Supplementary Materials

Table S1. TraCeR and BraCeR analysis.

Table S2. Number of cells in each cluster across different experimental conditions (10x RNAseq data).

Clusters	***rag1*$^{+/+}$ PBS-injected**	***rag1*$^{-/-}$ PBS-injected**	***rag1*$^{-/-}$ *V. anguillarum*-injected**	***rag1*$^{-/-}$ *A. simplex*-injected**
1	824	27 *(nitr-rorc+)*	195 ILC3-like *(nitr+rorc+il22+)*	972
2	745	1694	1263	677 ILC1-like *(ifng1-1+)*
3	718	123 *(nitr+rorc+)*	923	1689
4	529	558 ILC1-like *(ifng1-1+)*	652 ILC1-like *(ifng1-1+)*	70
5	246	205	132	108
6	243	375	85 *(foxp3a+)*	41 *(foxp3a+)*
7	229	19	129	31 *(nitr-rorc+)*
8	120	132	23	4
9	72	29	41	31 ILC2-like *(nitr+rorc+il13+)*
10	30	32	13 *(nitr-rorc+)*	3
11	-	14	28	-
12	-	3	3	-

Table S3. Index sorting data.

Explanatory note

Lymphocytes development occurs through several stages, at distinct body locations, each marked with specific gene expression patterns. To capture the diversity of lymphoid cell types, we purified and sequenced the RNA from single cells collected from primary lymphoid organs (kidney and thymus), secondary lymphoid organs (spleen) as well as barrier tissues (gut and gills). The cloud repository we generated (https://www.sanger.ac.uk/science/tools/lymphocytes/lymphocytes/) enables users to investigate the expression of genes of interest across the population of zebrafish lymphocytes. Available datasets includes Smart-seq2 data from kidney, thymus, spleen, guts and gills of healthy, unstimulated wild-type zebrafish as well as 10x datasets from gut of unstimulated zebrafish both wild-type and *rag1*$^{-/-}$ mutant and immune-challenged (*V. anguillarum*- or *A. simplex*-injected) *rag1*$^{-/-}$ mutant.

Users can select the data set of interest from the "Single cells in 3D space" menu on the left side of the web page. After selecting a relevant dataset, individual cells in 3D t-SNE plot will be visible in the middle of the screen. The scale for gene expression levels is displayed in the top-right corner of the 3D t-SNE plot. Users can explore the expression levels of a single gene or two genes by selecting a corresponding option in the menu on the left side of the web page.

By hovering the mouse cursor above a cell of interest in 3D t-SNE plot users can see the level of expression of the gene (for the single gene option) or the ratio between expression of two selected genes (for the two-genes option) in the single cell; as well as information about cluster and cell type the cell belongs to. Both size and colour of the cells in t-SNE plot can be modified by using the sliders on the left.

A graph showing the summary of the expression levels across the clusters in the selected dataset can be found below the 3D t-SNE plot. User can look at the distribution of clusters in 3D t-SNE plot by selecting the "Hide/Show Clusters" option under the "Visualisation" menu on the left side of the page.

The cloud repository offers a possibility to download the selected t-SNE plot as png by selecting "Download plot as a png" by hovering the mouse cursor over the top right space, above the scale for gene expression. Other options include zoom in, pan, orbital orientation *etc*.

