Extended Data Figs

Extended Data Fig. 1. High-quality full-length transcriptomes enable downstream molecular identification.

Extended Data Fig. 2. Circular phylogenetic tree of the subunits of GPIb-IX complex across species.

Extended Data Fig. 3. Sequence identities, exon-intron phases, and distribution of the GPIbβ and GPIX subunits across jawless species.

Extended Data Fig. 4. Endogenous validation showing the presence of the VLRB-GPIbβ-GPIX complex in native VLRB cells.

Extended Data Fig. 5. Multiple sequence alignment of the protein sequences of GPIb-IX and VLRB focusing on the N-terminal and C-terminal regions.

Extended Data Fig. 6. Luciferase reporter assay for LcBHLHE41, LcFOS and LcSTAT5A regulations of *LcGp1bb-1*, *LcGp9-1* and *LcVlrb* promoters.

Extended Data Fig. 7. Phylogenetic tree of the FOS-associated genes.

Extended Data Fig. 8. Supplementary analysis of developmental trajectories underlying VLRB heterogeneity.

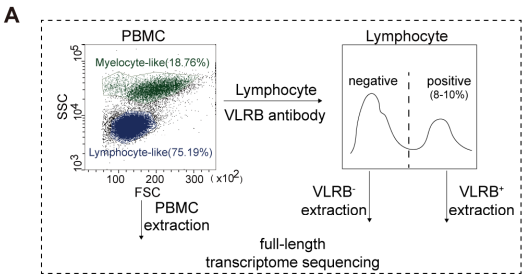
Supplementary Data 1. Summary of the basic characteristics of the 14 selected ITAM-containing candidates.

Supplementary Data 2. Accession numbers of sequences used for evolutionary analysis.

Supplementary Data 3. Analysis of promoter regions and predicted transcription factor binding sites.

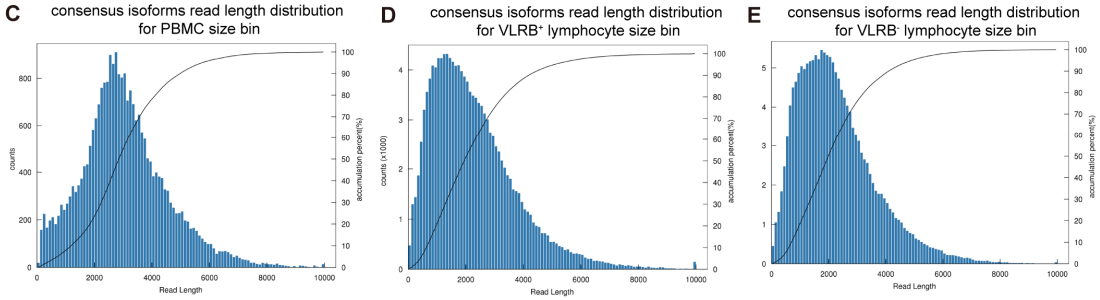
Supplementary Data 4. Probe sequences used in the DNA pull-down assays.

