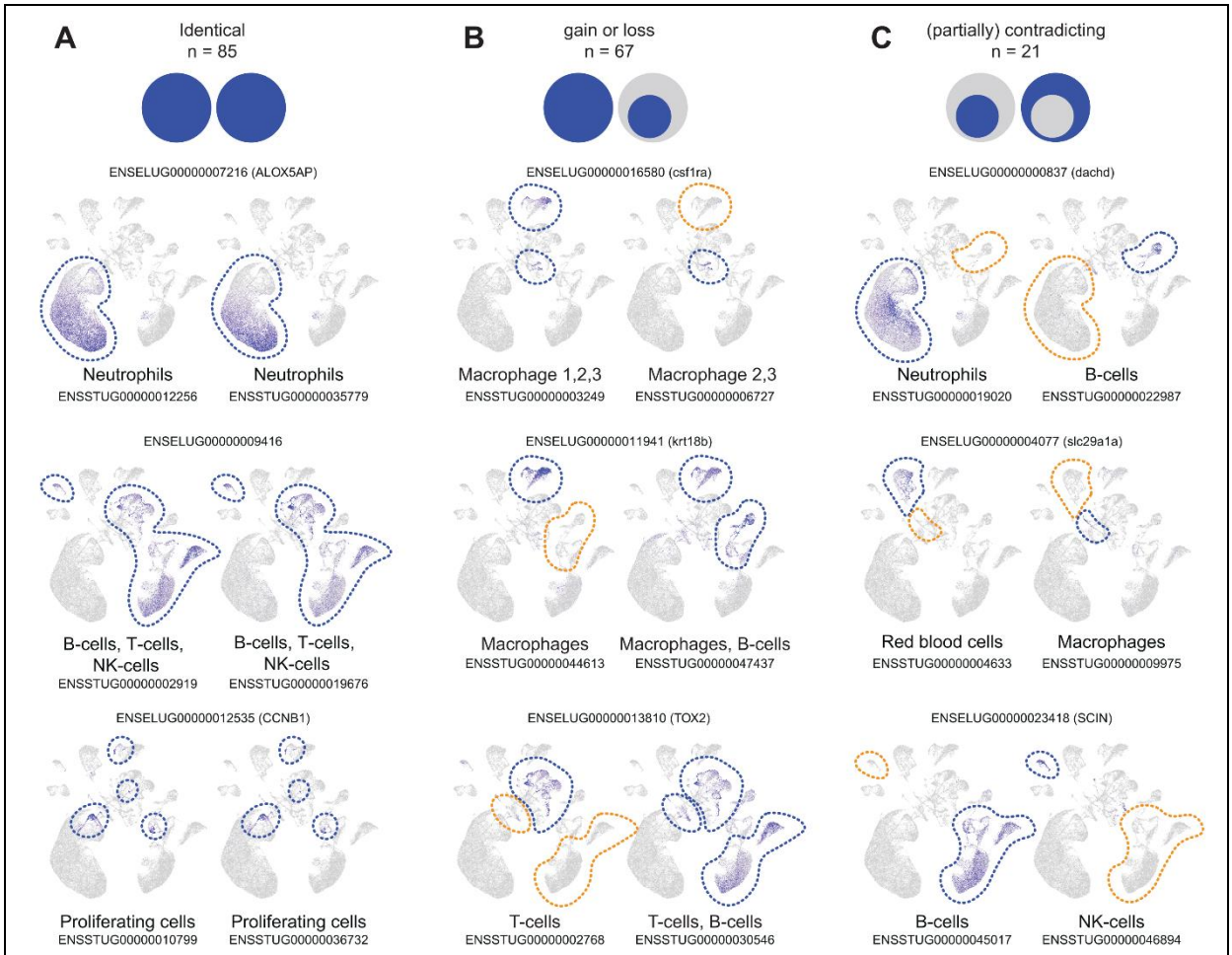

Figure 9. Examples of expression patterns of ohnologue pairs (putative SSR4-derived gene duplicates). **(A)**: Ohnologue pairs with identical expression patterns across clusters ('identical'); **(B)**: ohnologue pairs with similar expression across clusters except that one ohnologue is missing from certain clusters ('gain or loss'); **(C)**: ohnologue pairs with fully asymmetric expression patterns ('contradicting'). For each ohnologue pair, the ENSEMBL gene ID (and gene name where available) of the *E. lucius* homologue is given atop the UMAP projections. On UMAP projections, clusters with retained (or gained) expression relative to the other ohnologue are circled in blue, while clusters with missing expression relative to the other ohnologue are circled in orange.

We then examined the expression patterns of these pairs and assigned 173 of them to one of our categories (**Table S8**): (1) 'identical': both ohnologues show identical or near-identical expression patterns across clusters, suggesting that both have retained the same function (n=85, see examples in **Fig. 9A**); (2) 'gain or loss': both ohnologues have similar expression across clusters except that one ohnologue is missing from certain clusters (n=67, see examples in **Fig. 9B**), which may suggest either

subfunctionalisation or a transition to neofunctionalisation; (3) 'contradicting': each ohnologue is expressed in a different cluster or set of clusters, suggesting neofunctionalisation (n=15, see examples in **Fig. 9C**); or (4) 'partially contradicting': each ohnologue is expressed in a combination of the same and different clusters, suggesting the beginning of neofunctionalisation (n=6). Striking examples of putative neofunctionalisation included *dachd* of which one ohnologue was expressed widely in neutrophils and the other expressed in pre-B-cells. Evidence from *Drosophila* and mammals suggest that the Dach transcription factor (*dachd* homologue) is a master developmental regulator, forming complexes with chromatin to induce differentiation (Davis & Rebay, 2016). Meanwhile, *SCIN*, the mammalian homologue of which is an actin-severing protein (Ghoshdastider et al., 2013), showed one ohnologue expressed widely in B-cells and the other in NK-cells (**Fig. 9C**). Plots of all examined ohnolog pairs can be found in **Supplementary File 1**.

Environment of current or prior generations affects immune cell transcriptomes

Not only did our dataset comprise single-cell expression profiles from nine separate *S. trutta* individuals, but these individuals came from three distinct backgrounds which may affect transcriptional diversity via genetic and / or epigenetic differences. Briefly, 'wild' fish (N =3) were captured directly from a local stream in the canton of St. Gallen, Switzerland, 'farm' fish (N =3) were derived from at least two generations of rearing in an aquaculture facility, while 'mix' fish (N =3) were hatched in the aquaculture facility but derived from initially wild-caught parents. By comparing cluster-level expression profiles of wild fish (deemed as a control) with those of farm and mix fish, we sought to characterise the influence of multigenerational and single-generation hatchery rearing, respectively. Selected a cluster from each of the four main lineages (N1, M1, T5 and B3 for neutrophils, monocytes / macrophages, T-cells, and B-cells, respectively), we performed between-group differential expression analyses using the pseudobulk approach (Murphy et al., 2022).

Of the selected clusters, M1 showed the most abundant differences in gene expression between groups (235 DEGs at FDR < 0.05, of which 157 were DE in mix and 78 in farm; **Table S9**), followed by N1 (178 DEGs at FDR < 0.05, of which 95 were DE in mix and 83 in farm; **Table S10**) and B3 (121 DEGs at FDR < 0.05, of which 65 were DE in mix and 56 in farm; **Table S11**), and PCA plots of all three showed clear separation between wild and mix groups along PC1 and / or PC2 (**Fig. 10**). T5 meanwhile showed more modest differences (74 DEGs at FDR < 0.05, of which 35 were DE in mix and 39 in farm; **Table S12**) and less clear sample separation on PCA (**Fig. S7**).