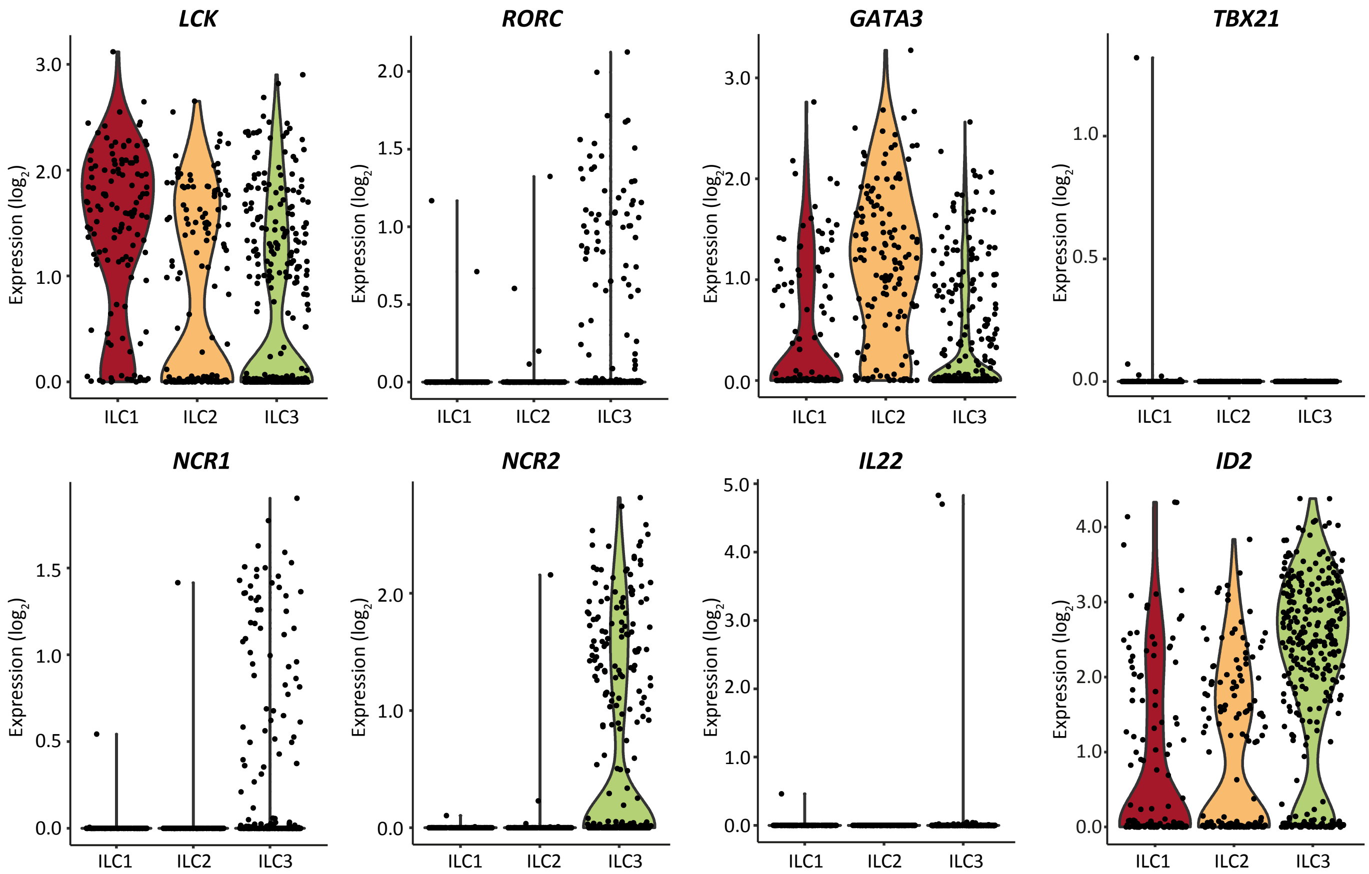


Figure S1. Expression of marker genes in human ILCs. Violin plots showing the expression of marker genes in individual human ILCs. No expression of IL13 was detected in this dataset. The expression values were calculated from publicly available dataset (33).

