**Unveiling Immune Complexity and Regenerative Dynamics in *Pleurodeles waltl*: A Cross-Species Single-Cell Analysis**

Course: **BINP51,45 credits**

Student: **Nikhilesh Vasantha Kumar**

Email : ni2651va-s@student.lu.se

Supervisor: **Nicholas Leigh**

Email : nicholas.leigh@med.lu.se

**Lund University 2024**

# **Abstract**

This study examines the immune cell repertoire of *Pleurodeles waltl*, an amphibian known for its exceptional regenerative capabilities, by comparing its immune system with those of *Homo sapiens* and *Xenopus laevis*. The focus of our analysis centres on adaptive immune cells, which play a crucial role in immune modulation to facilitate regeneration. We explored Differential Expression (DE) Analysis, Gene-Ontology (GO) enrichment analysis, and cross species mapping to try and identify various subsets of B and T cells in salamanders. Notably, our research has potentially uncovered Regulatory T cells, previously uncharacterised in these organisms. Our comparative analysis using the SAMap tool across *Pleurodeles*, *Homo sapiens*, and *Xenopus* *laevis* highlights evolutionary differences in immune cell clusters and facilitates an understanding of cross-species immunological relationships. The comparative approach has provided information on potential presence and absence of immune cells in salamanders and sets the stage for experimental follow-up. More broadly, this will facilitate a basic understanding of salamander immunity which will in turn inform on how the immune system impacts their regenerative capacity.

# 1.**Introduction**

## **1.1 Differences in regenerative capabilities in humans and salamanders**

Salamanders such as *Pleurodeles waltl* have excellent regenerative capabilities, which is in stark contrast to the limited regenerative abilities observed in humans. Salamanders can regenerate complex tissues and organs such as limbs, tail, tissues of the eye, and the heart (Joven et al., 2019). Human complex tissue regeneration is confined to rudimentary healing processes that often fail to restore full form and function (Gurtner et al., 2008).

## **1.2 Effects of the Immune system on regeneration.**

Both the innate and adaptive immune systems play a key role in the regenerative process. Systemic depletion of macrophages in axolotl during early blastema development prior to amputation leads to wound healing in axolotl, but not limb regeneration (Godwin et al., 2013). Notably, the presence of highly functional adaptive immunity has been found to have a negative corelation with an organism’s regenerative capacity (Julier et al., 2017). Therefore, understanding the roles of different immune cells specifically involved in adaptive immunity and their effects on the regenerative process plays a key role in understanding the differences in the regenerative capacity of different species, especially humans. (Menger et al., 2010). Salamanders are notably immunodeficient, and it has been suggested that this immunodeficiency is a key to their regenerative success (Bolaños-Castro et al., 2021).

## **1.3 Cross-species comparisons to analyse the immune cells involved in Regeneration.**

Evolution serves as a significant force shaping the characteristics of organisms, impacting not only their morphology, but also broadly influencing the cellular composition and gene expression profile of cells. While we know an immense amount about the immune systems of mammals (i.e., poor regenerators), we have scarce information on the immune systems of highly regenerative organisms. Traditional sequence-based methods of comparisons, such as BLAST, might lose critical information due to the swift evolution of immune genes. Therefore, ortholog-based approaches like Eggnog Mapper, which utilises precomputed orthologous groups and phylogenies from a database are essential to identify orthologous genes across species (Huerta-Cepas et al., 2019). This methodology ensures an accurate comparison of immune function-related genes and captures the evolutionary relationship and functional conservation of these genes across diverse species (Frazer et al., 2003). The choice to compare immune populations in *Homo sapiens* and *Pleurodeles* is due to the difference in regenerative capabilities between them. By examining both - the gene regulatory pathways and the immune populations within the immune systems of these species using scRNA-seq data, we can potentially uncover the immunological factors that contribute to regeneration in salamanders, while highlighting potential barriers to regeneration in humans. *Xenopus laevis* was also used for the comparison with *Pleurodeles* because of the close evolutionary relationship between them and the availability of better database resources to aid cross-species comparisons. Understanding these unique immune characteristics could reveal why adaptive immunity might negatively impact regeneration (Shanley et al., 2021), offering new perspectives on the mechanisms behind the regenerative processes of species with high regenerative capacity.

**2.Materials and Methods**

## **2.1 Collection of Data and Pre-processing**

### **2.1.1 Data Collection**

The raw gene expression count matrices of the spleen samples from *Pleurodeles waltl* were obtained from a previous research study conducted evaluating the genetic demultiplexing of the scRNA sequencing data (Cardiello et al., 2023). Additionally, the reference genome used for aligning and annotating the scRNA-seq data was also obtained from a previous study (Brown et al., 2022). The scRNA-seq dataset consisted of five biological replicates, which were divided into two count matrices. The first count matrix contained data from three replicates, comprising 19,045 cells with 65597 genes, while the second matrix contained data from two replicates, totalling 20,152 cells with 65597 genes. The *Homo sapiens* dataset used is the widely used Tabula sapiens immune dataset which contains 264,824 cells with proper cell type annotations for all the clusters. The *Xenopus laevis* dataset was taken from the work published by (Jiao et al., 2024) which contains 12,411 cells from four different biological replicates.