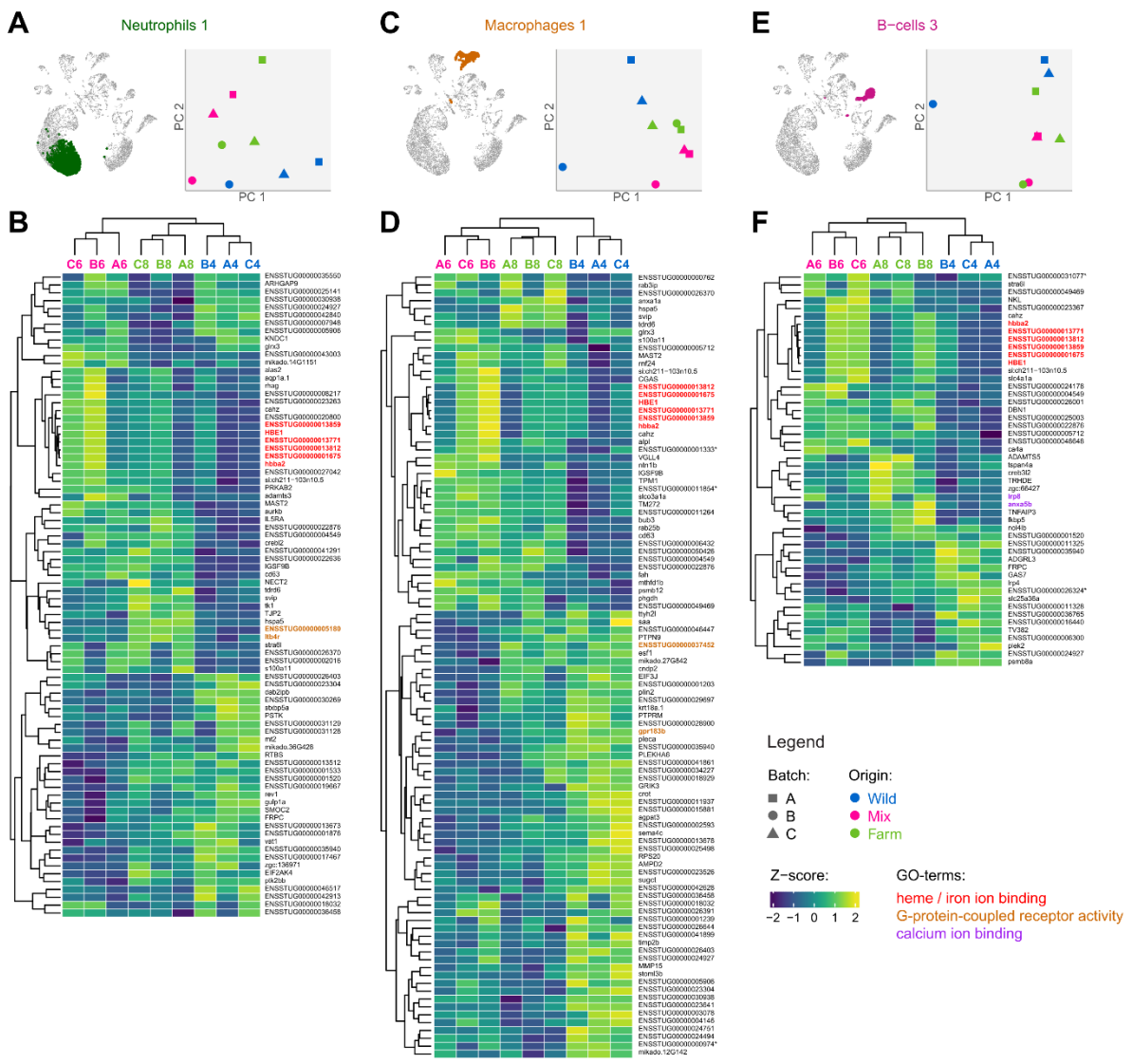

Figure 10. Pseudobulk differential gene expression profiles between fish of different origins, shown for three cell clusters. **(A, C, E)**: Clusters under interrogation are highlighted on UMAP plots for each of clusters N1, M1, and B3 alongside PCA plots derived from the top 500 most variably expressed genes (between individual pseudobulk samples) in each comparison. On PCA plots, point colour denotes individual origin (wild, farm, or mix), while point shape denotes the sampling batch. **(B, D, F)**: Heatmaps of genes found to be differentially expressed in either the 'farm' or 'mix' group compared to the 'wild' group. Dendrograms represent hierarchical clustering of genes and samples according to the Z-scores of the Log2(CPM) values (cell colour intensity). Sample IDs (top of each heatmap) are coloured according to the group from which the individual originated ('Origin'). Gene names are given where available from the ENSEMBL database, otherwise the ENSEMBL gene ID is shown. Gene names or IDs associated with particular enriched GO terms amongst DEGs are highlighted in bold and coloured according to GO term (red = 'heme binding' or 'iron ion transport', orange = 'G protein-coupled receptor activity', purple = 'calcium ion binding'). Note that only genes DE at FDR < 0.01 are shown, so not all genes comprising enriched GO terms are shown.

To infer potential functional implications of DEGs, we ran GO term enrichment analysis using the GO domain 'molecular function', which we considered more appropriate than biological processes or

cellular component given that the cell types had already been inferred. Surprisingly, despite as shown earlier, these clusters being highly distinct from each other in their overall transcriptional profiles, similar GO terms emerged amongst genes differentially expressed between groups in N1, M1, and B3 (**Table S13**). Specifically, the related functions 'heme binding' and 'iron ion binding' were enriched amongst genes upregulated in the mix group in all three aforementioned clusters due to differential expression of hemoglobin homologues. Meanwhile, 'G protein-coupled receptor activity' was enriched amongst genes upregulated in the farm group in N1 and B3, and amongst genes downregulated in the mix group in M1. In B3, 'calcium ion binding' was additionally enriched amongst genes upregulated in the farm group. No enriched GO terms were detected for genes differentially expressed in T5.

Discussion

Single cell gene expression profiling is already unlocking unprecedented insight into myriad biological processes in across animals and plants, including the genetic regulation of the immune system. Data from non-model organisms, in this case salmonids, can unlock deeper insights into the evolution of the immune system as well as insights of great value to, for example aquaculture. Indeed, commercial fish rearing environments may have profound effects on fish health and immunity, and therefore delineating the underlying transcriptional mechanisms may be of great benefit. A unique aspect of our study is our incorporation of inter-individual variation: we sampled nine fish from three different backgrounds and performed a single integrated scRNAseq analysis on the combined set of cell expression profiles to discern cell clusters and identify putative cell types. While undoubtedly introducing additional heterogeneity into the data due to potential genetic differences and environmental effects, this approach of combining cells from multiple individuals in silico meant that the putative cell types identified from the expression data were representative of fish from different backgrounds. Indeed, almost all clusters contained representative cells from all individuals and groups (**Fig. S5**), with the exception of one T-cell cluster (T7) which was represented almost exclusively by fish of the 'wild' group. Given that these individuals derived from rivers known previously to harbour fish pathogens (namely *Tetracapsuloides bryosalmonae*), the appearance of this origin-specific cell cluster may reflect the prior history of the population's exposure.

The expression patterns of immune cell markers derived from other species revealed that groups of clusters could be readily divided according to expression modules derived from these markers, suggesting that the roles of much of the immune gene repertoire are largely conserved. Indeed, our findings are broadly comparable in this regard to similar scRNAseq studies of other salmonid immune repertoires (Guslund et al., 2020; Perdiguero et al., 2024) . However, due to likely extensive functional divergence between salmonids and model organisms from which prior markers are derived, identifying cell types at finer scale is highly challenging. We therefore applied numerous methods to infer cell types: expression of prior markers and overlap of prior markers with cluster markers, GO-term enrichment, pseudotime analysis to infer terminally differentiated cells, and extensive literature searches. Indeed, we

found that several well-established markers were robust for broad cell lineages, for example the CD79 genes appeared to be reasonably robust markers of the B-cell lineage. Functional divergence however also means that many genes whose expression distinguishes salmonid immune lineages are likely to have no previously reported function in model species. Subsequently, we found that both broad immune cell lineages and, within these, sub-lineages, could be distinguished according to expression not only of prior markers, but of numerous genes with no reported immune function from model species. These included the expression of high choriolytic enzyme (HCE) – an enzyme otherwise involved in digestion of the chorion during hatching – marking non-terminally-differentiated neutrophils, while *epdl1* (ependymin) was a marker of a specific putative macrophage cluster (M3) despite its reported principal function in the teleost central nervous system (McDougall et al., 2018). Overall, our data not only offer robust markers for broad cell lineages, but also a range of markers or marker combinations for distinct sub-lineages which could be of great value in a diagnostic setting.

Given putative cluster identities it is possible to further interrogate the evolution of gene function across specific cell types. Given that gene duplicates arising from the SSR4 WGD event are understudied in relation to the immune system, we screened for patterns of putative functional divergence between duplicated gene pairs (ohnologues). Around half of the screened ohnologue pairs (85 out of 173) showed the same expression patterns, with the remaining pairs showing some degree of divergent expression patterns. The fate of SSR4-derived duplicates has been debated, with Sandve et al (2018) more recently concluding that the prevailing fate of ohnologue pairs is for one to be under stronger purifying selection than the other (Sandve et al., 2018). While relaxed selection on one copy may lead to neofunctionalisation, this may also lead to eventual pseudogenisation (indeed, this has been posited as the terminal fate of duplicated genes (Rastogi & Liberles, 2005), and asymmetric expression patterns could reflect a transition to either state. Among ohnologue pairs showing divergent expression patterns, the 'gain or loss' pattern was the most common (n = 67). While this would be consistent with subfunctionalisation, it is also feasible that such patterns reflect relaxed purifying selection in one copy which may, in millions of years, lead to the emergence of new function. We concede however that our analysis did not seek to rigorously test for functional divergence as we did not attempt to correlate expression levels across cell types with those of an outgroup species, nor did we test whether ohnologue pairs are under differential selection at the sequence level. Furthermore, given that our analysis was on the gene expression, rather than transcript expression level, we could not distinguish divergence of ohnologue function from differential isoform usage. However, our analysis revealed examples of putative functional divergence which suggest a more extensive study utilising data from multiple species would yield deeper insights into the fate and role of duplicated genes in the salmonid immune system. Indeed, the increasing availability of single-cell expression data from multiple non-model organisms will likely allow for studies of gene functional divergence in both unprecedented detail and evolutionary scale.