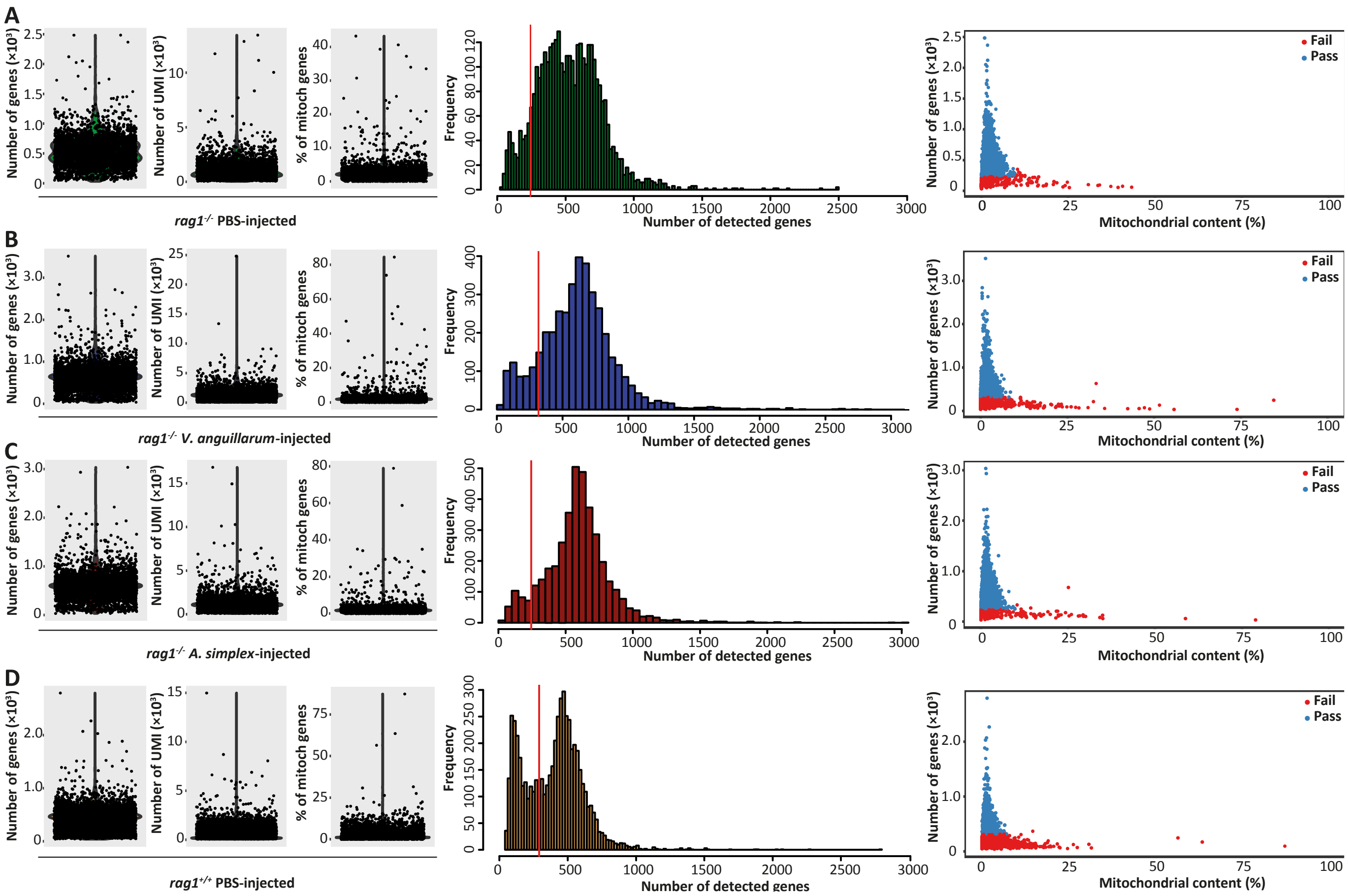


Figure S2. Quality control of 10x data. Number of detected genes, detected mRNA and the percentage of mitochondrial content in single cells from each of the 10x samples (left), histograms for the number of detected genes per sample (middle), QC plots with the number of detected genes versus the percentage of mitochondrial content. Cells that passed QC are coloured in blue, and cells that failed the QC are coloured in red (right). **A.** *rag1$^{-/-}$* PBS-injected zebrafish. **B.** *rag1$^{-/-}$ V. anguillarum*-injected zebrafish. **C.** *rag1$^{-/-}$ A. simplex*-injected zebrafish. **C.** *rag1$^{+/+}$* PBS-injected zebrafish.

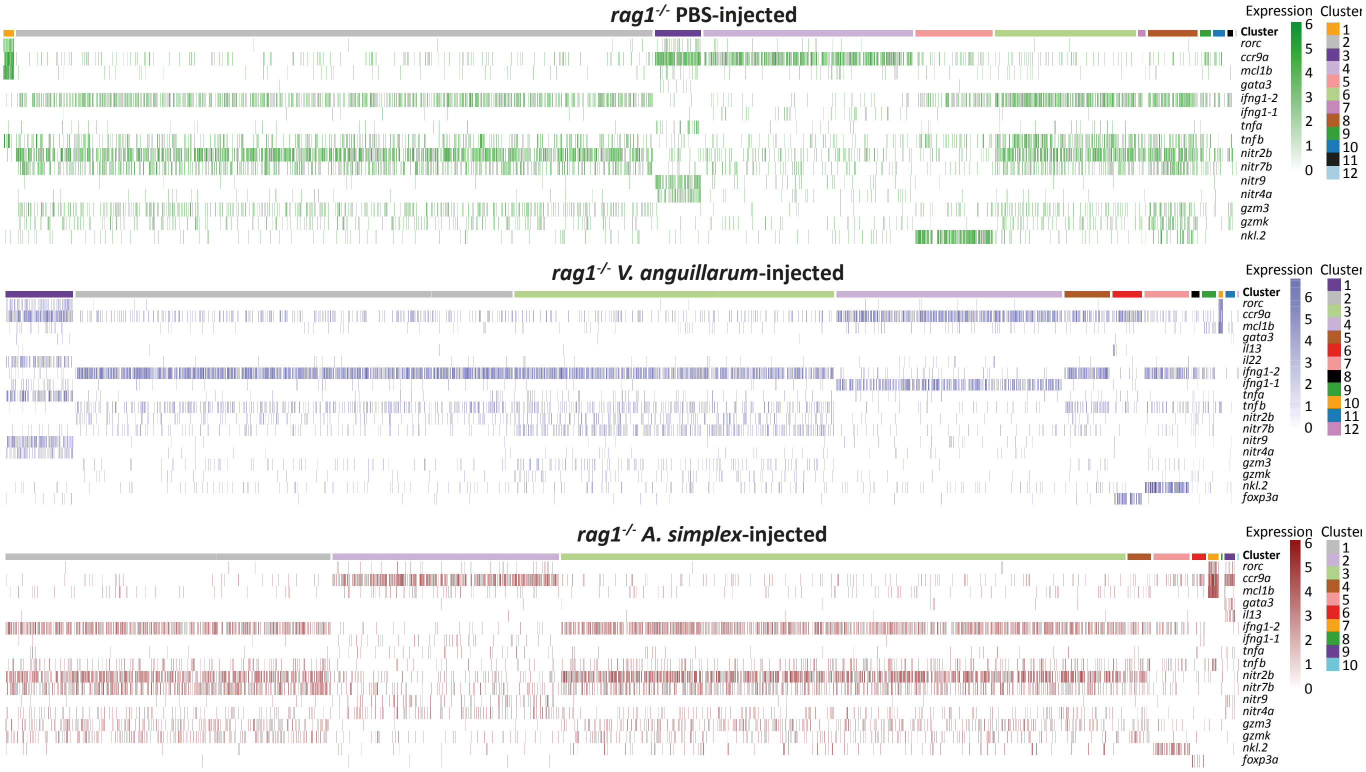
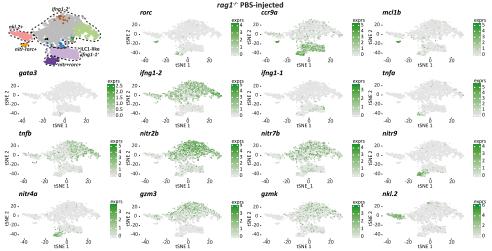
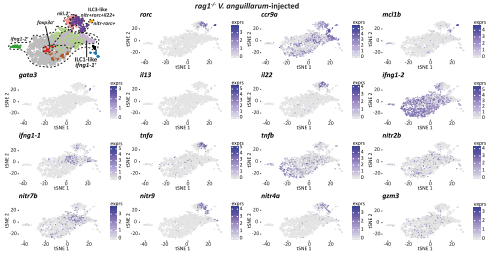