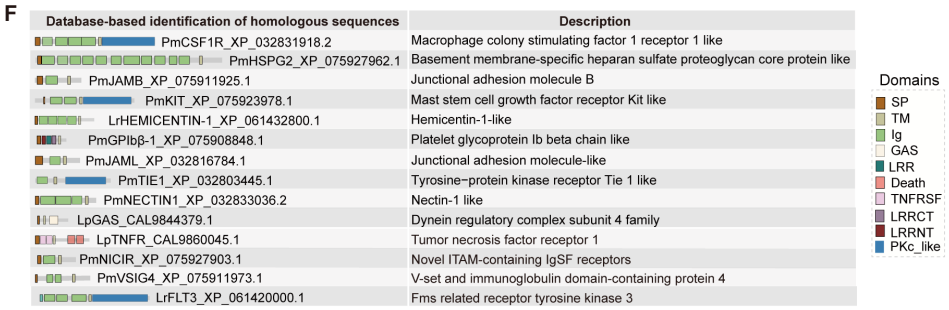
Extended Data Figs

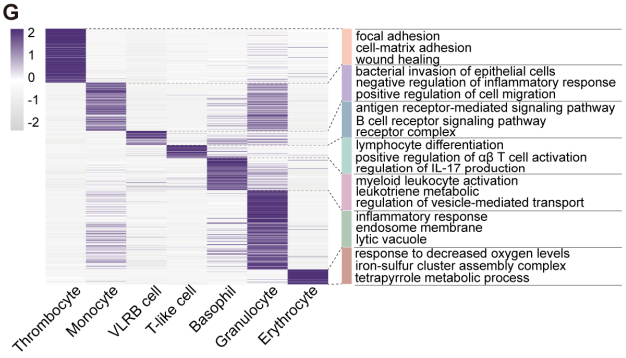


B

	PBMC	VLRB$^+$ Lymphocyte	VLRB$^-$ Lymphocyte
Number of Consensus sequences	25,824	135,465	165,538
Average read length of Consensus isoforms	3,059	2,285	2,283
high-quality isoforms(%)	98.51%	99.94%	99.93%
Number of predicted ORFs	15,353	61,972	77,592

