



# Diversity and functional specialization of oyster immune cells uncovered by integrative single cell level investigations

Sébastien de la Forest Divonne, Juliette Pouzadoux, Océane Romatif, C. Montagnani, Guillaume Mitta, Delphine Destoumieux-Garzon, Benjamin Gourbal, Guillaume Charrière, Emmanuel Vignal

## ► To cite this version:

Sébastien de la Forest Divonne, Juliette Pouzadoux, Océane Romatif, C. Montagnani, Guillaume Mitta, et al.. Diversity and functional specialization of oyster immune cells uncovered by integrative single cell level investigations. 2024. hal-04782332

**HAL Id: hal-04782332**

**<https://hal.science/hal-04782332v1>**

Preprint submitted on 14 Nov 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Title

Diversity and functional specialization of oyster immune cells uncovered by integrative single cell level investigations

# Authors

Sébastien de La Forest Divonne<sup>1</sup>, Juliette Pouzadoux<sup>1</sup>, Océane Romatif<sup>1</sup>, Caroline Montagnani<sup>1</sup>, Guillaume Mitta<sup>2</sup>, Delphine Destoumieux-Garzon<sup>1</sup>, Benjamin Gourbal<sup>3</sup>, Guillaume M. Charrière<sup>1#</sup>, Emmanuel Vignal<sup>1#\*</sup>

# Affiliations

<sup>1</sup> IHPE, Univ Montpellier, CNRS, Ifremer, Univ Perpignan Via Domitia, Montpellier, France

<sup>2</sup> Ifremer, IRD, Institut Louis-Malardé, Univ Polynésie française, UMR 241 SECOPOL, 98179 Taravao, Tahiti - Polynésie française, France

<sup>3</sup> IHPE, Univ Montpellier, CNRS, Ifremer, Univ Perpignan Via Domitia, Perpignan, France

# EV and GC co-supervised the work.

\* corresponding author: [emmanuel.vignal@umontpellier.fr](mailto:emmanuel.vignal@umontpellier.fr)

# Abstract

Mollusks are a major component of animal biodiversity and play a critical role in ecosystems and global food security. The Pacific oyster, *Crassostrea (Magallana) gigas*, is the most farmed bivalve mollusk in the world and is becoming a model species for invertebrate biology. Despite the extensive research on hemocytes, the immune cells of bivalves, their characterization remains elusive. Here we were able to extensively characterize the diverse hemocytes and identified at least seven functionally distinct cell types and three hematopoietic lineages. A combination of single-cell RNA sequencing, quantitative cytology, cell sorting, functional assays and pseudo-time analyses was used to deliver a comprehensive view of the

distinct hemocyte types. This integrative analysis enabled us to reconcile molecular and cellular data and identify distinct cell types performing specialized immune functions, such as phagocytosis, reactive oxygen species production, copper accumulation, and expression of antimicrobial peptides. This study emphasized the need for more in depth studies of cellular immunity in mollusks and non-model invertebrates and set the ground for further comparative immunology studies at the cellular level.

## Introduction

*Mollusca* is the second largest invertebrate phylum, after Arthropoda, and the largest marine phylum, comprising approximately 23 % of all known marine organisms (1). Among them, bivalves exhibit a high diversity and a rich evolutionary history (2). The Pacific oyster *Crassostrea (Magallana) gigas* (*C. gigas* - Thunberg, 1793) (NCBI:txid29159) is a sessile filter-feeding bivalve that thrives in a variety of stressful environments ranging from intertidal to deep-sea conditions (3). It is a key species for the aquaculture industry worldwide (4). Several infectious diseases affect *C. gigas* at different life stages, which impacts its production. Given the significant socio-economic value of this species, there has been an increased focus on understanding and mitigating these diseases (5). The causes of these mortalities can involve a variety of pathogens, including viruses, bacteria and parasites that can be responsible for the mortality events affecting *C. gigas* (6). One of the most extensively researched infectious diseases is POMS (Pacific Oyster Mortality Syndrome), a polymicrobial disease responsible for mass mortalities of juvenile oysters (7). The disease is triggered by the OsHV-1  $\mu$ Var herpesvirus, which alters the immune defenses of oysters, allowing the colonization of opportunistic bacteria, including *Vibrio*, that cause hemocyte lysis and bacteremia, ultimately leading to animal death (7). Other bacterial pathogens have also been identified as a contributing factor in mass mortalities of adult Pacific oysters in several countries. The most notable is *Vibrio aestuarianus* which affects adult oysters in Europe (8, 9). To date, the majority of oyster pathogens or opportunistic pathogens that have been