Finally, our profiling of individuals of three different origins allowed us not only to incorporate inter-individual variation but also explore variation resulting from individual history. Taking four distinct immune cell clusters as examples, we found evidence of substantial altered transcriptional regulation in fish derived reared for at least two generations in a farm (farm) and those which were reared for a single generation on a farm (mix) compared to those that were derived from natural streams (wild). In both contexts we identified genes involved in G protein-coupled receptor activity – a function closely linked with immune modulation, for example interleukin production (Saroz et al., 2019) and TLR signaling (Sharma et al., 2013) – to be differentially expressed in two of the four tested clusters. Meanwhile, in the mix context we found hemoglobin gene homologues to be upregulated in three of the four tested clusters. While surprising, the production of hemoglobin by nonerythroid cells is not unheard of (Ghosh et al., 2018) and there is now evidence for a role of heme in a variety of cellular processes, including potentially immune responses via its interaction with NO (Dunaway et al., 2024). Another plausible explanation is the contamination of target immune cell libraries by RNA from red blood cells in the processed material which, although enriched for immune cells nevertheless contained abundant RBCs. Indeed, the expression of hemoglobin homologues in neutrophils, macrophages and B-cells was highest in the 'mix' sample B2 (**Fig. 10**) which also had the highest proportion of RBCs (around 16%, see **Fig. S5**). Further investigation should focus on the extent to which transcriptional differences arise from environmental influences in the early growth environment or via natural selection on genetic variation which purges certain genotypes from the farm rearing environment. Indeed, the recurrence of similar gene sets differentially expressed across multiple cell types might imply the existence of fixed polymorphisms at cis- or trans-regulatory elements that affect the expression of a few genes in a systemic fashion. In any case, proliferation of genotypes with altered transcriptional regulation of the immune system may pose a potential risk to natural populations.

Conclusions

In conclusion, our data, while reflecting broad functional conservation of the immune system transcriptome in salmonids, also reveal numerous idiosyncrasies and evidence of functional divergence following whole genome duplication. Furthermore, reflecting the importance of studying both wild and farmed fish (from both ecological / evolutionary and aquacultural perspectives), our data captures diversity from both contexts, and subsequently reveals putative influences of rearing background on immune system function at the transcriptional level.

Materials and methods

Study design, sampling

Brown trout from different rearing histories and environments were used: wild group (wild), mix group (mix) and farm group (farm). The wild group consisted of young-of-the-year (≤ 1 year old) brown trout (F1 generation) that were fished during spring 2021 in two Swiss rivers (Brübach, coordinates between

47.471195 °N/ 9.136465 °E and 47.473303 °N/ 9.140526 °E; and Rörlibadbach between 47.473019 °N/ 9.136563 °E and 47.478731 °N/ 9.134755 °E. The parents (F0 generation) were born and raised in the wild. These waterbodies are known to be positive for Proliferative Kidney Disease, caused by a parasite *Tetracapsuloides bryosalmonae*, for generations (own observations). Although all animals were tested negative for T. bryosalmonae according to Bettge et al (Bettge et al., 2009), we cannot exclude that the F0 generation has faced an early infection or contact with *T. bryosalmonae*. The mix group (mix) originated from wild parents (F0 generation), female and male adult brown trout, which were fished in the same river system (Brübach) and spawned in autumn 2020. Fertilized eggs were transferred to a farm facility where they (F1 generation) were raised until spring 2021. The farm group (farm) originated from a Swiss farm, where two generations of brown trout were raised under farm conditions. Both mix and farm F1 generations were raised in the same hatchery facility (*Fischereizentrum Steinach*, Switzerland) from the stage of fertilized eggs until approximately three months of age-old fry. All animals used had the same genetic origin.

All F0 generations were transferred to the Institute for Fish and Wildlife Health (FIWI), University of Bern, Switzerland in May 2021, when fish were approximately 3 months of age. Each group of fish (wild, mix, farm) was distributed into six aquaria with equal fish numbers (n = 27 per aquaria). Each group consisted of three biological replicates. At arrival, two fish per aquarium were subjected to a health check, no pathogens were detected.

Fish were maintained at constant 16 °C with a flow-through water system and 12:12h light:dark photoperiod. The fish were fed a commercial diet twice a day with 2.5% of their body weight and additionally with live artemia in the first month after arrival. The fish were acclimatized to laboratory conditions for four months prior to sampling. Experiment was approved by the cantonal veterinary office (Bern, Switzerland) (Authorization number BE25/2021).

Cell enrichment, library preparation and sequencing

Immune cells were isolated from the brown trout head and trunk kidneys when they reached the age of approximately 8 months. Both head and trunk kidneys are centres of hematopoietic cell production, and therefore material from both organs were pooled for each biological replicate. Fish were euthanized with an immersion bath in overdosed MS222 (150µg/l tricaine methanesulfonate; MS 222®, Pharmaq). The kidneys were then dissected and processed for RNA sequencing as follows:

The kidney tissue was passed through a 105µm pore size mesh filter with Leibovitz's (L-15) medium (Thermofischer, Reinach, Switzerland) containing 10% fetal calf serum (FCS). The resulting cell suspension was then layered onto a sterile, isotonic Ficoll gradient (Ficoll-paque Plus, GE Healthcare Bio-Sciences AB, Sweden (with a density of 1.077 g/ml and spun at 750 x g for 40 minutes at 4°C to dissociate red blood cells from white blood cells. Leukocytes at the Ficoll/ medium interphase were

aspirated, washed in L-15 medium and centrifuged at 290 x g at 4°C for 10 minutes. The cell suspension was aspirated and resuspended in PBS and kept on ice.

Cell suspensions were checked for viability and concentration using an automated cell counter (DeNovix® CellDrop Automated Cell Counter, NC, USA). The samples were diluted to a concentration ranging between 700 and 1900 cells/µL, depending on the sample. The cell suspensions were sent to the Bern Genomics Facility for 10x Chromium sequencing library preparation. All samples were prepared to ensure a target cell recovery of 10,000 cells and were subjected to cell isolation on the 10x Genomics Chromium Controller instrument. To minimize cell death, all steps, including PBL isolation, sorting, and Chromium™ Single Cell isolation, were completed on the same morning.

A total of 2500 cells per sample were loaded into the chips of the Chromium™ Single Cell 3' Gel Beads Kit (Single Cell 3' v3.1 Dual Indexing) (10x Genomics) and processed on the Chromium Controller instrument to generate singe cell Gel Bead-In Emulsions (GEMs) following the manufacturer's instructions.

Next, GEMs were subjected to library construction using the Chromium™ Single Cell 3' Library Kit v3.1 (10x Genomics). In the first step, reverse transcription was performed, generating cDNA tagged with a cell-specific barcode and a unique molecular index (UMI) per transcript. Fragments were size-selected using SPRIselect magnetic beads (Beckman Coulter, Brea, CA, USA). Illumina sequencing adapters were then ligated to the size-selected fragments, and cleaned up using SPRIselect magnetic beads. The quality of the final library was assessed using an Agilent 2100 Bioanalyzer (Agilent technologies, Amstelveen, The Netherlands). The samples were subsequently sequenced using a NextSeq 550 instrument (Illumina, CA, US) with 150PE chemistry.