Figure S3. Expression of marker genes in *lck*+ cells collected from *rag1*$^{-/-}$ gut. Heatmaps showing expression of marker genes in each of the clusters for *lck*+ cells collected from *rag1*$^{-/-}$ gut, PBS-injected (top), *V. anguillarum-injected* (middle), *A. simplex-injected* (bottom).


$rag1^{-/-}$ PBS-injected


$rag1^{-/-}$ V. anguillarum-injected

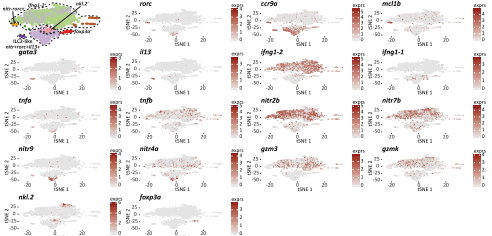

$rag1^{-/-}$ A. simplex-injected

Figure S4. Expression of marker genes in *lck*+ cells collected from *rag1*$^{-/-}$ gut. t-SNE plots showing expression of marker genes in each of the clusters for *lck*+ cells collected from *rag1*$^{-/-}$ gut, PBS-injected (top), *V. anguillarum*-injected (middle), *A. simplex*-injected (bottom) zebrafish. Expression levels are log2 scaled.