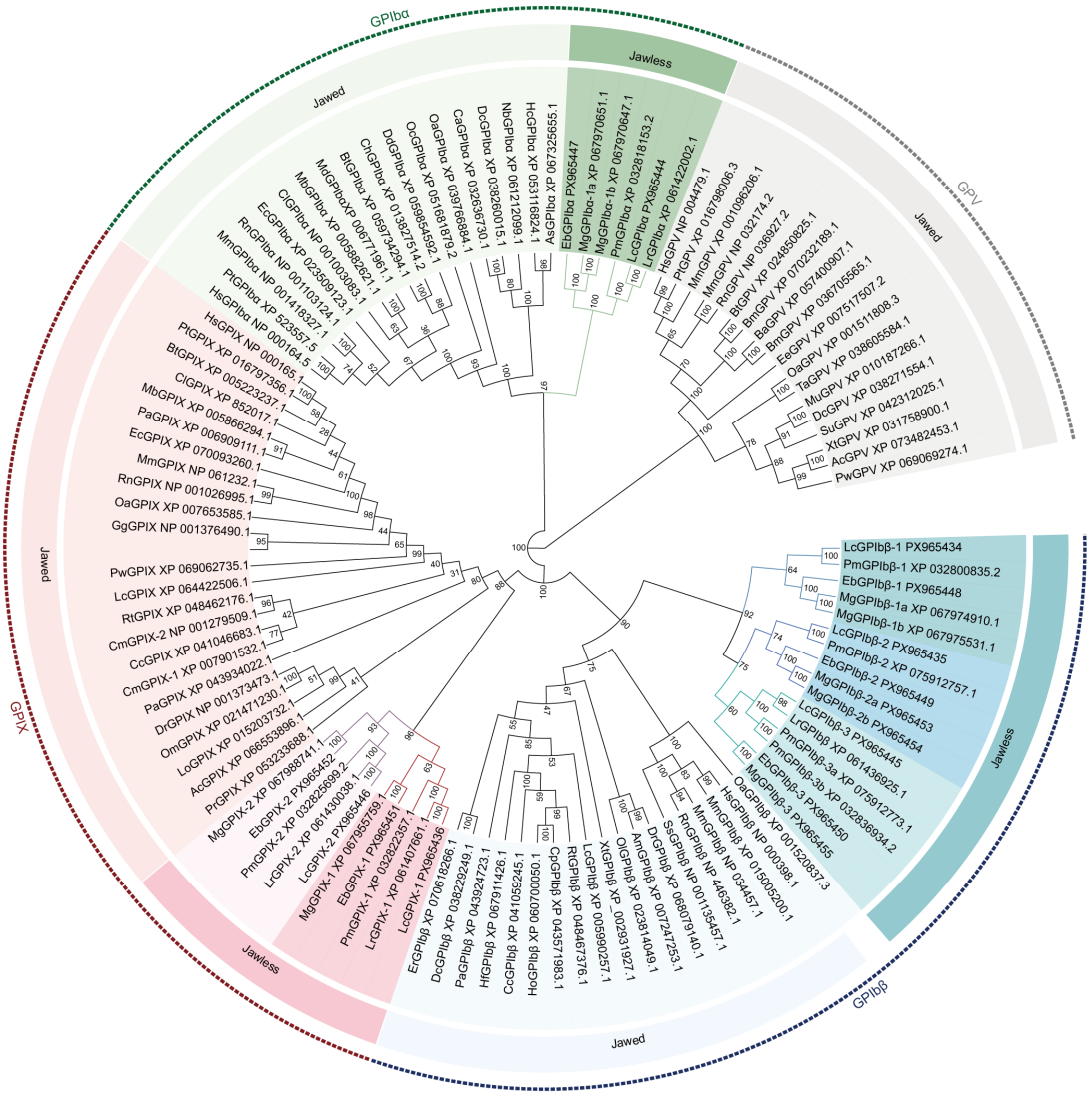






Extended Data Fig. 1 | High-quality full-length transcriptomes enable downstream molecular identification. A, Three sample sources of full-length transcriptome sequencing. Left panel shows the overall PBMC population as a flow cytometry scatter plot; right panel shows the proportion of LcVLRB antibody-positive and antibody-negative cells within the lymphocyte gate. **B,** Overview and quality assessment of three

full-length transcriptomes sequencing. ORF, open reading frame. **C-E,** Length distribution of consensus isoforms obtained from PBMC, $VLRB^+$ and $VLRB^-$ full-length transcriptome sequencing, respectively. The histogram shows the read length distribution of consensus isoforms, and the overlaid line indicates the accumulation percentage. **F,** Features and descriptions of corresponding molecules in lamprey protein database on NCBI for candidates screened from *L. camtschaticum*. Fourteen candidates identified in *L. camtschaticum* were searched against online databases. The table presents the corresponding molecules in other lamprey species, including their descriptions and structural information. **G,** GO and KEGG pathway analysis of differentially expressed genes across single-cell subpopulations of lamprey PBMCs ($log_2FC \geq 2$, pct.1 ≥ 0.1). GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes.

