Deciphering the evolution of vertebrate immune cell types with single-cell RNA-seq

Santiago J Carmona[1,2] and David Gfeller[1,2]

[1] Department of Oncology, Ludwig Institute for Cancer Research, University of Lausanne, Switzerland.
[2] Swiss Institute of Bioinformatics (SIB), Lausanne, Switzerland.

To whom correspondence should be addressed: david.gfeller@unil.ch, santiago.carmona@unil.ch

Abstract

Single-cell RNA-seq is revolutionizing our understanding of cell type heterogeneity in many fields of biology, ranging from neuroscience to cancer to immunology. In Immunology, one of the main promises of this approach is the ability to define cell types as clusters in the whole transcriptome space (i.e., without relying on specific surface markers), thereby providing an unbiased classification of immune cell types. So far, this technology has been mainly applied in mouse and human. However, technically it could be used for immune cell-type identification in any species without requiring the development and validation of species-specific antibodies for cell sorting. Here we review recent developments using single-cell RNA-seq to characterize immune cell populations in non-mammalian vertebrates, with a focus on zebrafish (*Danio rerio*). We advocate that single-cell RNA-seq technology is likely to provide key insights into our understanding of the evolution of the adaptive immune system.

Introduction - immune cell types in vertebrates

The adaptive immune system, as defined in mammals, is centered on lymphocytes bearing antigen receptors generated by recombination-activating gene (RAG)-mediated rearrangement. This highly complex system arose early in the evolution of jawed vertebrates approximately 500 million years ago (Cooper and Alder, 2006; Flajnik and Kasahara, 2010). The adaptive immune system is able to generate an enormous diversity of B and T cell receptors that can recognize and eliminate newly encountered pathogens and develop memory to protect against previously encountered pathogens. The adaptive immune system of jawed vertebrates has a functional counterpart in jawless vertebrates (lampreys and hagfish), involving somatic rearrangement of variable regions in Leucine-rich repeat proteins (Pancer et al., 2004). Mammalian species also have an innate immune system that does not possess the ability to somatically rearrange receptors but relies on largely expanded families of genes such Toll-like receptors (TLRs) or

Killer Cell Immunoglobulin Like Receptors (KIRs). The innate immune system is key for mounting a rapid response against pathogens and for modulating the adaptive immune response. The innate and adaptive immune systems further play central roles in wound healing, tissue remodeling and homeostasis.

Different types of immune responses are mediated by different types of immune cells. In mammals, immunologists have been able to study the immune system at a very high resolution, delineating more than 10 main cell types, including CD4/CD8 T Cells, B Cells, Dendritic Cells (DC), Macrophages, Granulocytes, Natural Killers (NK) or other Innate Lymphoid Cells (ILC). These can be further split into tens of distinct cellular sub-types or states (e.g. Th1, Th2, Th17 CD4 T Cells, M1/M2 polarized macrophages, >10 DCs subtypes, etc.). Immunologists take advantage of monoclonal antibodies for more than 300 cell surface markers that allow fine dissection of immune cell types (Clark et al., 2016). Once a large number of such markers are available, new populations can be discovered by exploring T cell subsets expressing different combinations of them (e.g. Lin- CD123+ CD127low innate lymphoid cells (Mora-Velandia et al., 2017)). This approach has proved very powerful in mouse and human and resulted in a detailed phenotypic and functional characterization of immune cell populations in these two species.

Historically, lymphocytes where first discriminated from myeloid cells among white blood cells in 1879 (Vivier et al., 2016). Almost a century later, B and T cells were distinguished in birds and mammals (Warner et al., 1962), and later NK cells were identified (Kiessling et al., 1975) as non-T, non-B lymphocytes with the capacity to spontaneously kill tumor cells without the need for prior immunization. In 1997, lymphoid-tissue-inducer (LTi) cells were identified (Mebius et al., 1997). More recently, new types of non-T, non-B lymphocytes were identified, and have been classified as 'innate lymphoid cells' including subsets ILC1, ILC2, ILC3, NK cells and LTi (Spits et al., 2013; Vivier et al., 2016).

Since many decades, people realized that studying the immune system of evolutionary distant vertebrates would be a powerful approach to define selection pressures that have shaped mechanisms, molecules and specialized cells or structures during the evolution of the adaptive immune system (Cooper and Alder, 2006; Flajnik and Kasahara, 2010; Pancer et al., 2004). Approaches based on anatomy and histology were historically the first to investigate immune cells in different species. Birds were often used for such studies (Cooper et al., 1965) and is noteworthy to recall that B cells were first discovered in birds as cells developing in Bursa of Fabricius, hence the "B" of B cells (Cooper et al., 1966). Jawed fish were observed to possess lymphocyte populations that are analogous to T cells, B cells and non-specific cytotoxic cells, macrophages and polymorphonuclear leukocytes (1996). Comparative studies also revealed that the same organs perform different roles in the development of immune cells in different species. For instance, in teleost fish kidneys play the role of bone marrow in mammalian species for the development of T and B cells (Zapata and Amemiya, 2000).

With the availability of genome sequencing technologies, most of our current understanding of immune cell types in non-mammalian species and their evolution comes from comparative genomic studies. These studies have provided very important insights into the evolution of the

vertebrates' immune system, including tracing the origin of RAG (Morales Poole et al., 2017) or the discovery of evolutionary distant cytokines, chemokines and their receptors (Jacobson et al., 2017). Many genes used as markers of mammalian immune cell populations are conserved in all jawed vertebrates, suggesting that the corresponding cell types are also conserved (Cooper and Alder, 2006; Flajnik and Kasahara, 2010; Vivier et al., 2016). Unfortunately, comparative genomics studies are limited to orthology-based relationships and cannot provide phenotypic or functional information about immune cell populations.

Gene expression analyses of tissues using cDNA microarrays and RNA-seq have also greatly contributed to our understanding of immune responses across multiple vertebrates. These include studying immune responses of teleost fish to pathogens (Sudhagar et al., 2018) or nutrition (Martin and Król, 2017), or the discovery of new cytokines (Jacobson et al., 2017). Unfortunately, in 'bulk' gene expression profiling of multicellular organisms, the mixing of numerous cell types results in the dilution and masking the heterogeneity of immune cell types. Furthermore, a deep understanding of the evolution of the immune system will ultimately require the study of individual immune cell types, their ontogeny, differentiation, function and interaction with other cell types.