### **2.1.2 Pre-processing**

To comprehensively annotate the genes identified in the *Pleurodeles* dataset, we utilised the Trinotate tool (Bryant et al., 2017). Trinotate streamlines the annotation process by integrating various annotation tools and databases, including BLASTX, HMMER, and PFAM, to identify homologous protein sequences, protein domains, and functional motifs. Additionally, Trinotate utilises data from public databases such as Swiss-Prot, EggNOG, and Gene Ontology (GO) to provide comprehensive functional and gene annotations for the identified genes in *Pleurodeles waltl.*

#### **2.1.3 CellBender**

We opted to utilise the CellBender tool for both the *Pleurodeles* datasets following a warning flagged by CellRanger, indicating a potential presence of high ambient RNA in the first scRNA-seq dataset. CellBender was chosen for its capability to effectively resolve technical artifacts and biases, particularly ambient RNA contamination, which can adversely affect data quality and downstream analyses (Fleming et al., 2023).

### **2.1.4 Dataset Integration and Filtering**

We filtered both the datasets to remove cells with a higher percentage of mitochondrial genes, we calculated the percentage of mitochondrial genes using a set of common mitochondrial genes as reference found in the mt-genome of *Pleurodeles* *waltl* (Brown et al., 2022) *.* Cells exceeding 10% threshold of mitochondrial gene expression were excluded from further analysis to mitigate potential biases introduced by cellular stress or damage. Additionally, cells with fewer than 400 genes detected were also removed from the datasets.

## **2.2 Dimensionality Reduction and Clustering**

We performed log-normalisation over the datasets and scaled them to reduce the cell-to-cell variance in gene expression levels. Dimensionality reduction was then conducted using Principal Component Analysis (PCA) to identify the principal components, which were key in identifying cell neighbours. 30 principal components were utilised for clustering which were determined by identifying the variance between them to select for the most informative ones. Finally, the Leiden algorithm was used for clustering, grouping cells with similar expression patterns.

**2.3 Dataset Integration using Harmony.**

After identifying the clusters and applying stringent quality-control within the *Pleurodeles waltl* datasets, we proceeded to integrate these datasets. The integration was based on common genes shared between the datasets, and Harmony was employed for this purpose, particularly focusing on biological replicates and the datasets being covariates for the batch-correction step. This iterative adjustment ensures that the final integrated dataset reflects true biological signals (Antonsson & Melsted, 2024)rather than technical variations which is crucial for cross species evaluation and immune population identification.

## **2.4 Differential Expression Analysis**

Following the integration and batch correction of the *Pleurodeles waltl* datasets using Harmony, we conducted a differential expression analysis to identify genes that show significant differences in expression across the identified cell populations. For this analysis, we employed the Wilcoxon rank-sum test, a non-parametric statistical test that compares the distributions of gene expression values between two groups of cells. It is acknowledged within the field that the pseudobulk approach, which aggregates single-cell data into pseudo-bulk samples per condition, often provides enhanced statistical power due to increased sample size per group. This method is particularly advantageous when comparing gene expression across different conditions or treatments. However, given the focus of our study on identifying unique or prominently upregulated genes within specific clusters rather than across conditions, the Wilcoxon rank-sum test was deemed a more appropriate approach. This approach allowed us to identify cell populations that might be immune cell populations based on uniquely expressed or significantly upregulated in specific clusters in the spleen of *Pleurodeles waltl*.

## **2.5 Gene-Ontology Enrichment Analysis**

Upon identifying potential adaptive immune clusters, including T cells and B cells, within the integrated *Pleurodeles waltl* spleen dataset, we further refined our analysis to explore the immune sub-populations of these cells. Targeted re-clustering of these adaptive immune cells was employed to identify the nuanced sub-types present and their respective gene expression profiles. Crucially, this stage of the study involved extensive preliminary work due to the lack of readily available Gene Ontology (GO) databases for *Pleurodeles waltl*. We undertook the substantial task of manually annotating the genes identified in our datasets, ensuring their reliable mapping to meaningful GO terms. Adapting GOATOOLS for use with a species not commonly linked to existing bioinformatics tools posed a significant challenge in performing GO Enrichment Analysis. Overcoming these obstacles not only advanced our understanding of the *Pleurodeles waltl* immune system but also demonstrated the robustness of our methodological approach in adapting and applying advanced bioinformatics tools in non-model organisms.