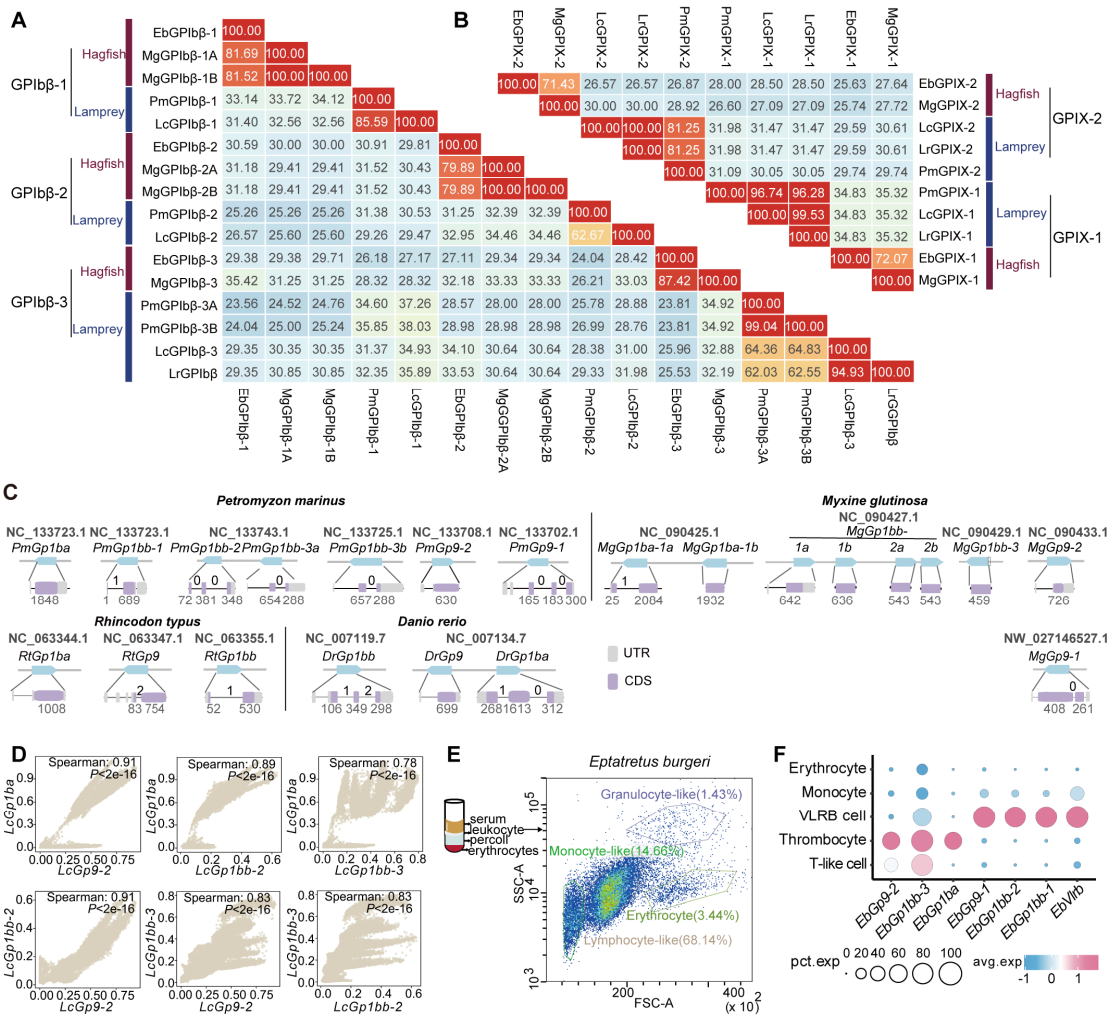


21

22 **Extended Data Fig. 2 | Circular phylogenetic tree of the subunits of GPIb-IX**

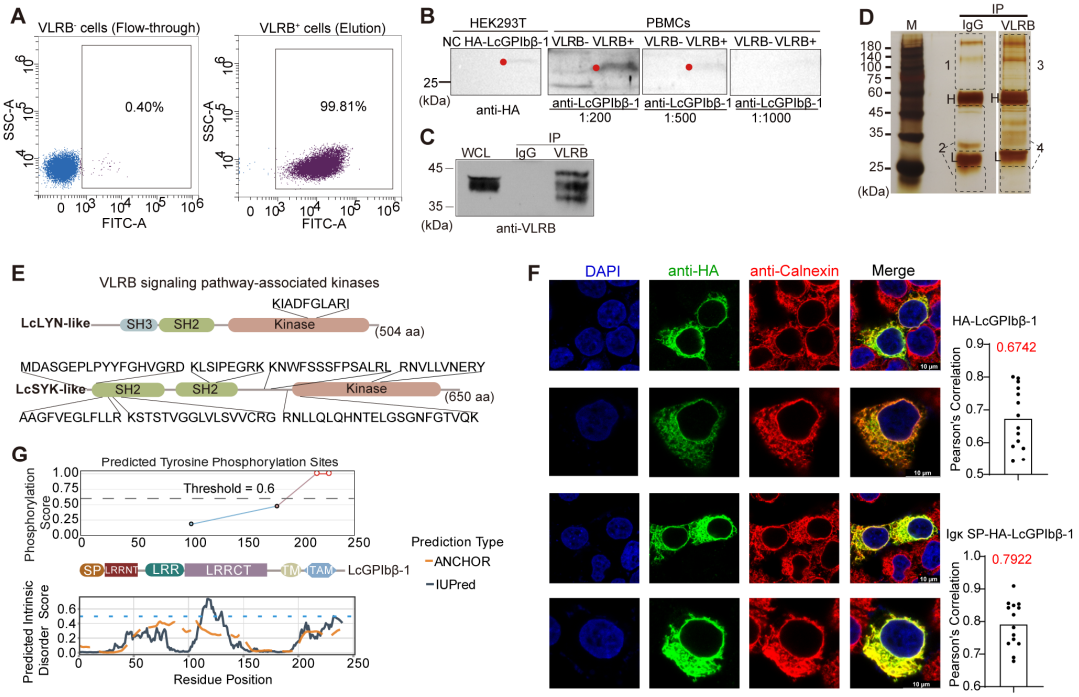
23 **complex across species.** The tree was constructed using IQ-TREE. Numbers on the

24 nodes indicate bootstrap values.