To overcome these limitations, different tools have been developed such as antibodies or transgenic lines. In the field of comparative immunology of jawed vertebrates, the teleost Zebrafish (*Danio rerio*) has become a reference species for most immunological studies in non-mammalian vertebrates. This organism presents many important advantages such as the possibility to conduct *in vivo* imaging and the availability of powerful genetic tools (Renshaw and Trede, 2012). The genome of zebrafish is also among the best annotated genomes of non-mammalian vertebrates (Howe et al., 2013).
Classical studies of hematopoiesis in zebrafish have defined major blood lineages. In the kidney marrow, the site of adult hematopoiesis, erythroid, lymphoid, and myeloid cells are present and can be separated based on light-scattering characteristics (Traver et al., 2003). However, subsets of cells within each of these lineages are poorly defined, primarily because of the lack of markers to distinguish them, therefore limiting our understanding of immune cell types.

Immune cell-specific monoclonal antibodies can be in principle developed for any species from which a marker gene can be cloned. For example, monoclonal antibodies have been recently developed for lampreys, enabling the study of novel lymphocytic lineages (Hirano et al., 2013) at the base of the vertebrates' phylogenetic tree. However, the high costs and technical challenges associated to the development of such tools for less studied species, in particular outside of human and mouse, significantly limit characterization of immune cells types. In the absence of appropriate antibodies, the use of fluorescent reporter lines has greatly contributed to the characterization of immune cells in zebrafish (Udvadia and Linney, 2003). Many lines are available where fluorescent proteins are expressed under the control of immune cells' specific promoters, such as *mpx* for neutrophils (Renshaw et al., 2006), *lck* for T cells and NK cells (Carmona et al., 2017; Langenau et al., 2004), *foxp3* for T regs (Kasheta et al., 2017) and *rag2* for B cells (Page et al., 2013). Isolation of these fluorescent cells of a particular lineage from whole organs by FACS enables population-based gene profiling (Rougeot et al., 2014) or

targeted single-cell gene expression analysis (using multiplex single-cell RT-PCR (Moore et al., 2016)). Thanks to the availability of reporter lines like these, it has been also possible to study transcriptional changes of immune cells during the course of infection (Saraceni et al., 2016).
However, few immune cell types are clearly defined in molecular terms, and therefore markers or promoters suitable for the prospective isolation of defined cell populations are available only for a handful of immune cell types in zebrafish, and are almost absent in other non-mammalian vertebrate species.

Recently, the development of sensitive and accurate single-cell RNA-sequencing (scRNA-seq) protocols have profoundly transformed our ability to comprehensively analyze immune cells and deconvolute their heterogeneity with virtually no need for cell sorting based on specific markers, opening the possibility to study immune cells in any vertebrate species.

Single-cell RNA-seq in immunology

Over the past years, numerous scRNA-seq protocols have been developed. Currently, all scRNA-seq protocols consist of first retro-transcription of mRNA in each cell into cDNA, second amplification of cDNA by polymerase chain reaction (PCR) or *in vitro* transcription, and finally sequencing of cDNA libraries, enabling quantification of the expression level of individual transcripts. Kolodziejczyk *et al.* (Kolodziejczyk et al., 2015) provide a comprehensive review of individual scRNA-seq protocols and their relative strengths and weaknesses. Different techniques can be classified based on the sequencing coverage of the transcripts. Methods such as Smart-Seq2 (Picelli et al., 2014) and CEL-seq2 (Hashimshony et al., 2016) enable full-length mRNA sequencing, allowing isoform quantification and *de novo* transcriptome assembly. Droplet-based methods such as 10X-Genomics (Zheng et al., 2017) and inDrops (Zilionis et al., 2016) generate only 3' terminal-end short sequencing reads, due to the integration of a sequencing priming site on the oligodT primer used for reverse transcription.
Droplet-based methods have a high throughput (tens of thousands of cells). Plate-based methods such as Smart-seq2 have lower throughput (hundreds of cells) but can be coupled with FACS index sorting, simultaneously producing light scattering and protein expression measurements along with single-cell transcriptomes.
Moreover, full-length mRNA sequencing protocols such as Smart-seq2 allow for coupled T and B cell receptor reconstruction and transcriptome quantification in the same cell, for species where VDJ loci have been annotated, opening the door for high throughput simultaneous analysis of T and B cell heterogeneity and clonality (Afik et al., 2017; Canzar et al., 2016; Eltahla et al., 2016; Stubbington et al., 2016).

Current scRNA-seq protocols amplify about 10 percent of the transcripts in each cell (Islam et al., 2014). Moreover, burst-like stochastic activation of transcription (Raj and van Oudenaarden, 2008) introduces additional intrinsic variability (gene 'drop-outs') in measurements of single-cell transcriptomes. This represents a limitation to detect low-abundance transcripts in a single cell. However, in practice, the large number of cells compensates this weakness and the full transcriptome of any cell type can be accurately reconstructed by computationally pooling (i.e.,

clustering) cells with similar transcriptomes (see later for estimations on the minimum number of cells for scRNA-seq required to sample a population of interest). As such, unsupervised clustering or other more supervised classification of cells into cell-types or states is a fundamental task in scRNA-seq data analysis. The goal is to detect clusters of cells that have sufficiently similar transcriptomes with other cells of the same cluster, and significantly different trasncriptomes from cells in other clusters. Unsupervised approaches use the whole transcriptome or the set of most variable genes, without additional information. More supervised approaches can pre-select specific markers based on a priori knowledge. However, as mentioned earlier, the high drop-out rate in scRNA-seq often results in inaccurate or incomplete classification when using only on a few markers (see example in Figure 1 and discussion in (Carmona et al., 2017)). In addition to the classical K-means and hierarchical clustering that perform well, tens of clustering methods have been applied to the task of clustering scRNAseq data (reviewed in (Andrews and Hemberg, 2018)). Very closely related to clustering are the concepts of dimensionality reduction and visualization with methods like Principle Component Analysis (PCA), Multi-Dimensional Scaling (MDS) or t-distributed Stochastic Neighbor Embedding (t-SNE), all of which are fundamental to summarize high-dimensional data produced by scRNA-seq assays and enable human interpretation of cell heterogeneity.