characterized in detail have been found to subvert hemocyte defenses for their own benefit. These include the OsHV-1  $\mu$ Var herpesvirus (10) and virulent *Vibrio* strains of the species *Vibrio crassostreae* and *Vibrio tasmaniensis* (11), *Vibrio aestuarianus* and *Vibrio harveyi* (12), highlighting the critical role of hemocytes in oyster immunity. The development of immune-based prophylactic treatments, such as immune-priming and immune-shaping, represents a promising avenue for enhancing the natural defenses of oysters against pathogens and increasing their survival rate. Nevertheless, the advancement of such therapies is still constrained by a dearth of knowledge regarding the underlying molecular and cellular mechanisms (13, 14).

The study of hemocytes has a long history, dating back to the 1970s (for review see (15)). Hemocytes are cellular effectors of the immune system. They engage in phagocytosis to engulf and destroy potential pathogens, neutralizing parasites by encapsulation, or preventing pathogen dissemination by cell aggregation and the release of extracellular DNA traps (16). Furthermore, they engage in the humoral response by releasing cytokines, antimicrobial peptides, and reactive oxygen species (ROS), which enable them to combat pathogens (17). In addition to their role in oyster immunity, hemocytes have been implicated in numerous physiological processes, including shell repair (18), wound healing, nutrient transport, and environmental contaminant removal (19). Despite this acquired knowledge, hemocytes remain an under-characterized population of circulating immune cells. The lack of a unified classification and of molecular and functional genetic tools hinders our understanding of lineage ontogeny and functional specialization. Several studies have proposed different classifications of hemocytes in the *Ostreidae* family, with 3 to 4 hemocyte types reported (15). These classifications are primarily based on either microscopic or flow cytometry analyses. In *C. gigas*, three primary hemocyte cell types have been classically identified : blast, hyalinocyte, and granulocyte cells. While the immune response of *C. gigas* has been extensively studied using classical transcriptomics at the whole animal, tissue, or circulating hemocyte levels, these approaches have failed to consider the diversity of hemocyte cell types

82 and lineages that underpin these responses (20). However, it is still imperative to accurately  
83 describe the diversity of these cells, understand their ontogeny, and delineate cell lineages to  
84 comprehend their specific roles. The advent of single-cell RNA sequencing (scRNA-seq)  
85 techniques has enabled the monitoring of global gene expression at the single-cell level with  
86 thousands of individual cells in a single experiment. This provides a unique opportunity to  
87 overcome these limitations and deepen our understanding of hemocyte diversity and function  
88 in bivalves. Recently scRNA-seq was used to provide a first molecular description of a  
89 hemocyte population in the oyster *Crassostrea hongkongensis* (21). However, in the absence  
90 of morphological and/or functional characterization studies, the authors could not deduce the  
91 hemocyte cell types to be matched to the transcriptomic profiles generated by scRNA-seq.  
92 While numerous transcriptomic analyses have been conducted on *C. gigas* hemocytes, none  
93 have adopted a single-cell approach. In this study, we present an integrative analysis of the  
94 diversity of *C. gigas* hemocytes at the single-cell level. To this end, we combined scRNA-seq  
95 from a pathogen-free adult oyster combined with cytological, cell fractionation and functional  
96 assays. This approach allowed us to create a comprehensive transcriptomic, cytological and  
97 functional atlas of hemocyte cell types. Our scRNA-seq analyses identified 7 distinct  
98 transcriptomic populations and functional annotation revealed distinct populations with  
99 specific functions, including phagocytosis, oxidative burst, energetic metabolism, enhanced  
100 transcription, translation and cell division. Quantitative cytology enabled the identification of  
101 7 morphologically distinct hemocyte cell types, which allowed us to reconcile molecular and  
102 cytological data. Density gradients were used to separate hemocyte cell types and qPCR or  
103 functional assay analyses were performed to validate cell type-specific markers. By employing  
104 this integrated approach, we could identify 1 type of hyalinocyte, 2 types of blasts and 4 types  
105 of granular cells. Furthermore, we identified cell types that perform antimicrobial functions  
106 through phagocytosis, ROS production, copper accumulation, and expression of antimicrobial  
107 peptides. Finally, trajectory analysis of scRNA-seq data combined with functional analysis  
108 revealed distinct differentiation pathways that may control hemocyte ontology and