Augmented gene annotation

To maximise the amount of usable scRNAseq reads, which depends on the proportion of reads that can be confidently aligned to the transcriptome, we augmented the existing fSalTru1.1.104 ENSEMBL annotation using two published bulk RNAseq datasets derived from *S. trutta* kidney tissue (Ahmad et al., 2021; Sudhagar et al., 2019). Transcripts were assembled using a combination of de novo and genome guided approaches and further processed using the Mikado pipeline (Venturini et al., 2018) to identify a nonredundant set of transcripts not represented in the prior annotation. This approach enabled us to add 15,095 additional transcripts and 1,200 additional gene models. For more details please see the supplementary methods. The additional transcript and gene models ('Mikado-derived genes') were appended to a modified version of the ENSEMBL gtf (fSalTru1.1.104) which had been filtered to retain only protein-coding genes, lncRNAs, and the IG- and TR- genes that undergo somatic recombination before transcription. The GTF file of additional (non-ENSEMBL) genes and transcripts along with the functional annotations of Mikado-derived genes can be found in **Supplementary File 2**.

scRNAseq read alignment and quantification

Raw FASTQ files from each of the nine samples were processed using CellRanger version 6.0.1 (10x Genomics, Inc.). The genome (fSalTru1.1, GCF_901001165.1) was indexed using 'cellranger mkref' with the augmented annotation file (see above), and reads were aligned to the genome using 'cellranger count' using non-default parameters –nosecondary and --include-introns, and otherwise default parameters.

Processing of count data and clustering

The filtered barcode matrices from CellRanger-count were processed with the Seurat R package (Butler et al., 2018), using the Reciprocal Principal Components Analysis (RPCA) approach to normalize expression estimates across the three sample collection batches. In the RPCA approach, normalization (by SCTransform) and PCA were initially performed separately on samples from each of three sampling batches before integrating according to anchor features. To avoid subsequent clustering being overly influenced by cell cycle stage, the anchor features upon which PCA was performed were filtered to remove genes with annotation related to 'cell cycle' (using information from ENSEMBL). Similarly, given that red blood cells likely contributed a significant source of variation, hemoglobin genes were also excluded from the PCA. However, genes excluded from PCA were retained in the final set of 2000 variable genes. For clustering, the FindNeighbors() function was run using the first 30 principal components. We then tested a series of resolution options for the FindClusters() step, from 0.1 to 1 in intervals of 0.1. Using clusters from all resolutions we then obtained a clustering tree using the clustree() function from the clustree package (Zappia & Oshlack, 2018), and obtained the SC3 index, an indicator of cluster stability, for each resolution. A resolution of 0.3 was found to have the highest SC3 value (**Fig. S8**) and we therefore chose this resolution for initial clusters (**Fig. S1**), but later performed additional subclustering of specific clusters in cases where subpopulations could be clearly distinguished from the expression of lineage markers (**Fig. S9**). No clear batch effects were evident from UMAP projections (**Fig. S10**).

Differential expression between clusters was performed by running the FindAllMarkers() function using the raw (non-integrated) expression values (given that the integrated values are statistically interdependent), producing a list of expression markers for each cluster (hereafter 'cluster markers'). For further analyses we principally considered cluster markers with log2FC >/= 1 and which were detected in at least 60% of cells of a given cluster ('high abundance' markers), except for when clusters could be notably distinguished by expression patterns restricted to a subset of cells in the cluster.

Assigning putative cluster identities

For assignment of clusters to broad immune cell classes, we examined the cluster-specific expression of putative cell type marker genes, a list of which was compiled a-priori. These genes (hereafter 'prior markers') were obtained by compiling a list of putative cell lineage markers from other species (humans in addition to some fish-specific markers), derived from the literature (**Table S1**). Putative brown trout

homologues of these prior markers were identified by queries to the ENSEMBL database using the biomaRt R package (Durinck et al., 2005) to obtain putative brown trout homologues in combination with reverse blast searches (ENSEMBL blast tool, TblastN, default parameters) towards the brown trout genome using human queries and manually annotated fish queries from the literature. Reverse BLAST hits were aligned towards human and mouse reference sequences and query sequences to evaluate the overall degree of amino acid conservation (MEGA7, clustalw aligner). A simple neighbor-joining tree (MEGA7, Poisson substitution model, pairwise deletion) was constructed for markers with more pronounced sequence divergence (data not shown). In cases where homology could not be determined the marker was not used. This yielded in a list of 283 homologues of 129 prior markers of major cell lineage groups (e.g. B-cells, neutrophils) and functional classes (e.g. antigen presenting cells, cytotoxic cells, proliferating cells). For prior marker homologues which were present amongst the 2000 variable genes, expression values were summarised across markers belonging to the same lineage group or functional class using the AddModuleScore() function in Seurat, deriving expression modules corresponding with each major lineage group or functional class. We note that the list of prior marker genes was intended as a starting point and was not intended to be an exhaustive list of putative markers.

After grouping the clusters into broad immune (or other) cell types (i.e. neutrophils, monocytes / macrophages, T-cells, B-cells, and others), we used the cluster markers to further identify distinguishing gene expression features and more specific putative function. For this we considered only cluster markers that were unique to the broad cell type under investigation (e.g. markers of both T-cell and B-cell clusters were not used, although markers of ambiguous cluster N2 were allowed to be considered as either neutrophil or B-cell markers). We also performed GO term enrichment analyses on the cluster markers (all markers with log2FC >/= 1 and 60% detection in the cluster, with no criteria for uniqueness to a broad cell type) using the enricher() function from the ClusterProfiler package and GO terms of both ENSEMBL and Mikado-derived genes.

Orthology analyses

To identify ohnologues amongst the 2000 variably expressed genes, *S. trutta* gene IDs were queried against the ENSEMBL database for homologues in the northern pike (*E. Lucius*) using the biomaRt package (Durinck et al., 2005). For genes with pike homologues, we filtered these to retain only those with a 2:1 ratio (trout:pike) and for which each member of the pair resided on a different trout chromosome. This resulted in a list of 152 putative ohnolog pairs amongst the variably expressed genes. UMAP plots for each pair were then visually assessed and assigned to one of the following four categories (similar to (Berthelot et al., 2014), but visually): 1) same pattern same levels (identical); 2) same pattern different levels; 3) gain/loss (with one ortholog expressed across a wider area than the other); 4) swap or partial swap (where orthologs have, at least partly, mutually exclusive expression patterns). Importantly, these assignments should be considered suggestive rather than final, particularly because gain/loss patterns may also result from differences in sequencing depth or sampling imbalances.

Pseudobulk analyses

Clusters to use for pseudobulk differential expression analyses were chosen based on the following criteria: that they were clearly distinguishable from other clusters (i.e., abundant unique high abundance markers and not overlapping other clusters on a UMAP plot), and that they contained a reasonable cell count in all individuals (minimum 45). This led to the selection of clusters N1, M1, T5, and B3. The choice of B3 was further motivated by its identification as the pro/pre-B-cell cluster and thus alterations in this cluster could potentially have knock-on effects on subsequently differentiated B-cells. After pooling non-normalised read counts from each cluster / individual combination, the resulting pseudobulk count matrices were processed for differential expression analyses using the edgeR package (Robinson et al., 2009). Genes were first filtered to retain only those with at least 10 counts in at least one of the three origin groups (wild, mix, farm) prior to TMM-normalisation. PCA was calculated using the plotMDS() function which uses the top 500 most variably expressed genes by default. Differential expression was calculated using the estimateDisp() and glmFit() functions with a model design matrix that included group as the main factor while also accounting for experimental batch [model.matrix(~batch+group)]. Genes were considered differentially expressed between groups at FDR < 0.05. GO term enrichment analysis (using the domain 'molecular function') was carried out using the enricher() function from the ClusterProfiler package (Yu et al., 2012) and GO terms of both ENSEMBL and Mikado-derived genes. For heatmaps, $\log_2$CPM values of differentially expressed genes were Z-score normalised and plotted using the Heatmap() function from the ComplexHeatmap package (Gu et al., 2016).

Pseudotime analyses

The integrated, normalized data (SCT assay data with 2000 features for 83,847 cells) was converted into a 'cell_data_set'-object with monocle (monocle3_1.3.4). After preprocessing the data (preprocess_cds) using the first 50 dimensions of the PCA a UMAP dimensionality reduction was performed. Cells were clustered with a resolution of 0.00001 which resulted in 24 clusters. This resolution was chosen because the number of cell clusters is close to the original 28. After running the 'learn_graph' function, subsets of clusters and starting points for trajectories were interactively selected based on prior biological knowledge. Two separate pseudotime analyses were performed: one on neutrophil clusters using the cluster corresponding with putative myeloid progenitors (cluster MP in main analysis) as the starting point and another on B-cells using the cluster corresponding with pre-/pro-B-cells (cluster B3 in the main analysis) as the starting point.

Data and code availability

All raw sequencing data will be deposited to the NCBI sequence read archive (SRA) upon publication. Code for reproducing the core analyses can be found in the following github repository: https://github.com/jamesord/browntrout_scseq.