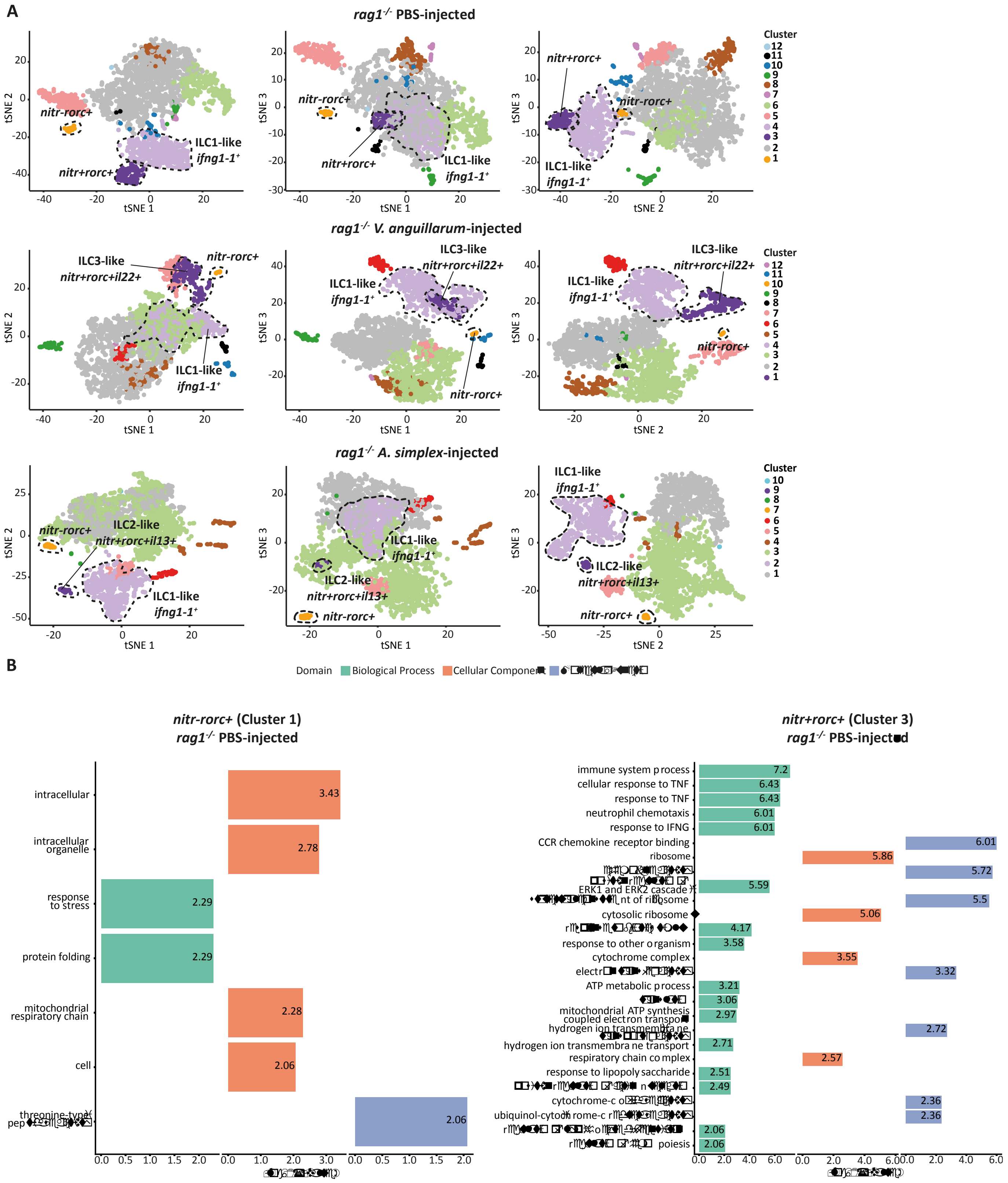


Figure S5. Identification of *nitr-rorc+* and *nitr+rorc+* population of cells. A. tSNE plots representing distribution of clusters of *lck+* cells from gut of *rag1*$^{-/-}$ zebrafish in different dimensions. **B.** GO terms characteristic for *nitr-rorc+* (Cluster 1) and *nitr+rorc+* (Cluster 3) cells in PBS-injected *rag1*$^{-/-}$ zebrafish.