**2.6 SAMap Analysis**

We utilised the Self-Assembling Manifold mapping (SAMap) analysis to align and compare single-cell RNA sequencing datasets from *Homo sapiens*, *Pleurodeles waltl*, and *Xenopus laevis*. The SAMap algorithm was applied to the integrated dataset using the Leiden cluster information, facilitating a cross-species comparison and integration of cellular landscapes. The integration and alignment process results in a joint manifold that represents a unified view of cellular clusters across species. This allows us to identify and explore conserved cellular clusters offering insights into functional similarities and differences between the species.

# 3.Results

## **3.1 Ensuring Data Integrity through Preprocessing and Quality Control**

Our analysis initiated with a detailed evaluation of data quality from spleen samples of *Pleurodeles waltl*, comprising raw scRNA-seq datasets with 39,197 cells and 65,597 genes each. Ambient RNA contamination was filtered using a trained CellBender model set at a 1.0% false positive rate [FPR] (Fig 1A & Fig 1B) efficiently filtered out 32.96% of counts (Fig 1C) in non-empty droplets for the first dataset and 4.31% of the non-cellular counts (Fig 1D) in the second dataset. We then filtered out the doublets using scvi-tools and souporcell, filtering out 5,179 cells in the first dataset and 3,216 cells in the second dataset. (Fig 1E & 1F). Additionally, cells with more than 10% of mitochondrial genes or having less than 400 genes and genes in less than three cells were filtered out for further analysis.

A screenshot of a graph

Description automatically generated

**Fig 1. Cellbender Model Training and Doublets Quality Control** (A)& (B) The Evidence Lower Bound (ELBO) is plotted against the number of epochs for both the training and test datasets, showcasing the optimisation of our model parameters over time, indicating stable convergence and suggesting adequate training without overfitting for dataset 1 and dataset 2, respectively. (C) & (D) Cell Probability Plot indicating the determination of barcodes that contain cells versus empty droplets for dataset 1 and 2 respectively. The plot displays UMI counts against barcode indices, with a fitted curve distinguishing the true signal (cell-containing barcodes) from the background noise (empty droplets), based on the cell probability estimates. (E) & (F) Venn diagram of the doublets predicted by the SCVI-model and souporcell algorithm based on barcodes for dataset 1 and 2, respectively.

## **3.2 Elucidating Cellular Heterogeneity: Dimensionality Reduction and UMAP Clustering Analyses and batch correction for biological replicates.**

A collage of images of different colors

Description automatically generatedFor the identification of cellular and specifically immune cell clusters, we employed principal component analysis (PCA). Following PCA, we utilised the Uniform Manifold Approximation and Projection (UMAP) for visualisation and the Leiden algorithm (Fig 2A & Fig 2B) for the actual clustering while keeping the neighbourhood size at 30 cells to calculate the neighbourhood graph during the clustering process. Figure 2C and 2D illustrate a notable pattern where cells from the same biological replicate tend to cluster together, suggesting a potential batch effect.

**Fig 2. Visualisation of *Pleurodeles* Cellular Clusters using UMAP for the spleen datasets** (A)&(B) shows the clustering information according to Leiden algorithm for dataset 1 and dataset 2 respectively. (C) & (D) shows the clustering patterns observed in each biological replicate in the respective dataset.

To address the batch effects observed, we initially integrated the datasets based on shared gene names. We then applied the Harmony algorithm for batch correction, considering both dataset origins and biological replicates as covariates. This integrated dataset exhibited improved clustering patterns when taking into context the biological replicates, as demonstrated in Figures 3A and 3B.

**A collage of images of different colors

Description automatically generatedFig 3. Clustering of the Integrated dataset after Harmony Batch-Correction** Clustering of the integrated dataset after batch correction shows less bias towards biological replicates(A) Clustering of the integrated dataset coloured according to biological replicates.(B) Clustering of the integrated dataset labelled according to the Leiden clustering algorithm.

## **3.4 Annotating Genes and Identifying Immune Clusters.**

Following the clustering of the integrated dataset, we utilised orthologous mappings which are eggnog annotations provided by the Trinotate tool to annotate the genes. Consequently, each gene annotation required the integration of sequence ortholog inferences derived from EggNOG-mapper accounting for the evolutionary lineage of *Pleurodeles waltl*. Utilising these annotated genes and corroborating with existing literature, we tried to pinpoint the adaptive immune clusters within the integrated dataset (Fig 4).

A collage of images of different colors

Description automatically generated**Fig 4. Expression of prominent Adaptive Immune Markers in the integrated dataset clusters.**

Plotting the expression of labelled adaptive immune marker genes on the dataset gives us information on the clustering of these cell groups. (A)*Cd79a* (B) *Cd79b* (C) *Cd3e* (D) *Cd4* (E) *Cd8a*. (F) Leiden clustering information of the Integrated dataset.