differentiation processes. Based on these findings, we propose a more comprehensive and up-to-date classification of *C. gigas* hemocytes, with a more accurate description of the different cell types, their potential ontology and a precise description of their sub-functionalization.

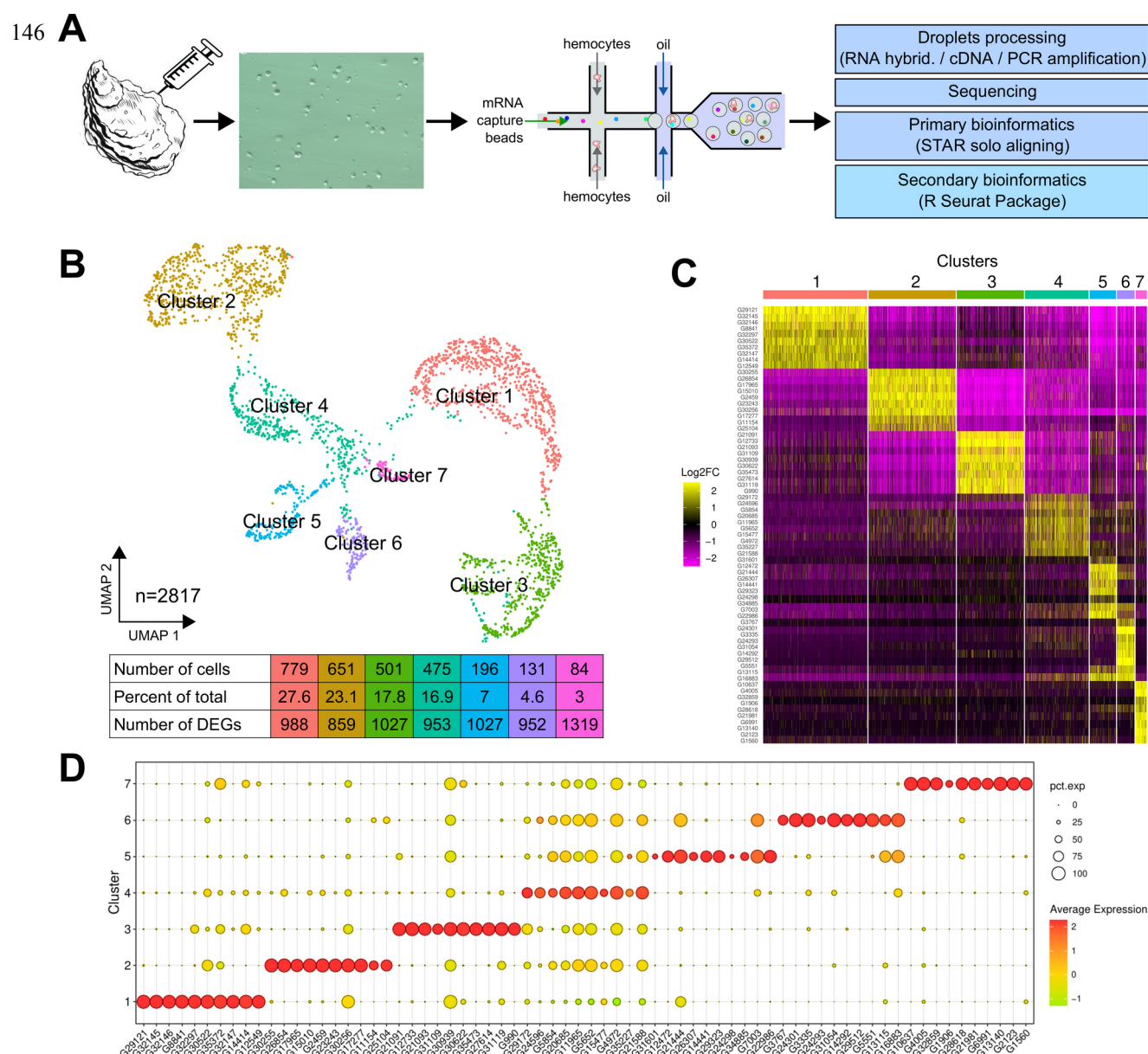
112

## 113 **Results**

### 114 **Single-cell RNA sequencing reveals 7 distinct transcriptomic clusters of circulating** 115 **immune cells in oysters.**

Oysters are known to exhibit a high degree of individual genetic polymorphism, including Copy Number Variation (CNV) and Presence Absence Variation (PAV) (22). To prevent misinterpretation of the single-cell transcriptomic data, and to characterize the hemocyte cell types and their heterogeneity, we sampled hemocytes from a unique pathogen-free animal (Ifremer Standardized Animal, 18-month-old) and applied single-cell drop-seq technology to 3,000 single hemocytes (**Fig. 1A**). The scRNA-seq library was generated and sequenced, resulting in 127,959,215 high-quality filtered reads available for single-cell analysis (ENA project accession number PRJEB74031). Primary bioinformatics analysis was performed using the STAR solo aligner software (23) against the *C. gigas* genome (Genbank reference GCA\_902806645.1) from the Roslin Institute (24). Of the 127,959,215 reads, 97 % showed a valid barcode and 89.2 % were successfully mapped to the genome with a saturation of 75.6 %. A total of 2,937 cells were profiled, yielding a median of 