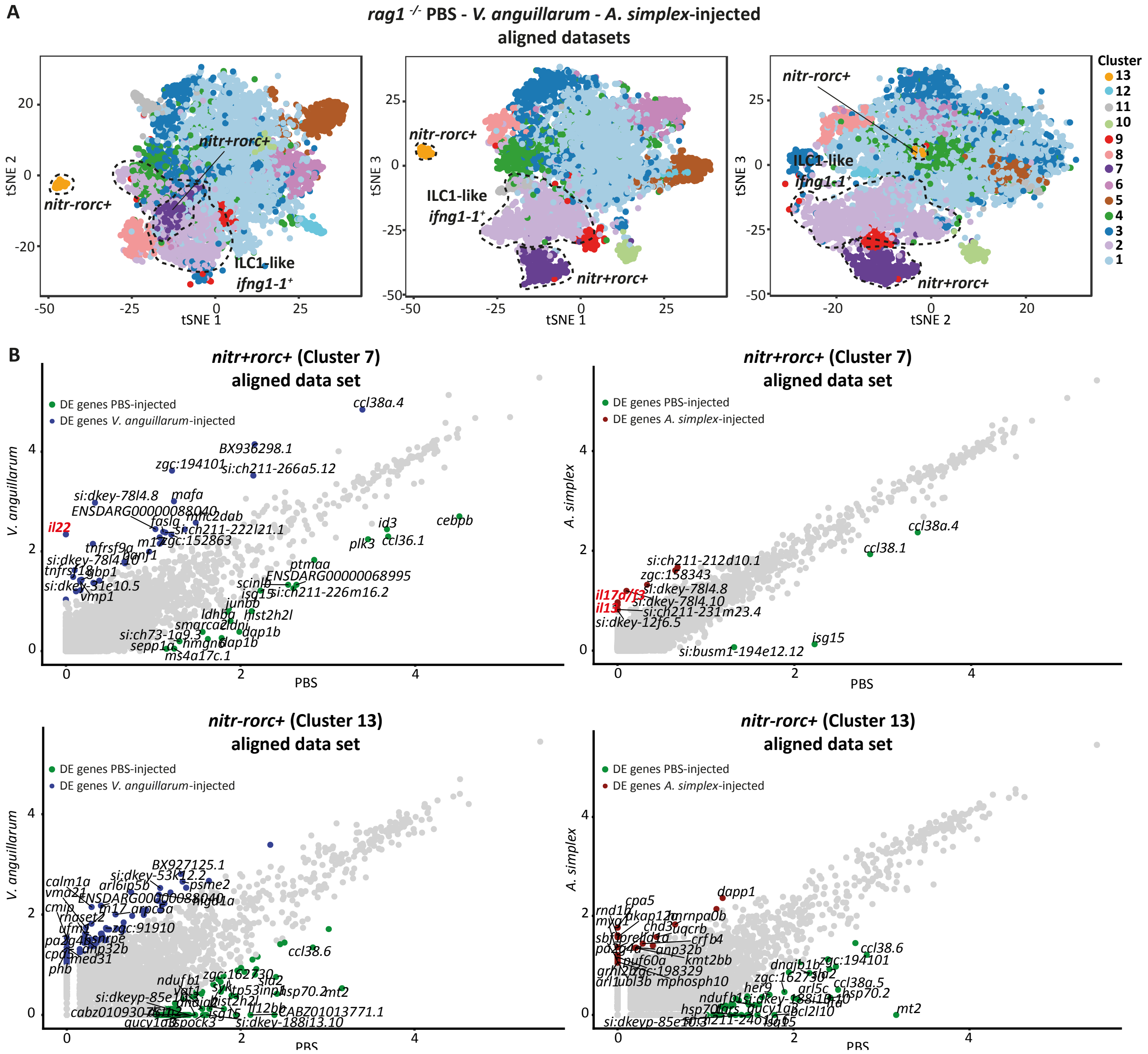


Figure S6. PBS - *V. anguillarum* - *A. simplex*-injected *rag1*$^{-/-}$ mutant zebrafish aligned data sets. A. tSNE plots representing distribution of aligned clusters in different dimensions. **B.** Regression plots showing top 20 DE genes between: *V. anguillarum* and PBS-injected zebrafish (left) and *A. simplex* and PBS-injected zebrafish (right) in *nitr+rorc+* cells (Cluster 7) and *nitr-rorc+* (Cluster 13).