Specifically, we examined the expression of notable B cell markers like *Cd79a* (Fig 4A) and *Cd79b* (Fig 4B) referencing key studies from (Nemazee & al., 2000; Reth, 2001) which are being expressed in clusters 0,1,2,3,5,7 and 16.Also T cell markers like *Cd3e* (Fig 4C) (Smith-Garvin et al., 2009), *Cd4* (Fig 4D) (Luckheeram et al., 2012; Zhu & Paul, 2008) and *CD8a* (Fig 4E) (Farber et al., 2014) are also being expressed in the clusters 4 and 5 .Although, there is a very minimal expression of *Cd4* in comparison to *Cd8* indicating a higher proportion of CD8+ T cells. Leveraging the clustering insights and expression information for B and T cells, we further subset these populations to identify and characterise their respective subclusters.

## **3.5 Identifying Immune subclusters and Differential Expression Analysis**

Building on the insights gleaned from the expression of marker genes, we segmented the clusters to delve into the underlying differences within the adaptive immune B and T cell populations of *Pleurodeles waltl*. For B cells, we sought to identify subclusters that could represent various stages of B cell maturation or differentiation, such as naive, memory, and plasma cells, each characterised by unique gene expression patterns. Similarly, for T cells, our analysis was geared towards distinguishing between different markers and functional states, such as effector, memory, and regulatory T cells. The detailed characterisation of these cell types would enrich our understanding of the immune mechanisms potentially linked to the regenerative prowess *of Pleurodeles waltl*, laying the groundwork for future research into regenerative processes.

**3.5.1 B cell subsets and Differential Expression Analysis**

A screenshot of a computer generated image

Description automatically generatedTo validate the B-cell subsets delineated from the integrated dataset, we examined the expression patterns of selected marker genes (Fig 5A-5E). Notably, we observed the expression of *Cd3e*, a gene not typically associated with B cells across our clusters (Fig 5D). This unexpected expression of a T-cell marker in B-cell clusters warranted a deeper investigation into the presence of this gene in our B-cell clusters. For instance, we noted the expression of *Cd22*, which is integral to the regulation of B cell functions. Moreover, we detected haemoglobin-related genes, such as *Hba1* and *Hba2*, typically linked to erythroid lineage which were also expressing *Cd79a* and *Cd79b*. We also identified elevated levels of ribosomal protein genes, like *Rpl10* and *Rpl18*, indicative of active protein synthesis which could imply heightened cellular activity within certain subclusters (Supplementary Fig 1). This comparison was not just limited to identifying uniquely expressed genes but extended to understanding the functional implications of these expression patterns in the subsequent analysis.

**Fig 5. Visualising B cell sub-cluster expression and differentially expressed genes.**

**(A)-(E)** These figures display UMAP (Uniform Manifold Approximation and Projection) plots. Each point on the UMAP plots represents a single cell, coloured by expression levels of specific genes: *Ptprc*(A*), Cd79a* (B), *Cd79b* (C), and *Cd3e* (D) and Leiden algorithm (E). The gradient from red to purple signifies expression levels from high to low.

**3.5.2 T cell subsets and Differential Expression Analysis**

To investigate the T-cell subsets identified from our integrated dataset, we analysed the expression of key T-cell markers (Fig 6A-6E). Our findings were particularly striking when we observed the expression of *Il2rb* (Figure 6D), an indicator for regulatory T cells(Tregs) which are cells known to play a documented role in tissue regeneration (Li et al., 2018a). In addition to these findings, we have identified markers implicated in T-cell signaling, such as PLD1, and protein synthesis, exemplified by TRNK1 (Supplementary Fig 2). To further investigate the role of these subclusters in both tissue regeneration and immune response, our subsequent step involved conducting gene ontology enrichment analysis for the identified subsets of T-cells and B-cells.

A group of images of a rainbow colored swirl

Description automatically generated with medium confidence**Fig 6 . Differential Expression Analysis and Expression profile of prominent T cell markers**

**(A)-(E)** These figures display UMAP (Uniform Manifold Approximation and Projection) plots. Each point on the UMAP plots represents a single cell, coloured by expression levels of specific genes: *Cd3e* (A), *Cd4* (B), *Cd8a* (C), and *Il2rb* (D) and Leiden algorithm (E). The gradient from yellow to purple signifies expression levels from high to low.

## **3.6 Gene-ontology Enrichment Analysis for B cell subclusters**

Building upon the differential expression profiles of B and T cell subclusters, we proceeded with a Gene-Ontology (GO) Enrichment Analysis which helps us uncover the broader biological contexts of the gene expression changes we observed and to gain insight into the pathways enriched in these B and T cell subclusters.

To construct a comprehensive GO framework applicable to *Pleurodeles waltl*, we generated a directed acyclic graph (DAG) of GO terms using EggNog GO annotation mappings. This tailored *Pleurodeles-*specific GO DAG provided a structured, hierarchical representation of the functional annotations that are relevant to *Pleurodeles waltl*, allowing for a more precise and meaningful enrichment analysis.