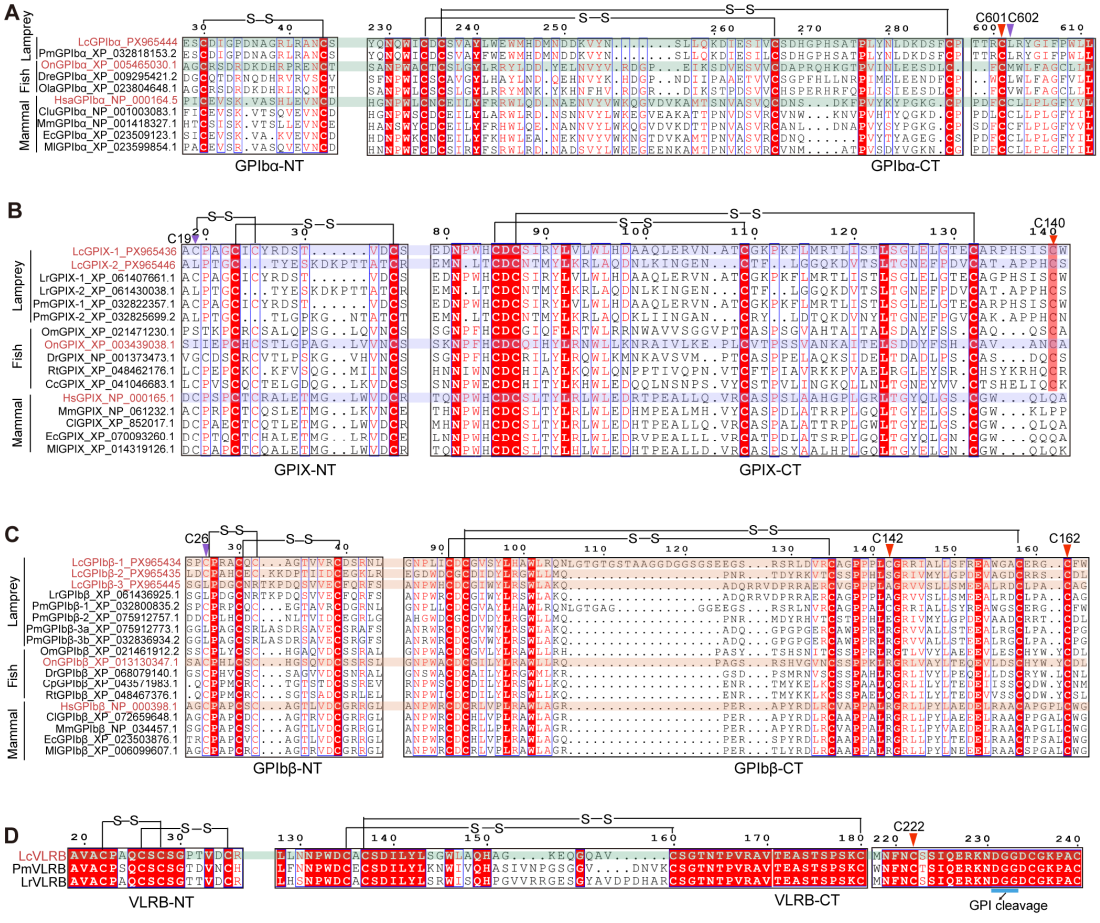
Extended Data Fig. 3 | Sequence identities, exon-intron phases, and distribution of the GPIbβ and GPIX subunits across jawless species. A-B, Similarity matrix heatmap of the GPIbβ and GPIX proteins. Similarity values are displayed within each box. The color scale of boxes ranges from blue (low similarity) to dark red (high similarity). Side bars in red and blue indicate subunits from hagfish and lamprey, respectively. **C,** Exon-intron phases of all subunits of the GPIb-IX complex in lamprey, hagfish, shark and zebrafish. Genes are depicted as light blue boxes with arrowheads showing their transcriptional direction along the chromosome. Gray boxes represent the UTR, untranslated region; Purple boxes mean CDS, coding region sequence; Black

dashed box in the *MgGp1bb-3* indicates that part of the gene sequence is missing; Black numbers above the gene structure indicate exon-intron phases, whereas gray numbers below represent the length distribution of CDS across exons. **D,** Scatterplots display the correlation among *LcGp1ba*, *LcGp9-2*, *LcGp1bb-2* and *LcGp1bb-3* expression. Spearman's rank correlation coefficient is shown, representing the strength and direction of the monotonic relationship between the two genes across cells. The corresponding *p*-value indicates the significance of the correlation. **E,** Flow cytometry scatter plot shows the overall PBMC population of *E. burgeri*. The proportions of different cell types are indicated by gating and displayed in the figure. **F,** Dot plot shows the expression pattern of all *EbGp1bb* and *EbGp9* genes in the hagfish.