Author contributions

Conceptualization – IAK, HS; Data Curation – JO; Formal Analysis – JO, SO, AB, MHS; Funding Acquisition – IAK, HS; Investigation (Experimentation, Data acquisition) HSM; Methodology (design and development) – HSM, HS, JO; MHS; Project Administration – JO, IAK, HS; Resources (provision of samples or materials) HSM; Supervision – IAK, HS; Visualization – JO, IAK, ST; Writing Draft – JO; Writing Review and Editing – IAK, HS, HSM, SO, ST, MHS

Funding sources

The work received funding from Project grant #212526 from the Swiss National Science Foundation awarded to IAK (https://data.snf.ch/grants/grant/212526) with additional funding from the Institute of Fish and Wildlife Health, University of Bern.

Conflict of interest statement

The authors declare no conflict of interest.

Acknowledgements

We thank the Fish Warden of the Canton of St Gallen and the staff and management of Fischereizentrum Steinach (Canton of St Gallen) for assistance with fish acquisition, Pamela Nicholson of the University of Bern Next Generation Sequencing Platform for advice regarding sample preparation and for performing 10x Chromium sequencing, and Dragan Stajic for insightful discussion. Calculations were performed on UBELIX, the high-performance computing cluster of the University of Bern.

References

Ahmad, F., Debes, P. V., Pukk, L., Kahar, S., Hartikainen, H., Gross, R., & Vasemägi, A. (2021). Know your enemy – transcriptome of myxozoan Tetracapsuloides bryosalmonae reveals potential drug targets against proliferative kidney disease in salmonids. *Parasitology*, *148*(6), 726–739. https://doi.org/10.1017/S003118202100010X

Ashman, L. K., Ferrao, P., Cole, S. R., & Cambareri, A. C. (1999). Effects of mutant c-Kit in early myeloid cells. *Leukemia & Lymphoma*, *34*(5–6), 451–461. https://doi.org/10.3109/10428199909058472

Athanasiadis, E. I., Botthof, J. G., Andres, H., Ferreira, L., Lio, P., & Cvejic, A. (2017). Single-cell RNA-sequencing uncovers transcriptional states and fate decisions in haematopoiesis. *Nature Communications*, *8*(1). https://doi.org/10.1038/S41467-017-02305-6

Azad, A. K., Rajaram, M. V. S., & Schlesinger, L. S. (2014). Exploitation of the Macrophage Mannose Receptor (CD206) in Infectious Disease Diagnostics and Therapeutics. *Journal of Cytology & Molecular Biology*, *1*(1), 1000003. https://doi.org/10.13188/2325-4653.1000003

Baer, J., Kugler, M., Schubert, M., Schotzko, N., Rösch, R., Vonlanthen, P., & DeWeber, J. T. (2023). A matter of time—Efficacy of whitefish stocking in a large pre-alpine lake. *Fisheries Management and Ecology*, *30*(6), 615–626. https://doi.org/10.1111/FME.12624

Bailey, C., Segner, H., Casanova-Nakayama, A., & Wahli, T. (2017). Who needs the hotspot? The effect of temperature on the fish host immune response to Tetracapsuloides bryosalmonae the

causative agent of proliferative kidney disease. *Fish & Shellfish Immunology*, *63*, 424–437. https://doi.org/10.1016/J.FSI.2017.02.039

Balke, N., Holtkamp, U., Hörl, W. H., & Tschesche, H. (1995). Inhibition of degranulation of human polymorphonuclear leukocytes by complement factor D. *FEBS Letters*, *371*(3), 300–302. https://doi.org/10.1016/0014-5793(95)00926-Z

Batista-Gonzalez, A., Vidal, R., Criollo, A., & Carreño, L. J. (2020). New Insights on the Role of Lipid Metabolism in the Metabolic Reprogramming of Macrophages. *Frontiers in Immunology*, *10*, 501852. https://doi.org/10.3389/FIMMU.2019.02993/BIBTEX

Berthelot, C., Brunet, F., Chalopin, D., Juanchich, A., Bernard, M., Noël, B., Bento, P., Da Silva, C., Labadie, K., Alberti, A., Aury, J. M., Louis, A., Dehais, P., Bardou, P., Montfort, J., Klopp, C., Cabau, C., Gaspin, C., Thorgaard, G. H., … Guiguen, Y. (2014). The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nature Communications 2014 5:1*, *5*(1), 1–10. https://doi.org/10.1038/ncomms4657

Bettge, K., Segner, H., Burki, R., Schmidt-Posthaus, H., & Wahli, T. (2009). Proliferative kidney disease (PKD) of rainbow trout: temperature- and time-related changes of Tetracapsuloides bryosalmonae DNA in the kidney. *Parasitology*, *136*(6), 615–625. https://doi.org/10.1017/S0031182009005800

Bi, L., Jia, S., Hu, W., Su, X., Chen, X., & Tang, H. (2022). Systematic analysis of prognostic significance, functional enrichment and immune implication of STK10 in acute myeloid leukemia. *BMC Medical Genomics*, *15*(1), 1–15. https://doi.org/10.1186/S12920-022-01251-7/FIGURES/8

Breitfeld, D., Ohl, L., Kremmer, E., Ellwart, J., Sallusto, F., Lipp, M., & Förster, R. (2000). Follicular B Helper T Cells Express Cxc Chemokine Receptor 5, Localize to B Cell Follicles, and Support Immunoglobulin Production. *The Journal of Experimental Medicine*, *192*(11), 1545. https://doi.org/10.1084/JEM.192.11.1545

Britt, E. C., Lika, J., Giese, M. A., Schoen, T. J., Seim, G. L., Huang, Z., Lee, P. Y., Huttenlocher, A., & Fan, J. (2022). Switching to the cyclic pentose phosphate pathway powers the oxidative burst in activated neutrophils. *Nature Metabolism 2022 4:3*, *4*(3), 389–403. https://doi.org/10.1038/s42255-022-00550-8

Burbach, B. J., Medeiros, R. B., Mueller, K. L., & Shimizu, Y. (2007). T-cell receptor signaling to integrins. *Immunological Reviews*, *218*(1), 65–81. https://doi.org/10.1111/J.1600-065X.2007.00527.X

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., & Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature Biotechnology 2018 36:5*, *36*(5), 411–420. https://doi.org/10.1038/nbt.4096

Chan, B., Lanyi, A., Song, H. K., Griesbach, J., Simarro-Grande, M., Poy, F., Howie, D., Sumegi, J., Terhorst, C., & Eck, M. J. (2003). SAP couples Fyn to SLAM immune receptors. *Nature Cell Biology 2003 5:2*, *5*(2), 155–160. https://doi.org/10.1038/ncb920

Chinen, T., Kannan, A. K., Levine, A. G., Fan, X., Klein, U., Zheng, Y., Gasteiger, G., Feng, Y., Fontenot, J. D., & Rudensky, A. Y. (2016). An essential role for the IL-2 receptor in Treg cell function. *Nature Immunology 2016 17:11*, *17*(11), 1322–1333. https://doi.org/10.1038/ni.3540

Colgan, T. J., Moran, P. A., Archer, L. C., Wynne, R., Hutton, S. A., McGinnity, P., & Reed, T. E. (2021). Evolution and Expression of the Immune System of a Facultatively Anadromous Salmonid. *Frontiers in Immunology*, *12*, 568729. https://doi.org/10.3389/FIMMU.2021.568729/XML/NLM

Crête-Lafrenière, A., Weir, L. K., & Bernatchez, L. (2012). Framing the Salmonidae Family Phylogenetic Portrait: A More Complete Picture from Increased Taxon Sampling. *PLOS ONE*, *7*(10), e46662. https://doi.org/10.1371/JOURNAL.PONE.0046662

Cuenca, M., Sintes, J., Lányi, Á., & Engel, P. (2019). CD84 cell surface signaling molecule: An emerging biomarker and target for cancer and autoimmune disorders. *Clinical Immunology*, *204*, 43–49. https://doi.org/10.1016/J.CLIM.2018.10.017

Davis, T. L., & Rebay, I. (2016). Master regulators in development: views from the Drosophila retinal determination and mammalian pluripotency gene networks. *Developmental Biology*, *421*(2), 93. https://doi.org/10.1016/J.YDBIO.2016.12.005

Dunaway, L. S., Loeb, S. A., Petrillo, S., Tolosano, E., & Isakson, B. E. (2024). Heme metabolism in nonerythroid cells. *Journal of Biological Chemistry*, *300*(4), 107132. https://doi.org/10.1016/J.JBC.2024.107132

Durinck, S., Moreau, Y., Kasprzyk, A., Davis, S., De Moor, B., Brazma, A., & Huber, W. (2005). BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics*, *21*(16), 3439–3440. https://doi.org/10.1093/BIOINFORMATICS/BTI525

Flies, A. S. (2020). Rewilding immunology Integrating comparative immunology can improve human, animal, and ecosystem health. *Science*, *369*(6499), 37–38. https://doi.org/10.1126/SCIENCE.ABB8664/SUPPL_FILE/ABB8664-FLIES-SM.PDF