The Gene-Ontology Enrichment Analysis for B cell subclusters provided insights into the functional characteristics and potential roles of these cells in *Pleurodeles waltl.* Across the clusters, we observed enriched GO terms related to fundamental biological processes such as cellular metabolic processes, RNA binding, and translation, indicating a high level of cellular activity and gene regulation within these immune cells. Clusters demonstrated unique signatures, such as Cluster 9's association with the cellular response to cold and poly-pyrimidine tract binding, suggesting a specialisation that could be linked to environmental adaptation or stress responses (Butler & Hannapel, 2012) (Fig 7C) . Cell stress has also been shown to upregulate *Cd3e* in B cells which may explain the expression of *Cd3e* in our B cell clusters (Nagel et al., 2014).Significantly, terms related to translational initiation and ribosomal activity, as seen in Clusters 0 and 4 (Fig 7A & 7B), point to active protein synthesis, possibly associated with antibody production or response to antigenic stimulation. This is in line with the known role of B cells in production of immunoglobulins. Cluster 11's profile is particularly intriguing (Fig 7D), with an enrichment in terms connected to leukocyte-mediated immunity and vesicle-related processes, possibly reflecting a role in the active secretion of immunomodulatory molecules or in the mechanisms of antigen processing and presentation (Rastogi et al., 2022).Moreover, some clusters exhibited enrichment in terms related to cellular response to stimulus and immune system processes, which aligns well with the expected activities of B cells in immune defence mechanisms (Fig 7D).

Several blue and white bars

Description automatically generated**Fig7 . Enriched Gene Ontology of prominent B cell subclusters** Depicts information of the predominant biological processes, cellular components, and molecular functions associated with each B cell subcluster. The x axis determines the negative logarithm of adjusted P value and the colour bar shaded from red to blue indicate the number of genes mapped to each GO term.

## **3.7 Gene-ontology Enrichment Analysis for T cell subclusters**

Similarly in T cells, we see a significant enrichment of GO terms associated with immune response, cell communication, and activation pathways, indicating the active involvement of the immune system. Clusters such as 0, 1, 3, and 4 emphasise various aspects of T cell functionality, from activation and differentiation to the regulation of immune responses. These clusters may represent different T cell subsets, each playing unique roles in the immune landscape of *Pleurodeles waltl*. For instance, terms related to lymphocyte activation and immune system processes are prevalent in cluster 0, aligning with the expected functional profiles of effector T cells, which are key players in immune defence. Cluster 1 shows a pattern consistent with T cell differentiation, possibly representing a transitional state towards a specific T cell lineage (Fig 8A) .Clusters 3 & 4 (Fig 8B & 8C ) feature a variety of genes associated with immune system regulation, suggesting a mixture of T cell subsets, including those akin to memory T cells, which are crucial for long-term immune surveillance and rapid response upon re-exposure to pathogens. Specifically in cluster 7 (Fig 8D) the prominence of GO terms related to cell communication and signalling suggests the management of complex immune interactions, which are characteristic of regulatory T cells (Tregs). These cells are known to modulate immune responses, maintain tolerance to self-antigens, and prevent autoimmune diseases (Kondělková et al., 2010) . Moreover, Tregs have been implicated in tissue repair and regeneration, as they can create an anti-inflammatory environment conducive to healing (Li et al., 2018b). Furthermore, the specific mention of "MHC class I receptor activity" within this cluster could point towards the interaction of these T cells with a wide variety of cell types, including non-immune cells, a critical feature for maintaining tissue integrity and promoting regenerative outcomes (Kikuchi, 2020).However, the GO terms suggested here could also be missing some information due to the lack of a proper mapping dataset. Despite these challenges, our approach has still provided valuable insights and a preliminary understanding of T cell and B cell sub-populations within *Pleurodeles waltl*. Recognizing the limitations of our current methods, it is necessary to conduct further targeted analyses to potentially uncover novel T cell and B cell subsets and their functions.

A close-up of several blue and white bars

Description automatically generated**Fig 8 . Enriched Gene Ontology of prominent T cell subclusters** Depicts information of the predominant biological processes, cellular components, and molecular functions associated with each T cell subcluster. The x axis determines the negative logarithm of adjusted P value and the colour bar shaded from red to blue indicate the number of genes mapped to each GO term.

## **3.7 Cross Species Comparisons of *Pleurodeles waltl* and identifying Immune Clusters**

Finally, to explore another means to annotate immune cell clusters identified in *Pleurodeles waltl* we conducted a cross-species comparative analysis. To do these, we compared our findings with data from *Homo sapiens* and *Xenopus* spleens. The UMAP visualisations provide an overview of the cellular overlap of the immune system between these species. In the comparison between *Pleurodeles waltl* and *Homo sapiens* (Fig 9A), there is a discernible overlap in certain immune cell clusters, suggesting conserved elements in the immune systems across vast evolutionary distances. When comparing *Pleurodeles waltl* to *Xenopus laevis* (Fig 9B) we observed an extensive pattern of overlap in all the clusters. The comprehensive UMAP (Fig 9C) incorporating all three species delineates the extent of immunological conservation across the species.