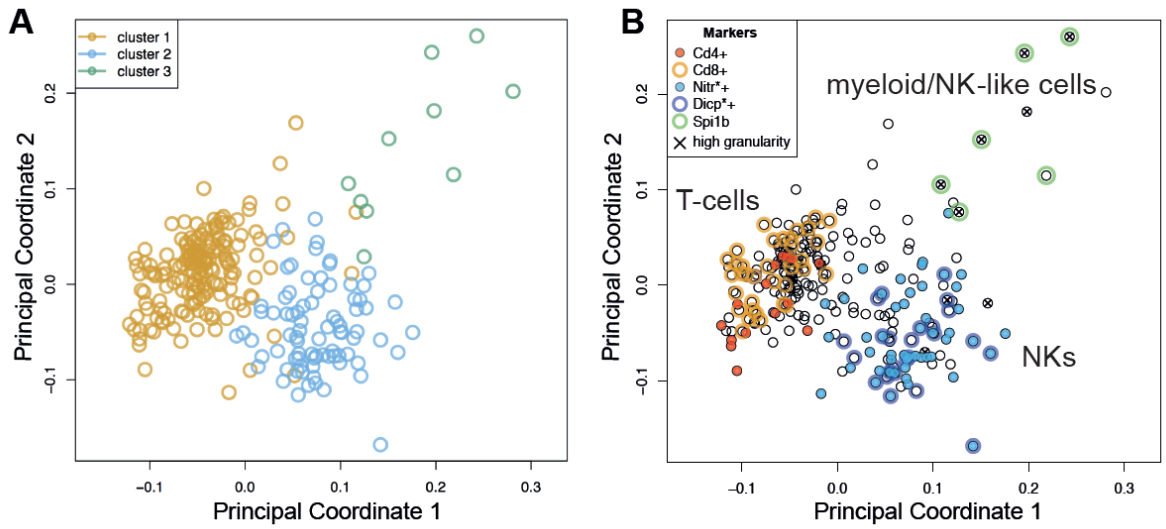

Figure 1. Low-dimensionality projection of single-cell RNA-seq data from zebrafish immune cells. **A** Multi-dimensional scaling (MDS) and clustering of lck+ cells from adult zebrafish spleens reveal three main clusters. **B** MDS projection where expression of specific marker genes revealed T cells, Natural Killers (NKs) and a smaller population of cells with features of both myeloid and NK cells.

scRNA-seq has been used to deconvolute immune cell type heterogeneity by identifying novel distinct immune cell subsets in health and diseases (Papalexi and Satija, 2017; Stubbington et al., 2017). Moreover scRNA-seq has the potential to reconstruct or predict developmental 'trajectories' for immune cells. Developmental trajectories are now routinely profiled with scRNA-seq. Such studies have led to a deeper understanding of the regulation of early myeloid

(Drissen et al., 2016), conventional dendritic cells (Schlitzer et al., 2015) and megakaryocytic differentiation (Psaila et al., 2016), as well as of fate decisions along T helper subtypes commitment during malaria infection (Lönnberg et al., 2017).

Single-cell RNA-seq of immune cells beyond mammals: case studies in zebrafish

Recently, we and others have successfully applied scRNA-seq to characterize immune cells in zebrafish (Athanasiadis et al., 2017; Carmona et al., 2017; Tang et al., 2017).

To this end, we took advantage of a zebrafish transgenic line expressing GFP under the control of the lymphocyte-specific transcription factor *lck* (Langenau et al., 2004). This line was previously thought to be T cell specific, and indeed enabled researchers to track zebrafish T cell development *in vivo*. However, *lck* in mouse and human is expressed in both T and NK cells, where it has an important role in activation by phosphorylating tyrosine residues within the immunoreceptor tyrosine-based activation motifs (ITAM) of the cytoplasmic tails of the TCR-gamma chains and CD3 subunits and of NK activating receptors (Binstadt et al., 1997). Therefore, we speculated that this expression pattern could be conserved in teleosts, and that a larger heterogeneity may exist among *lck*+ zebrafish lymphocytes than previously thought.

To address this question, our collaborators from the Cveijic and Teichmann groups performed scRNA-seq of *lck*+ zebrafish cells (Carmona et al., 2017). Briefly, spleens from adult Tg(*lck*:GFP) zebrafish were dissected and single GFP+ cells were FACS (index) sorted into 96-well plates containing lysing buffer, for subsequent cDNA library preparation. Of note, index sorting allows for parameters recorded on the cytometer to be linked to each sorted cell. Single-cell cDNA libraries were generated using the Smart-seq2 protocol (Picelli et al., 2014) and sequenced using 125-bp paired-end reads on Illumina HiSeq2000.
Gene expression was quantified as Transcripts per Million (TPMs) in each cell using the zebrafish reference transcriptome (Howe et al., 2013) and global similarities between all pairs of cells were computed using Pearson's correlation. Unsupervised clustering analysis suggested the presence of at least 3 distinct cell types (Carmona et al., 2017). Given the high-dimensionality of these data (~12,000 genes x 300 cells), dimensionality reduction methods are needed to aid visual interpretation of the cellular heterogeneity. Multi-Dimensional Scaling (MDS) was applied to the similarity matrix to obtain a low-dimensionality projection (different colors indicate the three clusters found in the original space) (Figure 1A). Remarkably, T cell specific genes such as Cd4 or Cd8 were specifically expressed in only one of the clusters (Figure 1B). Upon analysis of the differentially expressed genes among the three clusters, we found that the second cluster specifically expressed multiple members of the NITR and DICP innate immune receptors (Figure 1B) as well as the antimicrobial peptide NK-lysin 2 (Carmona 2017). Comparative genomics and functional studies suggested that these genes might be expressed by a fish immune cell type equivalent to mammalian natural killer cells (Haire et al., 2012; Pereiro et al., 2015; Yoder et al., 2004) (see also (Moore et al., 2016) for a description of NK-lysin 4 expression based on single-cell qPCR). These results suggest that cells in cluster 2 may correspond to a natural killer-like cell type in zebrafish. Interestingly, this NK signature was

confirmed in a *rag1*-/- knock out *lck*:GFP zebrafish line that fails to develop T cells (Tang et al., 2017), further supporting the identity of this lymphocytic subset.

The third and smaller cell cluster (green points Figure 1A), presented features of both natural killer and myeloid cells, including higher granularity and specific expression of the transcription factor *Spi1b,* the granulocyte/macrophage colony-stimulating factor receptor beta *(csf2rb),* the Src family tyrosine kinase *hck* along with NK lysins 3 and 4, perforin (*prf1.7*) and chemokines *ccl35.1* and *ccl33.3* (Carmona et al., 2017). Interestingly, a very similar population was described (Tang et al., 2017), indicating that this represents likely a novel sub-type of innate immune cells.

Remarkably, by clustering cells based on their transcriptional similarity and annotating the clusters based on the expression of a handful of marker genes with known or predicted function, scRNA-seq allowed us to obtain for the first time the full transcriptomes of T cells and Natural killers in an evolutionary distant vertebrate (Carmona et al., 2017). In addition, these data provided expression information about multiple cell-type specific cytokines, cytokine receptors and other immune-related genes. Evolutionary insights from these data are discussed later.