Förster, R., Mattis, A. E., Kremmer, E., Wolf, E., Brem, G., & Lipp, M. (1996). A Putative Chemokine Receptor, BLR1, Directs B Cell Migration to Defined Lymphoid Organs and Specific Anatomic Compartments of the Spleen. *Cell*, *87*(6), 1037–1047. https://doi.org/10.1016/S0092-8674(00)81798-5

Ghosh, A., Garee, G., Sweeny, E. A., Nakamura, Y., & Stuehr, D. J. (2018). Hsp90 chaperones hemoglobin maturation in erythroid and nonerythroid cells. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(6), E1117–E1126. https://doi.org/10.1073/PNAS.1717993115/SUPPL_FILE/PNAS.201717993SI.PDF

Ghoshdastider, U., Popp, D., Burtnick, L. D., & Robinson, R. C. (2013). The expanding superfamily of gelsolin homology domain proteins. *Cytoskeleton*, *70*(11), 775–795. https://doi.org/10.1002/CM.21149

Gillard, G. B., Grønvold, L., Røsæg, L. L., Holen, M. M., Monsen, Ø., Koop, B. F., Rondeau, E. B., Gundappa, M. K., Mendoza, J., Macqueen, D. J., Rohlfs, R. V., Sandve, S. R., & Hvidsten, T. R. (2021). Comparative regulomics supports pervasive selection on gene dosage following whole genome duplication. *Genome Biology*, *22*(1), 1–18. https://doi.org/10.1186/S13059-021-02323-0/FIGURES/3

Gu, Z., Eils, R., & Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*, *32*(18), 2847–2849. https://doi.org/10.1093/BIOINFORMATICS/BTW313

Guerra, F. M., Gommerman, J. L., Corfe, S. A., Paige, C. J., & Rottapel, R. (2012). Homeodomain-interacting protein kinase (HIPK)-1 is required for splenic B cell homeostasis and optimal T-independent type 2 humoral response. *PloS One*, *7*(4). https://doi.org/10.1371/JOURNAL.PONE.0035533

Guilliams, M., Ginhoux, F., Jakubzick, C., Naik, S. H., Onai, N., Schraml, B. U., Segura, E., Tussiwand, R., & Yona, S. (2014). Dendritic cells, monocytes and macrophages: a unified

nomenclature based on ontogeny. *Nature Reviews Immunology 2014 14:8*, *14*(8), 571–578. https://doi.org/10.1038/nri3712

Guslund, N. C., Solbakken, M. H., Brieuc, M. S. O., Jentoft, S., Jakobsen, K. S., & Qiao, S. W. (2020). Single-Cell Transcriptome Profiling of Immune Cell Repertoire of the Atlantic Cod Which Naturally Lacks the Major Histocompatibility Class II System. *Frontiers in Immunology*, *11*, 2602. https://doi.org/10.3389/FIMMU.2020.559555/BIBTEX

Harrison, D. J., Kharbanda, R., Cunningham, D. S., McLellan, L. I., & Hayes, J. D. (1989). Distribution of glutathione S-transferase isoenzymes in human kidney: basis for possible markers of renal injury. *Journal of Clinical Pathology*, *42*(6), 624–628. https://doi.org/10.1136/JCP.42.6.624

Harvie, E. A., & Huttenlocher, A. (2015). Neutrophils in host defense: new insights from zebrafish. *Journal of Leukocyte Biology*, *98*(4), 523–537. https://doi.org/10.1189/JLB.4MR1114-524R/-/DC1

Hashemzadeh Segherloo, I., Freyhof, J., Berrebi, P., Ferchaud, A. L., Geiger, M., Laroche, J., Levin, B. A., Normandeau, E., & Bernatchez, L. (2021). A genomic perspective on an old question: Salmo trouts or Salmo trutta (Teleostei: Salmonidae)? *Molecular Phylogenetics and Evolution*, *162*, 107204. https://doi.org/10.1016/J.YMPEV.2021.107204

Hitchcock, I. S., Hafer, M., Sangkhae, V., & Tucker, J. A. (2021). The thrombopoietin receptor: revisiting the master regulator of platelet production. *Platelets*, *32*(6), 770. https://doi.org/10.1080/09537104.2021.1925102

Hu, C., Zhang, N., Hong, Y., Tie, R., Fan, D., Lin, A., Chen, Y., Xiang, L. X., & Shao, J. Z. (2024). Single-cell RNA sequencing unveils the hidden powers of zebrafish kidney for generating both hematopoiesis and adaptive antiviral immunity. *ELife*, *13*. https://doi.org/10.7554/ELIFE.92424

Hu, Y., Zhang, Z., Kashiwagi, M., Yoshida, T., Joshi, I., Jena, N., Somasundaram, R., Emmanuel, A. O., Sigvardsson, M., Fitamant, J., El-Bardeesy, N., Gounari, F., Van Etten, R. A., & Georgopoulos, K. (2016). Superenhancer reprogramming drives a B-cell-epithelial transition and high-risk leukemia. *Genes and Development*, *30*(17), 1971–1990. https://doi.org/10.1101/GAD.283762.116/-/DC1

Karsunky, H., Merad, M., Cozzio, A., Weissman, I. L., & Manz, M. G. (2003). Flt3 Ligand Regulates Dendritic Cell Development from Flt3+ Lymphoid and Myeloid-committed Progenitors to Flt3+ Dendritic Cells In Vivo. *The Journal of Experimental Medicine*, *198*(2), 305. https://doi.org/10.1084/JEM.20030323

Kartal-Kaess, M., Cimmino, L., Infantino, S., Yabas, M., Zhang, J.-G., Tuohey, L., Gelman, I. H., Scott, J. D., Nutt, S. L., Enders, A., Tarlinton, D., & Dickins, R. A. (2012). RNAi Screening Identifies A Novel Role for A-Kinase Anchoring Protein 12 (AKAP12) in B Cell Development and Function. *Blood*, *120*(21), 855–855. https://doi.org/10.1182/BLOOD.V120.21.855.855

Keshav, S., Chung, P., Milon, G., & Gordon, S. (1991). Lysozyme is an inducible marker of macrophage activation in murine tissues as demonstrated by in situ hybridization. *Journal of Experimental Medicine*, *174*(5), 1049–1058. https://doi.org/10.1084/JEM.174.5.1049

Kim, J., Li, C. L., Chen, X., Cui, Y., Golebiowski, F. M., Wang, H., Hanaoka, F., Sugasawa, K., & Yang, W. (2023). Lesion recognition by XPC, TFIIH and XPA in DNA excision repair. *Nature 2023 617:7959*, *617*(7959), 170–175. https://doi.org/10.1038/s41586-023-05959-z

Kulkarni, R. (2022). Early Growth Response Factor 1 in Aging Hematopoietic Stem Cells and Leukemia. *Frontiers in Cell and Developmental Biology*, *10*, 925761. https://doi.org/10.3389/FCELL.2022.925761

Kuo, T. C., & Schlissel, M. S. (2009). Mechanisms controlling expression of the RAG locus during lymphocyte development. *Current Opinion in Immunology*, *21*(2), 173. https://doi.org/10.1016/J.COI.2009.03.008

Lecaudey, L. A., Schliewen, U. K., Osinov, A. G., Taylor, E. B., Bernatchez, L., & Weiss, S. J. (2018). Inferring phylogenetic structure, hybridization and divergence times within Salmoninae (Teleostei: Salmonidae) using RAD-sequencing. *Molecular Phylogenetics and Evolution*, *124*, 82–99. https://doi.org/10.1016/J.YMPEV.2018.02.022

Leung, L. L. K. (1984). Role of thrombospondin in platelet aggregation. *The Journal of Clinical Investigation*, *74*(5), 1764–1772. https://doi.org/10.1172/JCI111595

Liang, T., Guo, Y., Li, M., Ding, C., Sang, S., Zhou, T., Shao, Q., Liu, X., Lu, J., Ji, Z., Wang, T., & Kang, Q. (2020). Cytoskeleton protein 4.1R regulates B-cell fate by modulating the canonical NF-κB pathway. *Immunology*, *161*(4), 314–324. https://doi.org/10.1111/IMM.13250

Liu, X., Li, Y.-S., Shinton, S. A., Rhodes, J., Tang, L., Feng, H., Jette, C. A., Look, A. T., Hayakawa, K., & Hardy, R. R. (2017). Zebrafish B Cell Development without a Pre–B Cell Stage, Revealed by CD79 Fluorescence Reporter Transgenes. *The Journal of Immunology*, *199*(5), 1706–1715. https://doi.org/10.4049/JIMMUNOL.1700552/-/DCSUPPLEMENTAL