A screenshot of a graph

Description automatically generated

**Fig 9 . Comparative UMAP Analysis of Immune Cells Across Species**

Comparative analysis of immune cell clusters from *Pleurodeles waltl* (pw), humans (hs), and *Xenopus laevis* (xl), as visualised through UMAP projections.

(A) Pleurodeles waltl and human immune cells (B) *Pleurodeles waltl* and *Xenopus laevis* immune cells (C) A combined UMAP of all three species, illustrating the conservation and divergence of immune cell clusters across the taxa. (D) A stacked bar chart representation of the cell counts for corresponding clusters across species, providing a quantitative perspective on the comparative abundance and distribution of immune cells.

Using this comprehensive stitched UMAP we obtained gene mapping tables to identify genes that were responsible for the observed cluster alignments.

Using the cluster alignment information from SAMap and the cross-species gene pair information we can identify the cell type triangles between all the three species this process helped us to confirm and identify the presence of other immune clusters as well.

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

**Fig 10. Sankey plot mapping the clusters in Pleurodeles with humans and Xenopus.**

(A) Shows the mapping between the clusters in the integrated dataset of Pleurodeles to humans and Xenopus. (B) Shows the mapping of the subclustered clusters in the integrated dataset of Pleurodeles to humans and Xenopus

Upon visualisation of the Sankey plots (Fig 10A), which represent the cross-species cluster mappings, we gained insights into several key alignments, such as *Pleurodeles* clusters corresponding with human macrophages, various T cell subsets, and B cell lineages, affirming the evolutionary conservation of these pivotal immune components. Intriguingly, the mappings of clusters 5 and 9 presented anomalies, aligning with different cell types and suggesting a potential mixed cell population within these clusters. To resolve these ambiguities, we subset these clusters for a more focused mapping effort, aiming to identify the cell types and ensure an accurate cross-species comparative analysis.

These subset clusters especially cluster 5 showed split mapping between macrophages and B cells (Fig 10B) suggesting the presence of an intermediate B cell type which might be more involved in signalling and communication between adaptive and innate immune cells.

# **4.Discussion**

In our exploration of the immune landscape of *Pleurodeles waltl*, we have uncovered a rich diversity of immune cell populations, showcasing a level of complexity and potential functional diversity. This has helped us in our understanding of *Pleurodeles waltl's* immune cell repertoire, which is linked to its extraordinary regenerative capabilities, a subject of growing interest and potential translational value.

Particularly noteworthy is the discovery of a multitude of B cell subclusters, revealing a range of gene ontology terms indicative of various immunological functions, from antigen presentation to the synthesis of antibodies (as illustrated in Fig 7). These results parallel findings in mammals, where B cell diversity is essential for effective immunity (Meffre & O’Connor, 2019). A compelling aspect of our research lies in the implications of our discoveries. The potential identification of regulatory T cells within Cluster 7, for instance, warrants further investigation given the important role Tregs have in regeneration in other species. This has profound implications for the field of regenerative medicine, where understanding such links can inspire innovative therapeutic strategies (Leigh & Currie, 2022).

Our application of Gene-Ontology Enrichment Analysis offers a deeper layer of understanding, delving into the biological processes that drive gene expression within these cell clusters. This approach is reflective of the methodologies employed in human immunological research, such as the work by (Villani et al., 2017) which profiled human dendritic cell subsets to unveil their diverse roles in immunity. While this approach was informative it also gave us a lot of mixed signals mostly due to the absence of an established dataset and the incomplete gene annotation of our *Pleurodeles* scRNA datasets. This prompted us to further consider an analysis with closely related model organisms like *Xenopus* which helped us to understand a broader picture of the *Pleurodeles* immune cell landscape.

The incorporation of cross-species comparison using SAMap analysis emerged as an pivotal part of our investigation, drawing insightful parallels between the immune systems of *Pleurodeles waltl*,  *Homo sapiens*, and *Xenopus laevis*. The resulting UMAP plots (Figure 9) underscored both conserved and divergent immunological features, reflecting the evolutionary adaptation of immune repertoires. The gene triangles which are genes being commonly upregulated and differentially expressed in all the three organisms utilised by SAMap also gives us alignments of the immune clusters between the species, through the comparative analysis, certain immune clusters revealed expected alignments with their mammalian counterparts, suggesting evolutionary conserved mechanisms. However, some of these alignments were not always straightforward, as highlighted by discrepancies in cluster mappings (Figure 10A). This is clear in cluster 5’s mapping with both B-cells and macrophages. Such inconsistencies prompted further investigation, leading to a re-examination of these clusters by sub-clustering them to ensure accurate identification (Figure 10B). By understanding the nuanced interplay between diverse immune cell types, this study contributes to the broader discourse on the evolutionary underpinnings of immunity and regeneration, setting the stage for future explorations into therapeutic interventions rooted in an evolutionary context.