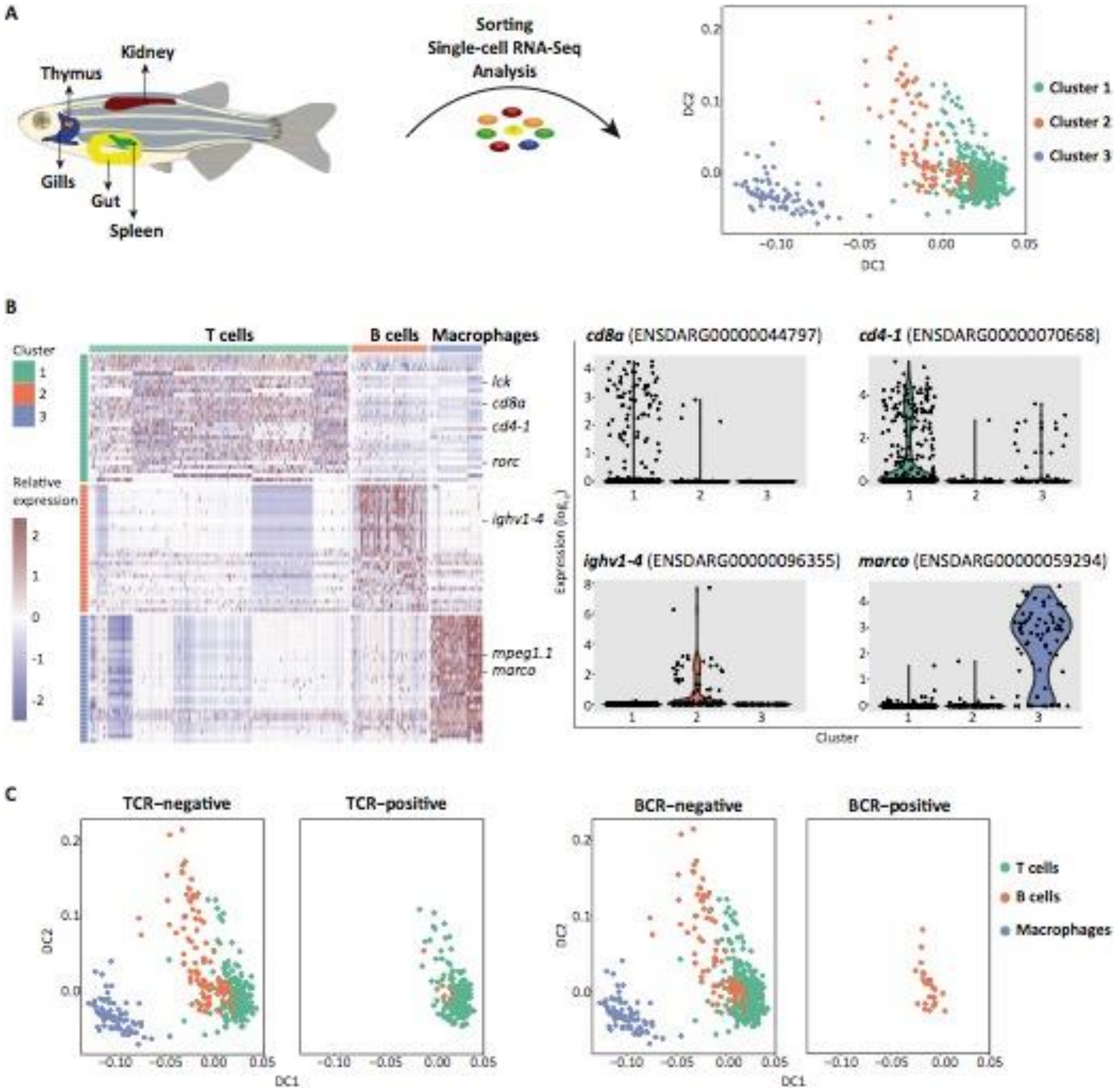


Figure S7. Identification of immune cell types in zebrafish during steady state haematopoiesis. A. Experimental strategy for sorting single cells from transgenic zebrafish lines. Cells were collected from different organs and transgenic lines, namely, *Tg(cd4-1:mCherry)*, *Tg(lck:EGFP)* and *Tg(mch2dab:GFP, cd45:DsRed)*. Single cells were index-sorted into 96-well plates and processed with standard Smart-seq2 protocol followed by computational analysis. **B.** Heatmap showing the expression level of marker genes for each of the identified cluster. Columns represent cells and rows marker genes. Violin plots showing the expression of signature genes in each cluster. **C.** TraCeR and BraCeR analysis showing cells positive and negative for TCR and BCR rearrangements across the cell types.

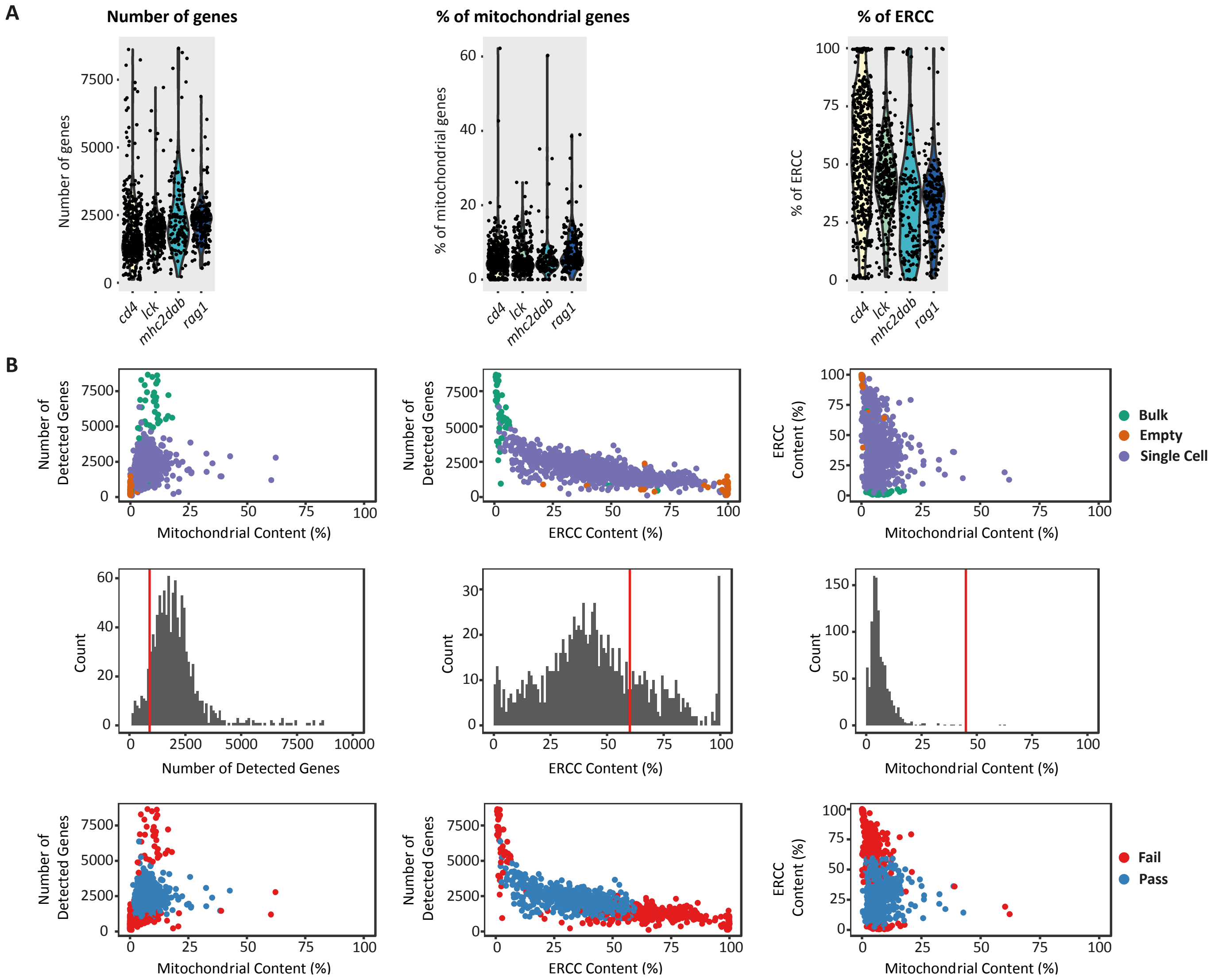


Figure S8. Quality control of SmartSeq2 data. A. Number of detected genes, mitochondrial and ERCC content in single cells from each of the transgenic lines. **B**. Plots for the number of detected genes, % ERCC and % mitochondria (MT) genes. Cells that passed QC (ERCC < 60 % MT < 45% and at least 900 detected genes) are coloured in blue. The bulk and empty wells were removed from the final plots.