In larger scale experiments, Langenau and colleagues (Tang et al., 2017) used scRNA-seq to study cell heterogeneity within zebrafish kidney marrow, the equivalent of mammalian bone marrow (BM) and site of larval and adult hematopoiesis. First, they isolated cells from reporter transgenic fish lines and performed Smart-seq2 to obtain gene signatures of defined lineages: Tg(*Runx1+23:*GFP) for hematopoietic stem and progenitor cells (HSPCs) (Tamplin et al., 2015); Tg(*cd41:*GFP), for both HSPCs and thrombocytes (Lin, 2005); Tg(*mpx:*GFP), for neutrophil/myeloid lineages (Mathias et al., 2006; Renshaw et al., 2006), Tg(*rag2:*GFP), for marrow-derived B cells (Page et al., 2013); and Tg(*lck:*GFP), for T and NK cells (Carmona et al., 2017; Langenau et al., 2004; Moore et al., 2016).
Having defined signatures for most hematopoietic cell lineages, the researchers then profiled almost 4,000 unlabeled single cells isolated from the kidney of wild-type zebrafish using InDrops scRNA-seq, obtaining an unbiased landscape of blood and kidney cells. Finally, Tang et al. were able to study immune-cell deficiencies within DNA–protein kinase catalytic subunit (*prkdc*) and interleukin-2 receptor γ a (*il2rga*) mutant fish, identifying blood cell losses in T, B, and natural killer cells. This highlights the power of scRNA-seq to address the impact of specific genetic mutants on different immune cell compartments.

Immune cells develop from multipotent stem cells through the process known as hematopoiesis. The molecular pathways governing hematopoiesis are highly conserved in vertebrates, where most (if not all) critical transcription factors involved in hematopoiesis in mammals have orthologs in fish (Davidson and Zon, 2004). In the classical model of hematopoiesis, an organized hematopoietic lineage tree starts with multipotent hematopoietic stem cells, and is then followed by oligopotent and unipotent progenitors. However, recent single-cell results challenged the classic model and proposed that hematopoietic progenitor cell types are very heterogeneous (Notta et al., 2016; Paul et al., 2015).

In zebrafish, Cvejic and colleagues applied single-cell RNA-seq to study thrombocyte lineage commitment, and were able to reconstruct a continuous lineage differentiation pathway (Macaulay et al., 2016). More recently, their study was extended to reconstruct differentiation trajectories of most hematopoiesis lineages by performing scRNA-seq of kidney-derived blood cells from eight different zebrafish transgenic reporter lines (Athanasiadis et al., 2017). Their results indicated that the haematopoietic program is highly conserved between zebrafish and higher vertebrates.
Altogether these different single-cell RNA-seq studies demonstrate that immune cell types can be identified by clustering cells based on transcriptional similarity in zebrafish and enabled detailed phenotypic characterization of distinct immune cell populations. Although several of these studies took advantage of reporter lines, the increase in throughput of scRNA-seq techniques will make it increasingly feasible to start from unsorted cell populations to decipher the complexity of immune cell types in any vertebrate species.

Envisioning immune cell type identification across vertebrates from blood samples

Building upon the success of these recent single-cell RNA-seq studies in zebrafish, we envision that a powerful approach to identify and characterize immune cells in any vertebrate species could be to perform scRNA-seq on whole blood cells or Peripheral blood mononuclear cells (PBMC)-enriched (e.g. by using Ficoll gradient). ScRNA-seq of PBMCs has allowed for detailed characterization of leucocytes heterogeneity in mammals from fresh (Zheng et al., 2017) and cryopreserved (Guillaumet-Adkins et al., 2017) samples. In species for which sufficient amounts of blood are not readily accessible, like zebrafish, dissection and sampling of immune tissues such as kidney marrow might be a viable alternative to obtain a high proportion of (unlabeled) immune cells (Tang et al., 2017). Different leukocyte isolation methods from peripheral blood have been employed for different fish species (Inoue et al., 2002; Pierrard et al., 2012), indicating that although erythrocytes are nucleated in non-mammalian species, physical separation of red blood cells from other cell types found in blood is feasible in most cases.

In species for which reference genomes or transcriptomes are available for sequence mapping or transcript quantification, high-throughput droplet-based scRNA-seq methods based on 3' counting, such as 10x Genomics or InDrops, represent an attractive approach. For other organisms with limited or lack of genomic information, full-length mRNA scRNA-seq protocols, such as Smart-seq2, enables *de novo* transcriptome assembly using tools such as Trinity (Grabherr et al., 2011) and Velvet (Schulz et al., 2012). RNA-seq followed by *de novo* transcriptome assembly has been successfully conducted in common (Ji et al., 2012) and crucian carp (Liao et al., 2013; Rhee et al., 2014), leading to the identification of hundreds of immune-related genes. Of note, it has been shown that representative assemblies can be generated with as few as 20 million reads from tissue samples (Francis et al., 2013), suggesting that the amount of reads obtained in standard scRNA-seq experiments is likely enough to reconstruct the full transcriptome.

Once reads have been mapped to their respective transcripts, clustering techniques should be used to group cells based on their transcriptional similarity. Representative transcriptome for

different cells types are then obtained by pooling together cells of each cluster. Cluster annotation should be done first by using existing markers derived from studies in mammalian. Given that most immune marker genes are conserved in jawed vertebrates, this approach is likely to work very well and was successfully applied in zebrafish (Carmona et al., 2017; Tang et al., 2017). Recently, we also proposed a similar approach for identifying ILCs in single-cell RNA-seq that relies on gene signatures that are conserved in most jawed vertebrates (Suffiotti et al., 2017). Whether totally new clusters will be left after annotating all those that display similarity with mammalian immune cell types is still unclear. However, it is likely that some clusters will show less striking resemblance to the mammalian counterpart, which can indicate distinct rates of evolution. In our work, for instance, we observed that many genes in the NK-like cell cluster (including predicted NK receptors NITR and DICP genes) were not conserved in mammalian species.