# **5.References**

Antonsson, S. E., & Melsted, P. (2024). Batch correction methods used in single cell RNA-sequencing analyses are often poorly calibrated. *BioRxiv*, 2024.03.19.585562. https://doi.org/10.1101/2024.03.19.585562

Bolaños-Castro, L. A., Walters, H. E., García Vázquez, R. O., & Yun, M. H. (2021). Immunity in salamander regeneration: where are we standing and where are we headed? *Developmental Dynamics : An Official Publication of the American Association of Anatomists*, *250*(6), 753. https://doi.org/10.1002/DVDY.251

Brown, T., Elewa, A., Iarovenko, S., Subramanian, E., Araus, A. J., Petzold, A., Susuki, M., Suzuki, K. T., Hayashi, T., Toyoda, A., Oliveira, C., Osipova, E., Leigh, N. D., Simon, A., & Yun, M. H. (2022). Sequencing and chromosome-scale assembly of the giant Pleurodeles waltl genome. *BioRxiv*, 2022.10.19.512763. https://doi.org/10.1101/2022.10.19.512763

Bryant, D. M., Johnson, K., DiTommaso, T., Tickle, T., Couger, M. B., Payzin-Dogru, D., Lee, T. J., Leigh, N. D., Kuo, T. H., Davis, F. G., Bateman, J., Bryant, S., Guzikowski, A. R., Tsai, S. L., Coyne, S., Ye, W. W., Freeman, R. M., Peshkin, L., Tabin, C. J., … Whited, J. L. (2017). A Tissue-Mapped Axolotl De Novo Transcriptome Enables Identification of Limb Regeneration Factors. *Cell Reports*, *18*(3), 762–776. https://doi.org/10.1016/J.CELREP.2016.12.063

Butler, N. M., & Hannapel, D. J. (2012). Promoter activity of polypyrimidine tract-binding protein genes of potato responds to environmental cues. *Planta*, *236*(6), 1747–1755. https://doi.org/10.1007/S00425-012-1726-7

Cardiello, J. F., Joven Araus, A., Giatrellis, S., Helsens, C., Simon, A., & Leigh, N. D. (2023). Evaluation of genetic demultiplexing of single-cell sequencing data from model species. *Life Science Alliance*, *6*(8), e202301979. https://doi.org/10.26508/lsa.202301979

Farber, D. L., Yudanin, N. A., & Restifo, N. P. (2014). Human memory T cells: generation, compartmentalization and homeostasis. *Nature Reviews. Immunology*, *14*(1), 24–35. https://doi.org/10.1038/NRI3567

Fleming, S. J., Chaffin, M. D., Arduini, A., Akkad, A. D., Banks, E., Marioni, J. C., Philippakis, A. A., Ellinor, P. T., & Babadi, M. (2023). Unsupervised removal of systematic background noise from droplet-based single-cell experiments using CellBender. *Nature Methods 2023 20:9*, *20*(9), 1323–1335. https://doi.org/10.1038/s41592-023-01943-7

Frazer, K. A., Elnitski, L., Church, D. M., Dubchak, I., & Hardison, R. C. (2003). Cross-Species Sequence Comparisons: A Review of Methods and Available Resources. *Genome Research*, *13*(1), 1–12. https://doi.org/10.1101/gr.222003

Godwin, J. W., Pinto, A. R., & Rosenthal, N. A. (2013). Macrophages are required for adult salamander limb regeneration. *Proceedings of the National Academy of Sciences*, *110*(23), 9415–9420. https://doi.org/10.1073/pnas.1300290110

Gurtner, G. C., Werner, S., Barrandon, Y., & Longaker, M. T. (2008). Wound repair and regeneration. *Nature*, *453*(7193), 314–321. https://doi.org/10.1038/nature07039

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., Mende, D. R., Letunic, I., Rattei, T., Jensen, L. J., von Mering, C., & Bork, P. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research*, *47*(D1), D309–D314. https://doi.org/10.1093/nar/gky1085

Jiao, A., Zhang, C., Wang, X., Sun, L., Liu, H., Su, Y., Lei, L., Li, W., Ding, R., Ding, C., Dou, M., Tian, P., Sun, C., Yang, X., Zhang, L., & Zhang, B. (2024). Single-cell sequencing reveals the evolution of immune molecules across multiple vertebrate species. *Journal of Advanced Research*, *55*, 73–87. https://doi.org/10.1016/J.JARE.2023.02.017