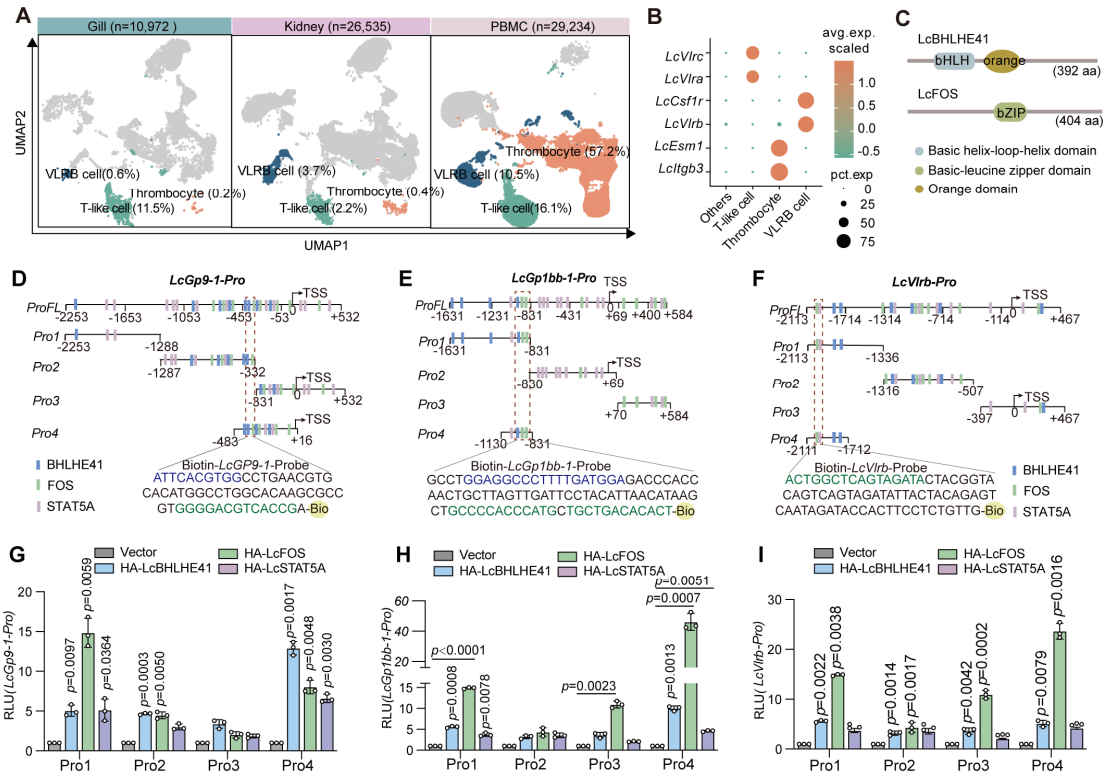


Extended Data Fig. 4 | Endogenous validation showing the presence of the VLRB-GPIbβ-GPIX complex in native VLRB cells. A, Efficiency of magnetic bead-based sorting measured by FACS analysis. Left panel shows fraction of VLRB antibody-positive cells in the flow-through, while right panel shows the fraction of VLRB antibody-positive cells recovered in the elution. **B,** Detection of specificity and titers for the LcGPIbβ-1 polyclonal antibody. VLRB$^-$ refers to cells collected from the flow-through during magnetic bead-based selection, while VLRB$^+$ refers to the positive cells obtained via magnetic bead-based selection. **C,** Validation of VLRB antibody specificity by Western blot in the PBMCs. Western blot analysis of whole cell lysates (WCL) and immunoprecipitated (IP) samples was used to validate antibody specificity, with proteins pulled down by mouse IgG used as a negative control. **D,** Silver staining of immunoprecipitated proteins for mass spectrometry analysis. Multiple gel slices, as indicated by the numbers in the figure, were collected to avoid contamination from

immunoglobulin heavy and light chains. H, heavy chains; L, light chains. **E,** The peptide information of LcLYN and LcSYK detected by mass spectrometry. **F,** Subcellular localization of the LcGPIbβ-1. HA-tagged LcGPIbβ-1 using its native (top two panels) or the human Igκ signal peptide (bottom two panels) were expressed in HEK293T cells. Cells were fixed and subjected to immunofluorescence using an anti-HA antibody(green). Nuclei were counterstained with DAPI (blue), and the endoplasmic reticulum was marked by anti-Calnexin antibody staining (red). The scatter-bar plot on the right shows the distribution of anti-HA-Calnexin colocalization coefficients across multiple fields. The red text indicates the mean coefficient. SP, signal peptide. **G,** Predicting features of the ITAM motifs of LcGPIbβ-1. The top panel exhibits the phosphorylation scores of predicted tyrosine sites. The bottom panel displays the predicted intrinsic disorder score of LcGPIbβ-1 using ANCHOR or IUPred. The medium shows the domain structure of LcGPIbβ-1 protein with residue positions corresponding to those indicated in the upper and lower panels.