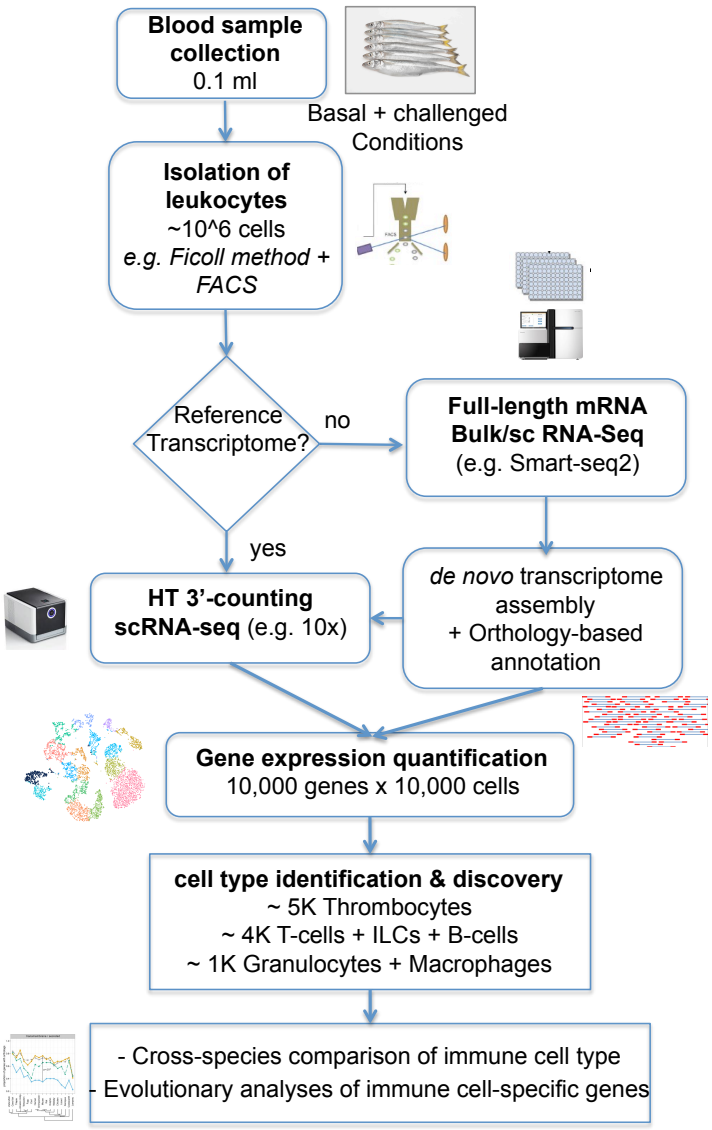

Figure 2. General pipeline for studying immune cell types at the single-cell gene expression level, starting from whole blood samples, in any organism with enough accessible blood.

In Figure 2, we delineate the main steps of this proposed strategy and the number of cells that should be considered as starting material. For reference, in the order of 10^7 leucocytes and thrombocytes per ml of blood can be typically recovered from Asian catfish (Pierrard et al., 2012). Teleost fish typically contain comparable amounts of thrombocytes and leucocytes, and among leucocytes, up to 90% can correspond to lymphocytes (Rey Vázquez and Guerrero, 2007). Therefore, in a plausible scenario, from 100 µl of blood of a fish with typical blood composition, we estimate than one million leucocytes and thrombocytes can be obtained. This amount is 2 orders of magnitude bigger than what is required to obtain 10,000 single-cell transcriptomes using 10x Genomics current technologies (for which the cell recovery rate is about 50% (Zheng et al., 2017). From these 10K cells, we would roughly obtain 5K thrombocytes, 4K lymphocytes and 1K of granulocytes and monocytes. As a reference, in zebrafish 3,000 single-cell transcriptomes obtained from kidney

marrow were sufficient to identify HSCs, thrombocytes, macrophages, neutrophils, different subsets of T cell and Natural killer cells, B cells, and erythrocytes, among others (Tang et al., 2017).
Overall, our estimates indicate that the approach outlined in Figure 2 is likely applicable to species from most vertebrate clades.

Consequences for our understanding of immune cell type evolution, with a focus on lymphocytes

Phylogenetic explorations of the adaptive immune system have relied on the identification of *TCR*, *BCR*, *RAG1*/*RAG2*, *MHC class I* and *class II* genes as the key elements of the adaptive immune system. This search has led to the identification of these genes in all jawed vertebrates that have been carefully examined, including cartilaginous fish, like sharks (Cannon et al., 2004; Cooper and Alder, 2006)

However, the potential for unbiased, marker-free identification of cell types as clusters in the transcriptome space in any organism, opens novel opportunities in comparative immunology, that go far beyond genomic conservation of immune genes.

With current scRNA-seq technologies, it is now virtually possible to ´discover´ *de novo* all of the main immune cell populations by sequencing thousands of cells and clustering them based on transcriptome similarity (Trapnell, 2015). By applying scRNA-seq beyond mammals, for instance, we have identified novel types of natural killer-like cells (Carmona et al., 2017). By anchoring on expected NK features, such as expression of innate immune receptors (e.g. *nitr2a*, *dicp1.1*), cytotoxic molecules (Perforin, granzymes), antimicrobial peptides (*nkl.2* in particular), and lack of T and B cell features, multiple novel NK-specific molecules were identified, including cytokines, cytokine receptors and other immune-related genes. Many of these genes have recently evolved by duplication and neofunctionalization, and do not necessarily have mammalian orthologs (Carmona et al., 2017). Indeed, changes in immune gene expression patterns (a kind of neofunctionalization), are likely to reflect adaptation to different habitats (Yang et al., 2016). Interestingly, zebrafish might represent a particularly attractive model to study NKs and other ILCs, considering the robust immunity of rag1-/- mutants that lack T and B cells, but are yet able to maintain a low mortality rate (Tokunaga et al., 2017).

Finally, once immune cell type-specific genes have been identified, genomic sequence analysis, such as ortholog identification and calculation of non-synonymous to synonymous substitution rates, can inform us about cell type-specific evolutionary rates. This approach was first introduced by ourselves and collaborators when characterizing T and NK cells in zebrafish (Carmona et al., 2017). A striking result from this study was the much lower conservation of trans-membrane NK specific proteins compared to other trans-membrane proteins, including those specific for T cells.

Conclusion

The last decade has witnessed an exponential growth of single-cell RNA-seq data (Svensson et al., 2018), but most of these data came from only a few big academic centers that had invested into this technology since the beginning. With the advent of streamlined protocols such as those available at 10X Genomics or InDrops, it is very likely that scRNA-seq will become available in many more places in the next few years and will be routinely performed to analyze gene expression at the single-cell level in many different tissues and species (Regev et al., 2017).

The use of this technology to characterize immune cell populations from a broad range of vertebrate species following the pipeline proposed in Figure 2 will likely impact our understanding of the evolution of immune cells. First, it could confirm the existence of mammalian immune cell types predicted based on the conservation of known markers in other non-mammalian species. This is likely the first and most important insight that will come from such studies. For instance, it could provide transcriptomic evidence about the conservation of ILCs (Suffiotti et al., 2017; Vivier et al., 2016). Similarly, it will enable to monitor gene expression changes in specific immune cell populations upon infections or other stimuli. Second, it could help predict new markers and study the evolution rate of immune cell types, for instance by assessing the sequence conservation across vertebrate species or dN/dS ratio of genes specifically expressed in a given cell type in a given species (Carmona et al., 2017). Third, it will help studying neo-functionalization events (i.e., genes that are not expressed in the same cell types as their mammalian orthologs). Finally, it may potentially reveal completely novel immune cell types that have no mammalian equivalent, providing similar insights into the forces that drive the evolution of the adaptive immune system as the discovery of the adaptive immune system in jawless vertebrates (Pancer et al., 2004). For these different reasons, we anticipate that, in the next few years, comparative immunology will strongly benefit from technological developments in scRNAseq that so far were mainly applied to mouse and human samples.