Joven, A., Elewa, A., & Simon, A. (2019). Model systems for regeneration: salamanders. *Development*, *146*(14). https://doi.org/10.1242/dev.167700

Julier, Z., Park, A. J., Briquez, P. S., & Martino, M. M. (2017). Promoting tissue regeneration by modulating the immune system. *Acta Biomaterialia*, *53*, 13–28. https://doi.org/10.1016/J.ACTBIO.2017.01.056

Kikuchi, K. (2020). New function of zebrafish regulatory T cells in organ regeneration. *Current Opinion in Immunology*, *63*, 7–13. https://doi.org/10.1016/J.COI.2019.10.001

Kondělková, K., Vokurková, D., Krejsek, J., Borská, L., Fiala, Z., & Ctirad, A. (2010). Regulatory T cells (TREG) and their roles in immune system with respect to immunopathological disorders. *Acta Medica (Hradec Kralove)*, *53*(2), 73–77. https://doi.org/10.14712/18059694.2016.63

Leigh, N. D., & Currie, J. D. (2022). Rebuilding limbs, one cell at a time. *Developmental Dynamics*, *251*(9), 1389–1403. https://doi.org/10.1002/DVDY.463

Li, J., Tan, J., Martino, M. M., & Lui, K. O. (2018a). Regulatory T-Cells: Potential Regulator of Tissue Repair and Regeneration. *Frontiers in Immunology*, *9*(MAR), 1. https://doi.org/10.3389/FIMMU.2018.00585

Li, J., Tan, J., Martino, M. M., & Lui, K. O. (2018b). Regulatory T-Cells: Potential Regulator of Tissue Repair and Regeneration. *Frontiers in Immunology*, *9*(MAR), 1. https://doi.org/10.3389/FIMMU.2018.00585

Luckheeram, R. V., Zhou, R., Verma, A. D., & Xia, B. (2012). CD4 +T cells: Differentiation and functions. *Clinical and Developmental Immunology*, *2012*. https://doi.org/10.1155/2012/925135

Meffre, E., & O’Connor, K. C. (2019). Impaired B-cell tolerance checkpoints promote the development of autoimmune diseases and pathogenic autoantibodies. *Immunological Reviews*, *292*(1), 90–101. https://doi.org/10.1111/IMR.12821

Menger, B., Vogt, P. M., Kuhbier, J. W., & Reimers, K. (2010). Applying Amphibian Limb Regeneration to Human Wound Healing. *Annals of Plastic Surgery*, *65*(5), 504–510. https://doi.org/10.1097/SAP.0b013e3181d376f9

Nagel, A., Möbs, C., Raifer, H., Wiendl, H., Hertl, M., & Eming, R. (2014). CD3-Positive B Cells: A Storage-Dependent Phenomenon. *PLoS ONE*, *9*(10). https://doi.org/10.1371/JOURNAL.PONE.0110138

Nemazee, D., & al., et. (2000). B-cell-receptor-dependent positive and negative selection in immature B cells. *Curr. Top. Microbiol. Immunol.*, *245*, 57–71.

Rastogi, I., Jeon, D., Moseman, J. E., Muralidhar, A., Potluri, H. K., & McNeel, D. G. (2022). Role of B cells as antigen presenting cells. *Frontiers in Immunology*, *13*, 954936. https://doi.org/10.3389/FIMMU.2022.954936/BIBTEX

Reth, M. (2001). Oligomeric antigen receptors: A new view on signaling for the selection of lymphocytes. *Trends in Immunology*, *22*(7), 356–360. https://doi.org/10.1016/S1471-4906(01)01964-0

Shanley, L. C., Mahon, O. R., Kelly, D. J., & Dunne, A. (2021). Harnessing the innate and adaptive immune system for tissue repair and regeneration: Considering more than macrophages. *Acta Biomaterialia*, *133*, 208–221. https://doi.org/10.1016/J.ACTBIO.2021.02.023

Smith-Garvin, J. E., Koretzky, G. A., & Jordan, M. S. (2009). T cell activation. *Annual Review of Immunology*, *27*, 591–619. https://doi.org/10.1146/ANNUREV.IMMUNOL.021908.132706

Villani, A. C., Satija, R., Reynolds, G., Sarkizova, S., Shekhar, K., Fletcher, J., Griesbeck, M., Butler, A., Zheng, S., Lazo, S., Jardine, L., Dixon, D., Stephenson, E., Nilsson, E., Grundberg, I., McDonald, D., Filby, A., Li, W., De Jager, P. L., … Hacohen, N. (2017). Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science*, *356*(6335). https://doi.org/10.1126/SCIENCE.AAH4573/SUPPL\_FILE/AAH4573\_VILLANI\_SM.PDF

Zhu, J., & Paul, W. E. (2008). CD4 T cells: fates, functions, and faults. *Blood*, *112*(5), 1557–1569. https://doi.org/10.1182/BLOOD-2008-05-078154