Lobón-Cerviá, J., & Sanz, N. (2017). Brown Trout: Life history, ecology and management. *Brown Trout: Life History, Ecology and Management*, 1–862. https://doi.org/10.1002/9781119268352

Ludwig, J., Federico, G., Prokosch, S., Küblbeck, G., Schmitt, S., Klevenz, A., Gröne, H.-J., Nitschke, L., & Arnold, B. (2015). Dickkopf-3 Acts as a Modulator of B Cell Fate and Function. *The Journal of Immunology*, *194*(6), 2624–2634. https://doi.org/10.4049/JIMMUNOL.1402160

Macintyre, A. N., Finlay, D., Preston, G., Sinclair, L. V., Waugh, C. M., Tamas, P., Feijoo, C., Okkenhaug, K., & Cantrell, D. A. (2011). Protein Kinase B Controls Transcriptional Programs that Direct Cytotoxic T Cell Fate but Is Dispensable for T Cell Metabolism. *Immunity*, *34*(2), 224. https://doi.org/10.1016/J.IMMUNI.2011.01.012

Macqueen, D. J., & Johnston, I. A. (2014). A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1778). https://doi.org/10.1098/RSPB.2013.2881

Malmstrøm, M., Matschiner, M., Tørresen, O. K., Star, B., Snipen, L. G., Hansen, T. F., Baalsrud, H. T., Nederbragt, A. J., Hanel, R., Salzburger, W., Stenseth, N. C., Jakobsen, K. S., & Jentoft, S. (2016). Evolution of the immune system influences speciation rates in teleost fishes. *Nature Genetics 2016 48:10*, *48*(10), 1204–1210. https://doi.org/10.1038/ng.3645

Marić, A., Srećković Batoćanin, D., Škraba Jurlina, D., Brkušanin, M., Karanović, J., Kanjuh, T., Nikolić, V., Mrdak, D., & Simonović, P. (2023). A treatise about reliability in dating events of evolutionary history of brown trout Salmo cf. trutta (Actinopterygii) at Western Balkans: Impassable barriers, isolation of populations and assistance of geological timeframe. *ACTA ICHTHYOLOGICA ET PISCATORIA*, *53*(1), 1–18. https://doi.org/10.3897/AIEP.53.97702

McDougall, C., Hammond, M. J., Dailey, S. C., Somorjai, I. M. L., Cummins, S. F., & Degnan, B. M. (2018). The evolution of ependymin-related proteins. *BMC Evolutionary Biology*, *18*(1), 1–13. https://doi.org/10.1186/S12862-018-1306-Y/FIGURES/5

Michaud, D., Mirlekar, B., Steward, C., Bishop, G., & Pylayeva-Gupta, Y. (2022). B Cell Receptor Signaling and Protein Kinase D2 Support Regulatory B Cell Function in Pancreatic Cancer. *Frontiers in Immunology*, *12*, 745873. https://doi.org/10.3389/FIMMU.2021.745873/FULL

Milgroom, M. G. (2023). Vertebrate Immune Systems. *Biology of Infectious Disease*, 123–141. https://doi.org/10.1007/978-3-031-38941-2_9

Mirete-Bachiller, S., Olivieri, D. N., & Gambón-Deza, F. (2021). Immunoglobulin T genes in Actinopterygii. *Fish & Shellfish Immunology*, *108*, 86–93. https://doi.org/10.1016/J.FSI.2020.11.027

Murphy, A. E., Skene, N. G., Arising, F., Zimmerman, K. D., Langefeld, M. A. &, & Nature, C. D. (2022). A balanced measure shows superior performance of pseudobulk methods in single-cell RNA-sequencing analysis. *Nature Communications 2022 13:1*, *13*(1), 1–4. https://doi.org/10.1038/s41467-022-35519-4

Nagasawa, T., Kawaguchi, M., Nishi, K., & Yasumasu, S. (2022). Molecular evolution of hatching enzymes and their paralogous genes in vertebrates. *BMC Ecology and Evolution*, *22*(1), 9. https://doi.org/10.1186/S12862-022-01966-2

Nangaku, M., Alpers, C. E., Pippin, J., Shankland, S. J., Kurokawa, K., Adler, S., Morgan, B. P., Johnson, R. J., & Couser, W. G. (1998). CD59 protects glomerular endothelial cells from immune-mediated thrombotic microangiopathy in rats. *Journal of the American Society of Nephrology : JASN*, *9*(4), 590–597. https://doi.org/10.1681/ASN.V94590

Novoa, B., Pereiro, P., López-Muñoz, A., Varela, M., Forn-Cuní, G., Anchelin, M., Dios, S., Romero, A., Martinez-López, A., Medina-Gali, R. M., Collado, M., Coll, J., Estepa, A., Cayuela, M. L., Mulero, V., & Figueras, A. (2019). Rag1 immunodeficiency-induced early aging and senescence in zebrafish are dependent on chronic inflammation and oxidative stress. *Aging Cell*, *18*(5), e13020. https://doi.org/10.1111/ACEL.13020

Palikova, M., Papezikova, I., Markova, Z., Navratil, S., Mares, J., Mares, L., Vojtek, L., Hyrsl, P., Jelinkova, E., & Schmidt-Posthaus, H. (2017). Proliferative kidney disease in rainbow trout (Oncorhynchus mykiss) under intensive breeding conditions: Pathogenesis and haematological and immune parameters. *Veterinary Parasitology*, *238*, 5–16. https://doi.org/10.1016/J.VETPAR.2017.03.003

Perdiguero, P., Jiménez-Barrios, P., Morel, E., Abós, B., & Tafalla, C. (2024). Single-cell atlas of rainbow trout peripheral blood leukocytes and profiling of their early response to infectious pancreatic necrosis virus. *Frontiers in Immunology*, *15*, 1404209. https://doi.org/10.3389/FIMMU.2024.1404209/BIBTEX

Qu, J., Jin, J., Zhang, M., & Ng, L. G. (2023). Neutrophil diversity and plasticity: Implications for organ transplantation. *Cellular & Molecular Immunology 2023 20:9*, *20*(9), 993–1001. https://doi.org/10.1038/s41423-023-01058-1

Rastogi, S., & Liberles, D. A. (2005). Subfunctionalization of duplicated genes as a transition state to neofunctionalization. *BMC Evolutionary Biology*, *5*, 28. https://doi.org/10.1186/1471-2148-5-28

Rehberger, K., Wernicke von Siebenthal, E., Bailey, C., Bregy, P., Fasel, M., Herzog, E. L., Neumann, S., Schmidt-Posthaus, H., & Segner, H. (2020). Long-term exposure to low 17α-ethinylestradiol (EE2) concentrations disrupts both the reproductive and the immune system of juvenile rainbow trout, Oncorhynchus mykiss. *Environment International*, *142*, 105836. https://doi.org/10.1016/J.ENVINT.2020.105836

Reth, M., & Nielsen, P. (2014). Signaling Circuits in Early B-Cell Development. *Advances in Immunology*, *122*, 129–175. https://doi.org/10.1016/B978-0-12-800267-4.00004-3

Robertson, F. M., Gundappa, M. K., Grammes, F., Hvidsten, T. R., Redmond, A. K., Lien, S., Martin, S. A. M., Holland, P. W. H., Sandve, S. R., & Macqueen, D. J. (2017). Lineage-specific rediploidization is a mechanism to explain time-lags between genome duplication and

evolutionary diversification. *Genome Biology*, *18*(1), 1–14. https://doi.org/10.1186/S13059-017-1241-Z/FIGURES/7

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2009). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139. https://doi.org/10.1093/BIOINFORMATICS/BTP616

Sadiku, P., Willson, J. A., Ryan, E. M., Sammut, D., Coelho, P., Watts, E. R., Grecian, R., Young, J. M., Bewley, M., Arienti, S., Mirchandani, A. S., Sanchez Garcia, M. A., Morrison, T., Zhang, A., Reyes, L., Griessler, T., Jheeta, P., Paterson, G. G., Graham, C. J., … Walmsley, S. R. (2021). Neutrophils Fuel Effective Immune Responses through Gluconeogenesis and Glycogenesis. *Cell Metabolism*, *33*(2), 411-423.e4. https://doi.org/10.1016/J.CMET.2020.11.016

Sandve, S. R., Rohlfs, R. V., & Hvidsten, T. R. (2018). Subfunctionalization versus neofunctionalization after whole-genome duplication. *Nature Genetics 2018 50:7*, *50*(7), 908–909. https://doi.org/10.1038/s41588-018-0162-4