References

Afik, S., Yates, K.B., Bi, K., Darko, S., Godec, J., Gerdemann, U., Swadling, L., Douek, D.C., Klenerman, P., Barnes, E.J., et al. (2017). Targeted reconstruction of T cell receptor sequence from single cell RNA-seq links CDR3 length to T cell differentiation state. Nucleic Acids Res. *45*, e148.

Andrews, T.S., and Hemberg, M. (2018). Identifying cell populations with scRNASeq. Mol. Aspects Med. *59*, 114–122.

Athanasiadis, E.I., Botthof, J.G., Andres, H., Ferreira, L., Lio, P., and Cvejic, A. (2017). Single-cell RNA-sequencing uncovers transcriptional states and fate decisions in haematopoiesis. Nat. Commun. *8*.

Binstadt, B.A., Brumbaugh, K.M., and Leibson, P.J. (1997). Signal transduction by human NK cell MHC-recognizing receptors. Immunol. Rev. *155*, 197–203.

Cannon, J.P., Haire, R.N., Rast, J.P., and Litman, G.W. (2004). The phylogenetic origins of the antigen-binding receptors and somatic diversification mechanisms. Immunol. Rev. *200*, 12–22.

Canzar, S., Neu, K.E., Tang, Q., Wilson, P.C., and Khan, A.A. (2016). BASIC: BCR assembly from single cells. Bioinformatics btw631.

Carmona, S.J., Teichmann, S.A., Ferreira, L., Macaulay, I.C., Stubbington, M.J.T., Cvejic, A., and Gfeller, D. (2017). Single-cell transcriptome analysis of fish immune cells provides insight into the evolution of vertebrate immune cell types. Genome Res. *27*, 451–461.

Clark, G., Stockinger, H., Balderas, R., van Zelm, M.C., Zola, H., Hart, D., and Engel, P. (2016). Nomenclature of CD molecules from the Tenth Human Leucocyte Differentiation Antigen Workshop. Clin. Transl. Immunol. *5*, e57.

Cooper, M.D., and Alder, M.N. (2006). The Evolution of Adaptive Immune Systems. Cell *124*, 815–822.

Cooper, M.D., Peterson, R.D., and Good, R.A. (1965). DELINEATION OF THE THYMIC AND BURSAL LYMPHOID SYSTEMS IN THE CHICKEN. Nature *205*, 143–146.

Cooper, M.D., Raymond, D.A., Peterson, R.D., South, M.A., and Good, R.A. (1966). The functions of the thymus system and the bursa system in the chicken. J. Exp. Med. *123*, 75–102.

Davidson, A.J., and Zon, L.I. (2004). The “definitive” (and ’primitive’) guide to zebrafish hematopoiesis. Oncogene *23*, 7233–7246.

Drissen, R., Buza-Vidas, N., Woll, P., Thongjuea, S., Gambardella, A., Giustacchini, A., Mancini, E., Zriwil, A., Lutteropp, M., Grover, A., et al. (2016). Distinct myeloid progenitor–differentiation pathways identified through single-cell RNA sequencing. Nat. Immunol. *17*, 666–676.

Eltahla, A.A., Rizzetto, S., Pirozyan, M.R., Betz-Stablein, B.D., Venturi, V., Kedzierska, K., Lloyd, A.R., Bull, R.A., and Luciani, F. (2016). Linking the T cell receptor to the single cell transcriptome in antigen-specific human T cells. Immunol. Cell Biol. *94*, 604–611.

Flajnik, M.F., and Kasahara, M. (2010). Origin and evolution of the adaptive immune system: genetic events and selective pressures. Nat. Rev. Genet. *11*, 47–59.

Francis, W.R., Christianson, L.M., Kiko, R., Powers, M.L., Shaner, N.C., and D Haddock, S.H. (2013). A comparison across non-model animals suggests an optimal sequencing depth for de novo transcriptome assembly. BMC Genomics *14*, 167.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. *29*, 644–652.

Guillaumet-Adkins, A., Rodríguez-Esteban, G., Mereu, E., Mendez-Lago, M., Jaitin, D.A., Villanueva, A., Vidal, A., Martinez-Marti, A., Felip, E., Vivancos, A., et al. (2017). Single-cell transcriptome conservation in cryopreserved cells and tissues. Genome Biol. *18*.

Haire, R.N., Cannon, J.P., O'Driscoll, M.L., Ostrov, D.A., Mueller, M.G., Turner, P.M., Litman, R.T., Litman, G.W., and Yoder, J.A. (2012). Genomic and functional characterization of the diverse immunoglobulin domain-containing protein (DICP) family. Genomics *99*, 282–291.

Hashimshony, T., Senderovich, N., Avital, G., Klochendler, A., de Leeuw, Y., Anavy, L., Gennert, D., Li, S., Livak, K.J., Rozenblatt-Rosen, O., et al. (2016). CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. Genome Biol. *17*.

Hirano, M., Guo, P., McCurley, N., Schorpp, M., Das, S., Boehm, T., and Cooper, M.D. (2013). Evolutionary implications of a third lymphocyte lineage in lampreys. Nature *501*, 435–438.

Howe, K., Clark, M.D., Torroja, C.F., Torrance, J., Berthelot, C., Muffato, M., Collins, J.E., Humphray, S., McLaren, K., Matthews, L., et al. (2013). The zebrafish reference genome sequence and its relationship to the human genome. Nature *496*, 498–503.

Inoue, T., Moritomo, T., Tamura, Y., Mamiya, S., Fujino, H., and Nakanishi, T. (2002). A new method for fish leucocyte counting and partial differentiation by flow cytometry. Fish Shellfish Immunol. *13*, 379–390.

Islam, S., Zeisel, A., Joost, S., La Manno, G., Zajac, P., Kasper, M., Lönnerberg, P., and Linnarsson, S. (2014). Quantitative single-cell RNA-seq with unique molecular identifiers. Nat. Methods *11*, 163–166.

Jacobson, G., Muncaster, S., Mensink, K., Forlenza, M., Elliot, N., Broomfield, G., Signal, B., and Bird, S. (2017). Omics and cytokine discovery in fish: Presenting the Yellowtail kingfish (Seriola lalandi) as a case study. Dev. Comp. Immunol. *75*, 63–76.

Ji, P., Liu, G., Xu, J., Wang, X., Li, J., Zhao, Z., Zhang, X., Zhang, Y., Xu, P., and Sun, X. (2012). Characterization of Common Carp Transcriptome: Sequencing, De Novo Assembly, Annotation and Comparative Genomics. PLoS ONE *7*, e35152.

Kasheta, M., Painter, C.A., Moore, F.E., Lobbardi, R., Bryll, A., Freiman, E., Stachura, D., Rogers, A.B., Houvras, Y., Langenau, D.M., et al. (2017). Identification and characterization of T reg–like cells in zebrafish. J. Exp. Med. *214*, 3519–3530.