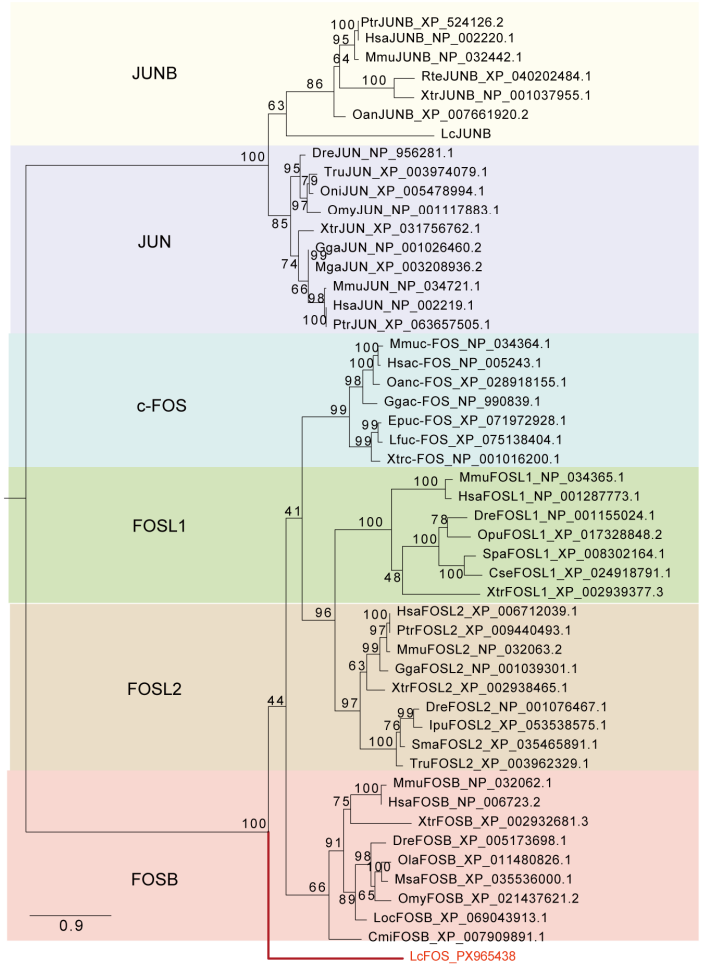


Extended Data Fig. 5 | Multiple sequence alignment of the protein sequences of GPIb-IX and VLRB focusing on the N-terminal and C-terminal regions. A-D, Multiple sequence alignment of GPIbα, GPIX, GPIbβ and VLRB focusing on the NT and CT. Red triangle indicate positions where key cysteine residues are required for hemostatic GPIb-IX complex or VLRB-GPIbβ-GPIX complex formation. Purple triangles indicate positions at which cysteine residues have been gained or lost in certain species relative to mammals. Blue box indicates the most probable GPI cleavage sites for VLRB proteins in the lamprey. The sequence names highlighted in red represent the species sequences used in the Fig. 3J-M.