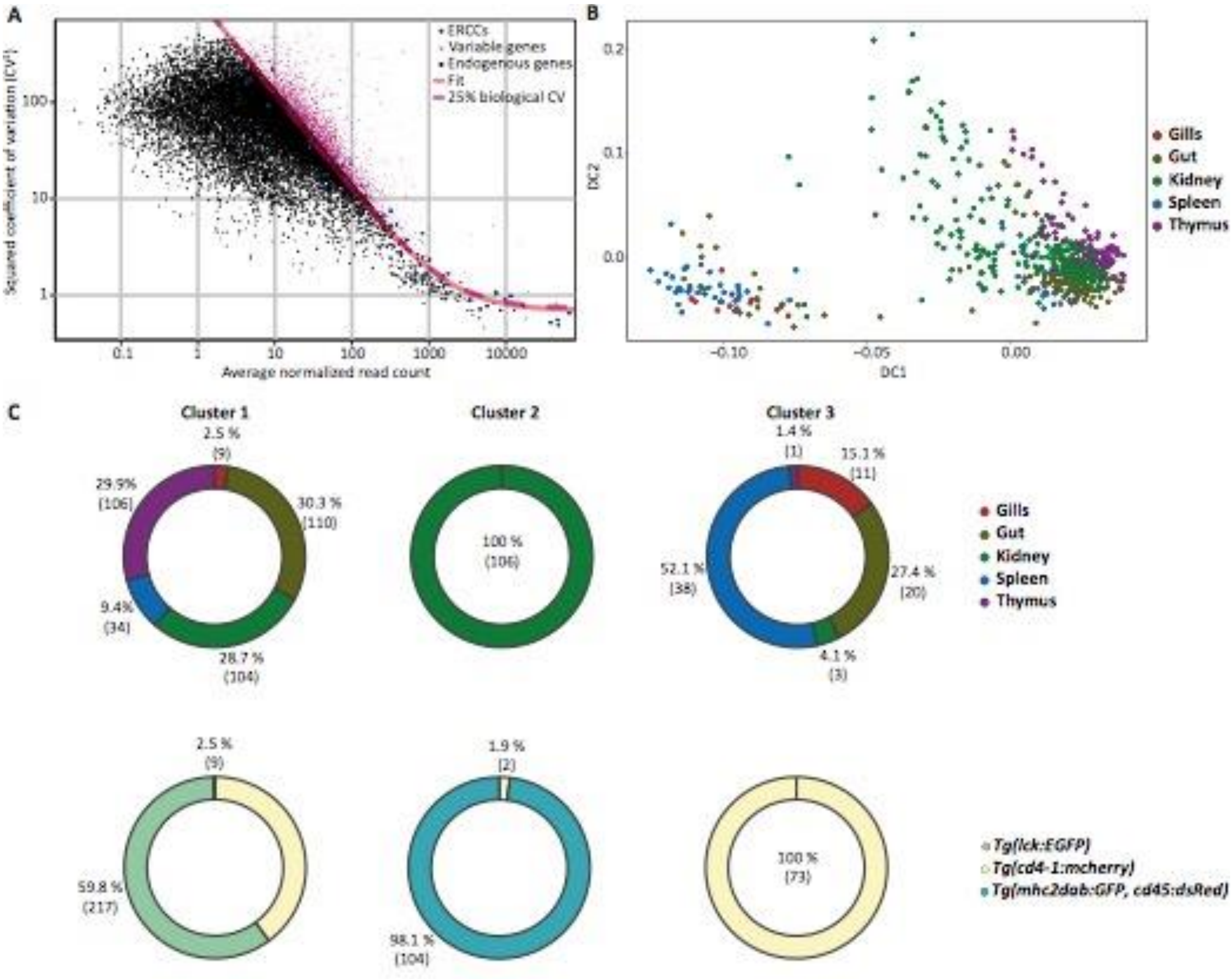


Figure S9. Clustering of Smart-seq2 data from *Tg(lck:EGFP)*, *Tg(cd4-1:mcherry)* and *Tg(mhc2dab:GFP, cd45:dsRed)* zebrafish transgenic lines. A. Brennecke method was used for identification of highly variable genes. The orange solid line shows the curve of the technical noise fit and the pink dashed line shows the position of genes with 20% biological coefficient of variance (CV). Blue dots indicate ERCCs, pink dots indicate highly variable genes, black dots represent the rest of genes expressed in the data set. **B.** Diffusion map showing origin of cells in each of the clusters. **C.** Pie charts with the absolute number (in brackets) and percentage of cells that originated from different organs (upper row) and from different transgenic lines *(Tg(lck:EGFP)*, *Tg(cd4-1:mcherry)*, *Tg(mhc2dab:GFP, cd45:dsRed*)) across all clusters (bottom row).