Kiessling, R., Klein, E., and Wigzell, H. (1975). "Natural" killer cells in the mouse. I. Cytotoxic cells with specificity for mouse Moloney leukemia cells. Specificity and distribution according to genotype. Eur. J. Immunol. *5*, 112–117.

Kolodziejczyk, A.A., Kim, J.K., Svensson, V., Marioni, J.C., and Teichmann, S.A. (2015). The Technology and Biology of Single-Cell RNA Sequencing. Mol. Cell *58*, 610–620.

Langenau, D.M., Ferrando, A.A., Traver, D., Kutok, J.L., Hezel, J.-P.D., Kanki, J.P., Zon, L.I., Look, A.T., and Trede, N.S. (2004). In vivo tracking of T cell development, ablation, and engraftment in transgenic zebrafish. Proc. Natl. Acad. Sci. *101*, 7369–7374.

Liao, X., Cheng, L., Xu, P., Lu, G., Wachholtz, M., Sun, X., and Chen, S. (2013). Transcriptome Analysis of Crucian Carp (Carassius auratus), an Important Aquaculture and Hypoxia-Tolerant Species. PLoS ONE *8*, e62308.

Lin, H.-F. (2005). Analysis of thrombocyte development in CD41-GFP transgenic zebrafish. Blood *106*, 3803–3810.

Lönnberg, T., Svensson, V., James, K.R., Fernandez-Ruiz, D., Sebina, I., Montandon, R., Soon, M.S.F., Fogg, L.G., Nair, A.S., Liligeto, U.N., et al. (2017). Single-cell RNA-seq and computational analysis using temporal mixture modeling resolves T_H1/T_{FH} fate bifurcation in malaria. Sci. Immunol. *2*, eaal2192.

Macaulay, I.C., Svensson, V., Labalette, C., Ferreira, L., Hamey, F., Voet, T., Teichmann, S.A., and Cvejic, A. (2016). Single-Cell RNA-Sequencing Reveals a Continuous Spectrum of Differentiation in Hematopoietic Cells. Cell Rep. *14*, 966–977.

Martin, S.A.M., and Król, E. (2017). Nutrigenomics and immune function in fish: new insights from omics technologies. Dev. Comp. Immunol. *75*, 86–98.

Mathias, J.R., Perrin, B.J., Liu, T.-X., Kanki, J., Look, A.T., and Huttenlocher, A. (2006). Resolution of inflammation by retrograde chemotaxis of neutrophils in transgenic zebrafish. J. Leukoc. Biol. *80*, 1281–1288.

Mebius, R.E., Rennert, P., and Weissman, I.L. (1997). Developing lymph nodes collect CD4+CD3- LTbeta+ cells that can differentiate to APC, NK cells, and follicular cells but not T or B cells. Immunity *7*, 493–504.

Moore, F.E., Garcia, E.G., Lobbardi, R., Jain, E., Tang, Q., Moore, J.C., Cortes, M., Molodtsov, A., Kasheta, M., Luo, C.C., et al. (2016). Single-cell transcriptional analysis of normal, aberrant, and malignant hematopoiesis in zebrafish. J. Exp. Med. *213*, 979–992.

Morales Poole, J.R., Paganini, J., and Pontarotti, P. (2017). Convergent evolution of the adaptive immune response in jawed vertebrates and cyclostomes: An evolutionary biology approach based study. Dev. Comp. Immunol. *75*, 120–126.

Mora-Velandia, L.M., Castro-Escamilla, O., Méndez, A.G., Aguilar-Flores, C., Velázquez-Avila, M., Tussié-Luna, M.I., Téllez-Sosa, J., Maldonado-García, C., Jurado-Santacruz, F., Ferat-Osorio, E., et al. (2017). A Human Lin-CD123+CD127lowPopulation Endowed with ILC Features and Migratory Capabilities Contributes to Immunopathological Hallmarks of Psoriasis. Front. Immunol. *8*, 176.

Notta, F., Zandi, S., Takayama, N., Dobson, S., Gan, O.I., Wilson, G., Kaufmann, K.B., McLeod, J., Laurenti, E., Dunant, C.F., et al. (2016). Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. Science *351*, aab2116-aab2116.

Page, D.M., Wittamer, V., Bertrand, J.Y., Lewis, K.L., Pratt, D.N., Delgado, N., Schale, S.E., McGue, C., Jacobsen, B.H., Doty, A., et al. (2013). An evolutionarily conserved program of B-cell development and activation in zebrafish. Blood *122*, e1–e11.

Pancer, Z., Amemiya, C.T., Ehrhardt, G.R.A., Ceitlin, J., Larry Gartland, G., and Cooper, M.D. (2004). Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey. Nature *430*, 174–180.

Papalexi, E., and Satija, R. (2017). Single-cell RNA sequencing to explore immune cell heterogeneity. Nat. Rev. Immunol. *18*, 35–45.

Paul, F., Arkin, Y., Giladi, A., Jaitin, D.A., Kenigsberg, E., Keren-Shaul, H., Winter, D., Lara-Astiaso, D., Gury, M., Weiner, A., et al. (2015). Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors. Cell *163*, 1663–1677.

Pereiro, P., Varela, M., Diaz-Rosales, P., Romero, A., Dios, S., Figueras, A., and Novoa, B. (2015). Zebrafish Nk-lysins: First insights about their cellular and functional diversification. Dev. Comp. Immunol. *51*, 148–159.

Picelli, S., Faridani, O.R., Björklund, Å.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. Nat. Protoc. *9*, 171–181.

Pierrard, M.-A., Roland, K., Kestemont, P., Dieu, M., Raes, M., and Silvestre, F. (2012). Fish peripheral blood mononuclear cells preparation for future monitoring applications. Anal. Biochem. *426*, 153–165.

Psaila, B., Barkas, N., Iskander, D., Roy, A., Anderson, S., Ashley, N., Caputo, V.S., Lichtenberg, J., Loaiza, S., Bodine, D.M., et al. (2016). Single-cell profiling of human megakaryocyte-erythroid progenitors identifies distinct megakaryocyte and erythroid differentiation pathways. Genome Biol. *17*.

Raj, A., and van Oudenaarden, A. (2008). Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. Cell *135*, 216–226.

Regev, A., Teichmann, S.A., Lander, E.S., Amit, I., Benoist, C., Birney, E., Bodenmiller, B., Campbell, P., Carninci, P., Clatworthy, M., et al. (2017). The Human Cell Atlas. ELife *6*.

Renshaw, S.A., and Trede, N.S. (2012). A model 450 million years in the making: zebrafish and vertebrate immunity. Dis. Model. Mech. *5*, 38–47.

