## Supplementary figures



**Supplementary figure 1.** Fresh Atlantic cod blood and spleen samples were flow-sorted into four populations each based on forward scatter and side scatter gating. Representative images of HE-stained cells from blood sub-populations (B1-B4) are shown on the left, while spleen sub-populations (S1-S4) are shown on the right. The percentage of the populations are as follows: B1-3%, B2-28%, B3-13%, B4-4%, S1-18%, S2-8%, S3-8%, S4-1.5%. Scale bars 10 µm



Cell barcodes sorted by number of reads (descending)

**Supplementary figure 2.** Cumulative distribution of reads used to identify STAMPs in a pool of amplified beads. Each number, 1-11, represents the sample identity. Drop-seq involves the generation of single-cell RNA profiles by adding a single cell to a single droplet containing a single bead. The large majority of amplified beads are not encapsulated with a cell and so are not exposed to a single cell's RNA, only the ambient RNA present in solution following droplet breakage. To identify the cell barcodes which do correspond to STAMPs, cell barcodes from each sample used in the experiment are arranged in decreasing number of reads, and the cumulative fraction of reads is plotted. The predicted number of individual cells calculated from the counted and aliquoted beads for each sample was estimated during the laboratory procedure. However, there is no clear inflection point at these estimated numbers. To be sure to include all salient information it was decided to include the first 600-3000 cell barcodes (shown by the black line), to be further filtered at later steps.



**Supplementary figure 3.** Feature scatter (top) and violin plots (bottom) showing the 11 samples before (left) and after (right) quality control cut offs are applied. Cells with a gene count (nFeature\_RNA) of fewer than 150 and more than 1500, cells with a total number of molecules of more than 4000 (nCount\_RNA) were filtered away in order to remove low-quality cells and possible cell multiplets.



## Supplementary figure 4.

Elbow plot showing a ranking of principle components based upon the percentage of variance explained by each. Here we observe an 'elbow' around PC25-30, suggesting that the majority of true signal is captured in the first 30 PCs.

	Sample number	Tissue	Library size (millions of fragments)	Average mapping (%)	Reads mapped (millions)	Number of cells included after filtering	Average number of mapped reads per cell	Average number of mapped transcripts per cell*	Average number of mapped genes per cell
Fish 1	1	Spleen	114	41	47	1153	4157	1356	628
	2	Spleen	51	73	37	327	5922	1381	557
	3	S3 sorted spleen	116	53	61	2102	6083	1059	430
	4	B1 sorted blood	60	48	29	1226	4985	656	318
Fish 2	5	Spleen	66	70	46	758	5333	637	306
	6	PBL	60	66	40	1780	2120	583	340
	7	S3 sorted spleen	15	52	8	230	2610	793	283
	8	B1 sorted blood	23	66	15	256	2794	730	332
	9	P1 sorted PBL	18	79	14	197	6882	753	437
	10	P2 sorted PBL	7	73	5	60	5021	731	418
	11	P3 sorted PBL	22	46	10	91	5231	1113	503
			Total = 552	Average = 61	Total = 312	Total = 8180	**Weighted average = 4695	**Weighted average = 884	**Weighted average = 409

**Supplementary table 1.** Overview of sample origin and sequencing results. Fresh Atlantic cod blood and spleen samples were taken from two Atlantic cod and were each flow-sorted into four populations based on forward scatter and side scatter gating: B1-B4 for the blood, and S1-S4 for the spleen. The populations of interest (B1, containing lymphocytes, and S3, containing myeloid cells) were sent for sequencing. Peripheral blood leukocytes (PBL) were obtained by centrifugation and also flow-sorted and the populations P1 (containing mostly lymphocytes), P2 (containing mostly lymphocytes and thrombocytes) and P3 (containing mostly myeloid cells).were sent for sequencing. An average of 61% of sequencing reads were mapped to the Atlantic cod genome (gadmor3). Following drop-seq computational protocol pipeline and the Seurat package filtering steps gene data from 8180 cells are included, with an average of 409 genes expressed per cell. \*Number of transcripts per cell given are after accounting for PCR amplification \*\*Averages are weighted by number of cells per sample.



**Supplementary figure 5.** Cell types from unsorted blood and spleen visualised in two dimensions using UMAP. Cells are from pilot study using wild-caught Atlantic cod in Oslo Fjord. Putative cell cluster labels are based upon the differential gene expression of known markers in mammals. The same parameters used in the main study have been applied: cells with a gene count of fewer than 150 or a gene count of more than 1500, and cells with a total number of molecules of more than 4000 were filtered away in order to remove low-quality cells and possible cell multiplets. Genes expressed in less than 5 cells were excluded. 30 principal components were included and the resolution for clustering was 0.35. The T cells and erythrocyte cell clusters are split into two subclusters at this resolution due to batch effects.



**Supplementary figure 6.** Plot showing the distribution of unsorted spleen, sorted spleen and sorted blood cells from fish 1 in red (sample 1-4) and fish 2 in blue (samples 5, 7 and 8) across the clusters. The 'other' cells in grey are from fish 2, and represent the peripheral blood leukocytes (PBL, sample 6) and sorted PBL (sample 9-11).



**Supplementary figure 7.** Average expression of MHCI genes across each cell cluster. Error bars show standard deviation.