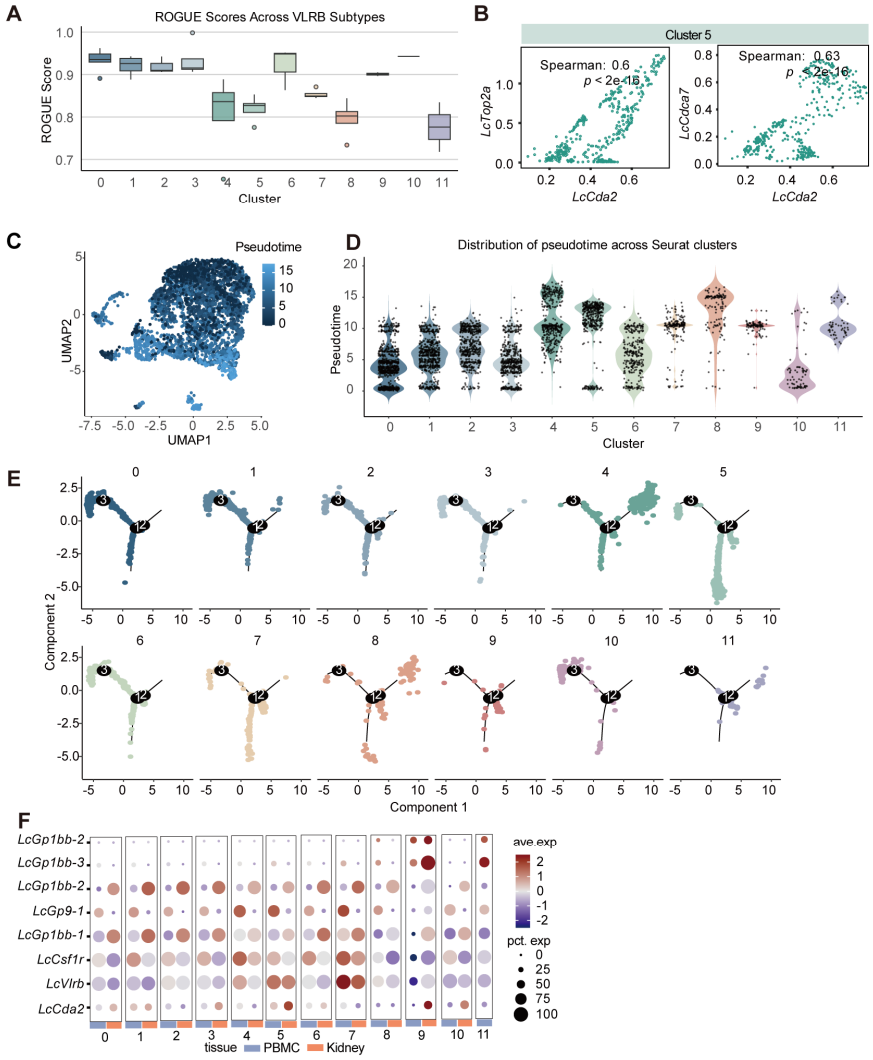


Extended Data Fig. 6 | Luciferase reporter assay for LcBHLHE41, LcFOS and LcSTAT5A regulations of *LcGp1bb-1*, *LcGp9-1* and *LcVlrb* promoters. A, UMAP visualization of integrated single-cell datasets from the gill, kidney and PBMC. Interest cells are colored according to their annotated cell type. **B,** Dot plots of curated marker genes across integrated datasets cell type. **C,** Structural domains architecture of LcBHLHE41 and LcFOS. **D-F,** Schematic of the full-length and truncated *LcGp9-1*, *LcGp1bb-1* and *LcVlrb* promoter used in the study. The probes used in competition assay were shown in the bottom, texts with blue and green colors correspond to BHLHE41 and FOS binding motif. **G-I,** Luciferase activity analyses in HEK293T cells transfected with a luciferase reporter for the *LcGp9-1*, *LcGp1bb-1* and *LcVlrb* with truncted promoter together with three HA-tagged selected TFs. Data are presented as

mean ± SD of three independent biological replicates (n = 3). Statistical significance between groups was determined using two-tailed unpaired Welch's t-test.



Extended Data Fig. 7 | Phylogenetic tree of the FOS-associated genes. Colored blocks indicate different molecular subgroups, and numbers at nodes represent bootstrap support values. LcFOS investigated in this study was highlighted in red.



Extended Data Fig. 8 | Supplementary analysis of developmental trajectories underlying VLRB heterogeneity. A, Purity evaluation of VLRB cell subpopulations, with each point representing a sample. The center line indicates the median ROUGE value. The lower and upper hinges represent the 25th and 75th percentiles, respectively. **B,** Scatter plots displaying the correlation of *LcCda2* with *LcTop2a* and *LcCdca7*. Spearman's rank correlation coefficient is shown, representing the strength and direction of the monotonic relationship between the two genes pairs in VLRB subset cluster 5. The corresponding *p*-value indicates the significance of the correlation. **C,** UMAP visualization of cells colored according to inferred pseudotime. **D,** Violin plot

displaying the pseudotime progression of cells within each subpopulation. **E,** The facet plot presenting the pseudotime trajectories for each VLRB subpopulation. **F,** Dot plot showing selected molecule expression across VLRB subpopulations between kidney and PBMC tissue.