Saroz, Y., Kho, D. T., Glass, M., Graham, E. S., & Grimsey, N. L. (2019). Cannabinoid Receptor 2 (CB2) Signals via G-alpha-s and Induces IL-6 and IL-10 Cytokine Secretion in Human Primary Leukocytes. *ACS Pharmacology and Translational Science*, *2*(6), 414–428. https://doi.org/10.1021/ACSPTSCI.9B00049/ASSET/IMAGES/MEDIUM/PT9B00049_M001.GIF

Schmidt-Posthaus, H., Steiner, P., Müller, B., & Casanova-Nakayama, A. (2013). Complex interaction between proliferative kidney disease, water temperature and concurrent nematode infection in brown trout. *Diseases of Aquatic Organisms*, *104*(1), 23–34. https://doi.org/10.3354/DAO02580

Serwas, N. K., Hoeger, B., Ardy, R. C., Stulz, S. V., Sui, Z., Memaran, N., Meeths, M., Krolo, A., Yüce Petronczki, Ö., Pfajfer, L., Hou, T. Z., Halliday, N., Santos-Valente, E., Kalinichenko, A., Kennedy, A., Mace, E. M., Mukherjee, M., Tesi, B., Schrempf, A., … Boztug, K. (2019). Human DEF6 deficiency underlies an immunodeficiency syndrome with systemic autoimmunity and aberrant CTLA-4 homeostasis. *Nature Communications 2019 10:1*, *10*(1), 1–15. https://doi.org/10.1038/s41467-019-10812-x

Sharma, N., Akhade, A. S., & Qadri, A. (2013). Sphingosine-1-phosphate suppresses TLR-induced CXCL8 secretion from human T cells. *Journal of Leukocyte Biology*, *93*(4), 521–528. https://doi.org/10.1189/JLB.0712328

Solbakken, M. H., Voje, K. L., Jakobsen, K. S., & Jentoft, S. (2017). Linking species habitat and past palaeoclimatic events to evolution of the teleost innate immune system. *Proceedings. Biological Sciences*, *284*(1853). https://doi.org/10.1098/RSPB.2016.2810

Song, W., Liu, C., Seeley-Fallen, M. K., Miller, H., Ketchum, C., & Upadhyaya, A. (2013). Actin-mediated feedback loops in B-cell receptor signaling. *Immunological Reviews*, *256*(1), 177–189. https://doi.org/10.1111/IMR.12113

Strepparava, N., Segner, H., Ros, A., Hartikainen, H., Schmidt-Posthaus, H., & Wahli, T. (2018). Temperature-related parasite infection dynamics: the case of proliferative kidney disease of brown trout. *Parasitology*, *145*(3), 281–291. https://doi.org/10.1017/S0031182017001482

Sudhagar, A., Ertl, R., Kumar, G., & El-Matbouli, M. (2019). Transcriptome profiling of posterior kidney of brown trout, Salmo trutta, during proliferative kidney disease. *Parasites and Vectors*, *12*(1), 1–17. https://doi.org/10.1186/S13071-019-3823-Y/FIGURES/6

Suurväli, J., Garroway, C. J., & Boudinot, P. (2022). Recurrent expansions of B30.2-associated immune receptor families in fish. *Immunogenetics*, *74*(1), 129–147. https://doi.org/10.1007/S00251-021-01235-4

Swann, J. B., Holland, S. J., Petersen, M., Pietsch, T. W., & Boehm, T. (2020). The immunogenetics of sexual parasitism. *Science*, *369*(6511). https://doi.org/10.1126/SCIENCE.AAZ9445/SUPPL_FILE/AAZ9445_TABLES_S1_TO_S8.ZIP

Swanson, J. A., Johnson, M. T., Beningo, K., Post, P., Mooseker, M., & Araki, N. (1999). A contractile activity that closes phagosomes in macrophages. *Journal of Cell Science*, *112 (Pt 3)*(3), 307–316. https://doi.org/10.1242/JCS.112.3.307

Uhlen, M., Karlsson, M. J., Zhong, W., Tebani, A., Pou, C., Mikes, J., Lakshmikanth, T., Forsström, B., Edfors, F., Odeberg, J., Mardinoglu, A., Zhang, C., von Feilitzen, K., Mulder, J., Sjöstedt, E., Hober, A., Oksvold, P., Zwahlen, M., Ponten, F., … Brodin, P. (2019). A genome-wide transcriptomic analysis of protein-coding genes in human blood cells. *Science*, *366*(6472). https://doi.org/10.1126/SCIENCE.AAX9198/SUPPL_FILE/AAX9198_UHLEN_SM.PDF

Veličković, T., Snoj, A., Simić, V., Šanda, R., Vukić, J., Barcytė, D., Stanković, D., & Marić, S. (2023). A new perspective on the molecular dating of the brown trout complex with an extended phylogeographic information on the species in Serbia. *Contributions to Zoology*, *92*(4), 362–389. https://doi.org/10.1163/18759866-BJA10046

Venturini, L., Caim, S., Kaithakottil, G. G., Mapleson, D. L., & Swarbreck, D. (2018). Leveraging multiple transcriptome assembly methods for improved gene structure annotation. *GigaScience*, *7*(8), 1–15. https://doi.org/10.1093/GIGASCIENCE/GIY093

Wang, L., Shen, S., Xiao, H., Ding, F., Wang, M., Li, G., & Hu, F. (2019). ARHGAP24 inhibits cell proliferation and cell cycle progression and induces apoptosis of lung cancer via a STAT6-WWP2-p27 axis. *Carcinogenesis*, *41*(5), bgz144. https://doi.org/10.1093/CARCIN/BGZ144

Wang, Y., Chiappetta, G., Guérois, R., Liu, Y., Romero, S., Boesch, D. J., Krause, M., Dessalles, C. A., Babataheri, A., Barakat, A. I., Chen, B., Vinh, J., Polesskaya, A., & Gautreau, A. M. (2023). PPP2R1A regulates migration persistence through the NHSL1-containing WAVE Shell Complex. *Nature Communications 2023 14:1*, *14*(1), 1–18. https://doi.org/10.1038/s41467-023-39276-w

Wilker, P. R., Kohyama, M., Sandau, M. M., Albring, J. C., Nakagawa, O., Schwarz, J. J., & Murphy, K. M. (2008). Transcription factor Mef2c is required for B cell proliferation and survival after antigen receptor stimulation. *Nature Immunology*, *9*(6), 603. https://doi.org/10.1038/NI.1609

Wu, L., Gao, A., Li, L., Chen, J., Li, J., & Ye, J. (2021). A Single-Cell Transcriptome Profiling of Anterior Kidney Leukocytes From Nile Tilapia (Oreochromis niloticus). *Frontiers in Immunology*, *12*, 783196. https://doi.org/10.3389/FIMMU.2021.783196/BIBTEX

Yazawa, R., Cooper, G. A., Beetz-Sargent, M., Robb, A., McKinnel, L., Davidson, W. S., & Koop, B. F. (2008). Functional adaptive diversity of the Atlantic salmon T-cell receptor gamma locus. *Molecular Immunology*, *45*(8), 2150–2157. https://doi.org/10.1016/J.MOLIMM.2007.12.007

Yu, G., Wang, L. G., Han, Y., & He, Q. Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS : A Journal of Integrative Biology*, *16*(5), 284. https://doi.org/10.1089/OMI.2011.0118

Yuan, W., Payton, J. E., Holt, M. S., Link, D. C., Watson, M. A., DiPersio, J. F., & Ley, T. J. (2007). Commonly dysregulated genes in murine APL cells. *Blood*, *109*(3), 961–970. https://doi.org/10.1182/BLOOD-2006-07-036640

Zappia, L., & Oshlack, A. (2018). Clustering trees: a visualization for evaluating clusterings at multiple resolutions. *GigaScience*, *7*(7), 1–9. https://doi.org/10.1093/GIGASCIENCE/GIY083

Zhang, H., Wada, J., Hida, K., Tsuchiyama, Y., Hiragushi, K., Shikata, K., Wang, H., Lin, S., Kanwar, Y. S., & Makino, H. (2001). Collectrin, a collecting duct-specific transmembrane glycoprotein, is

a novel homolog of ACE2 and is developmentally regulated in embryonic kidneys. *The Journal of Biological Chemistry*, *276*(20), 17132–17139. https://doi.org/10.1074/JBC.M006723200

Zhao, X., Huang, W., Shi, Y., Guo, J., Xiao, H., Ji, N., Feng, J., Dang, H., & Zou, J. (2022). PLAAT1 inhibits type I interferon response via degradation of IRF3 and IRF7 in Zebrafish. *Frontiers in Immunology*, *13*, 979919. https://doi.org/10.3389/FIMMU.2022.979919/BIBTEX