Renshaw, S.A., Loynes, C.A., Trushell, D.M.I., Elworthy, S., Ingham, P.W., and Whyte, M.K.B. (2006). A transgenic zebrafish model of neutrophilic inflammation. Blood *108*, 3976–3978.

Rey Vázquez, G., and Guerrero, G.A. (2007). Characterization of blood cells and hematological parameters in Cichlasoma dimerus (Teleostei, Perciformes). Tissue Cell *39*, 151–160.

Rhee, J.-S., Jeong, C.-B., Kim, D.-H., Kim, I.-C., Lee, Y.S., Lee, C., and Lee, J.-S. (2014). Immune gene discovery in the crucian carp Carassius auratus. Fish Shellfish Immunol. *36*, 240–251.

Rougeot, J., Zakrzewska, A., Kanwal, Z., Jansen, H.J., Spaink, H.P., and Meijer, A.H. (2014). RNA sequencing of FACS-sorted immune cell populations from zebrafish infection models to identify cell specific responses to intracellular pathogens. Methods Mol. Biol. Clifton NJ *1197*, 261–274.

Saraceni, P.R., Romero, A., Figueras, A., and Novoa, B. (2016). Establishment of Infection Models in Zebrafish Larvae (Danio rerio) to Study the Pathogenesis of Aeromonas hydrophila. Front. Microbiol. *7*, 1219.

Schlitzer, A., Sivakamasundari, V., Chen, J., Sumatoh, H.R.B., Schreuder, J., Lum, J., Malleret, B., Zhang, S., Larbi, A., Zolezzi, F., et al. (2015). Identification of cDC1- and cDC2-committed DC progenitors reveals early lineage priming at the common DC progenitor stage in the bone marrow. Nat. Immunol. *16*, 718–728.

Schulz, M.H., Zerbino, D.R., Vingron, M., and Birney, E. (2012). Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinforma. Oxf. Engl. *28*, 1086–1092.

Spits, H., Artis, D., Colonna, M., Diefenbach, A., Di Santo, J.P., Eberl, G., Koyasu, S., Locksley, R.M., McKenzie, A.N.J., Mebius, R.E., et al. (2013). Innate lymphoid cells--a proposal for uniform nomenclature. Nat. Rev. Immunol. *13*, 145–149.

Stubbington, M.J.T., Lönnberg, T., Proserpio, V., Clare, S., Speak, A.O., Dougan, G., and Teichmann, S.A. (2016). T cell fate and clonality inference from single-cell transcriptomes. Nat. Methods *13*, 329–332.

Stubbington, M.J.T., Rozenblatt-Rosen, O., Regev, A., and Teichmann, S.A. (2017). Single-cell transcriptomics to explore the immune system in health and disease. Science *358*, 58–63.

Sudhagar, A., Kumar, G., and El-Matbouli, M. (2018). Transcriptome Analysis Based on RNA-Seq in Understanding Pathogenic Mechanisms of Diseases and the Immune System of Fish: A Comprehensive Review. Int. J. Mol. Sci. *19*.

Suffiotti, M., Carmona, S.J., Jandus, C., and Gfeller, D. (2017). Identification of innate lymphoid cells in single-cell RNA-Seq data. Immunogenetics *69*, 439–450.

Svensson, V., Vento-Tormo, R., and Teichmann, S.A. (2018). Exponential scaling of single-cell RNA-seq in the past decade. Nat. Protoc. *13*, 599–604.

Tamplin, O.J., Durand, E.M., Carr, L.A., Childs, S.J., Hagedorn, E.J., Li, P., Yzaguirre, A.D., Speck, N.A., and Zon, L.I. (2015). Hematopoietic stem cell arrival triggers dynamic remodeling of the perivascular niche. Cell *160*, 241–252.

Tang, Q., Iyer, S., Lobbardi, R., Moore, J.C., Chen, H., Lareau, C., Hebert, C., Shaw, M.L., Neftel, C., Suva, M.L., et al. (2017). Dissecting hematopoietic and renal cell heterogeneity in adult zebrafish at single-cell resolution using RNA sequencing. J. Exp. Med. *214*, 2875–2887.

Tokunaga, Y., Shirouzu, M., Sugahara, R., Yoshiura, Y., Kiryu, I., Ototake, M., Nagasawa, T., Somamoto, T., and Nakao, M. (2017). Comprehensive validation of T- and B-cell deficiency in rag1-null zebrafish: Implication for the robust innate defense mechanisms of teleosts. Sci. Rep. *7*.

Trapnell, C. (2015). Defining cell types and states with single-cell genomics. Genome Res. *25*, 1491–1498.

Traver, D., Paw, B.H., Poss, K.D., Penberthy, W.T., Lin, S., and Zon, L.I. (2003). Transplantation and in vivo imaging of multilineage engraftment in zebrafish bloodless mutants. Nat. Immunol. *4*, 1238–1246.

Udvadia, A.J., and Linney, E. (2003). Windows into development: historic, current, and future perspectives on transgenic zebrafish. Dev. Biol. *256*, 1–17.

Vivier, E., van de Pavert, S.A., Cooper, M.D., and Belz, G.T. (2016). The evolution of innate lymphoid cells. Nat. Immunol. *17*, 790–794.

Warner, N.L., Szenberg, A., and Burnet, F.M. (1962). The immunological role of different lymphoid organs in the chicken. I. Dissociation of immunological responsiveness. Aust. J. Exp. Biol. Med. Sci. *40*, 373–387.

Yang, J., Chen, X., Bai, J., Fang, D., Qiu, Y., Jiang, W., Yuan, H., Bian, C., Lu, J., He, S., et al. (2016). The Sinocyclocheilus cavefish genome provides insights into cave adaptation. BMC Biol. *14*, 1.

Yoder, J.A., Litman, R.T., Mueller, M.G., Desai, S., Dobrinski, K.P., Montgomery, J.S., Buzzeo, M.P., Ota, T., Amemiya, C.T., Trede, N.S., et al. (2004). Resolution of the novel immune-type receptor gene cluster in zebrafish. Proc. Natl. Acad. Sci. *101*, 15706–15711.

Zapata, A., and Amemiya, C.T. (2000). Phylogeny of lower vertebrates and their immunological structures. Curr. Top. Microbiol. Immunol. *248*, 67–107.

Zheng, G.X.Y., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., et al. (2017). Massively parallel digital transcriptional profiling of single cells. Nat. Commun. *8*, 14049.

Zilionis, R., Nainys, J., Veres, A., Savova, V., Zemmour, D., Klein, A.M., and Mazutis, L. (2016). Single-cell barcoding and sequencing using droplet microfluidics. Nat. Protoc. *12*, 44–73.

(1996). Fish physiology. Vol. 15: The fish immune system: organism, pathogen and environment (New York: Acad. Press).