1,578 genes and 4,412 unique molecular identifiers (UMIs) per cell among the 23,841 total genes detected, with a sequencing saturation of 75.6 % (**Supp. Table 1**). Secondary bioinformatic data processing was conducted using the Seurat R package (version 4.3.0) (25). The data set was filtered to remove data corresponding to empty droplets or cell doublets. Cells with a gene number between 750 and 4,000 and less than 5 % mitochondrial genes were retained. After quality control processing, 120 cells corresponding to empty droplets and cell doublets were removed, and 2,817 cells were processed for data normalization. Finally, we performed linear dimensional reduction and clustering on the 3,000 most variable genes from 2,817 cells

(Supp. Fig. S1). Dimension reduction and clustering led to identifying 7 transcriptomic clusters, within which hemocytes were distributed. These 7 different clusters represented 27.6, 23.1, 17.8, 16.9, 7, 4.6 and 3 % of the total cells (Fig. 1B). For each transcriptomic cluster, a pattern of over- and under-represented transcripts in each cell was identified (Fig. 1C). Average Log2FC values and percentage of expression in each cluster relative to all other clusters were calculated for the ten most differentially expressed genes (Fig. 1D and Table 1). Clear transcriptomic cell clusters were detected as well as specific gene markers for each cluster (Supp. Data 1). Only the transcriptomic signature of cluster 4 was less contrasting than that of the other clusters (Fig. 1D).



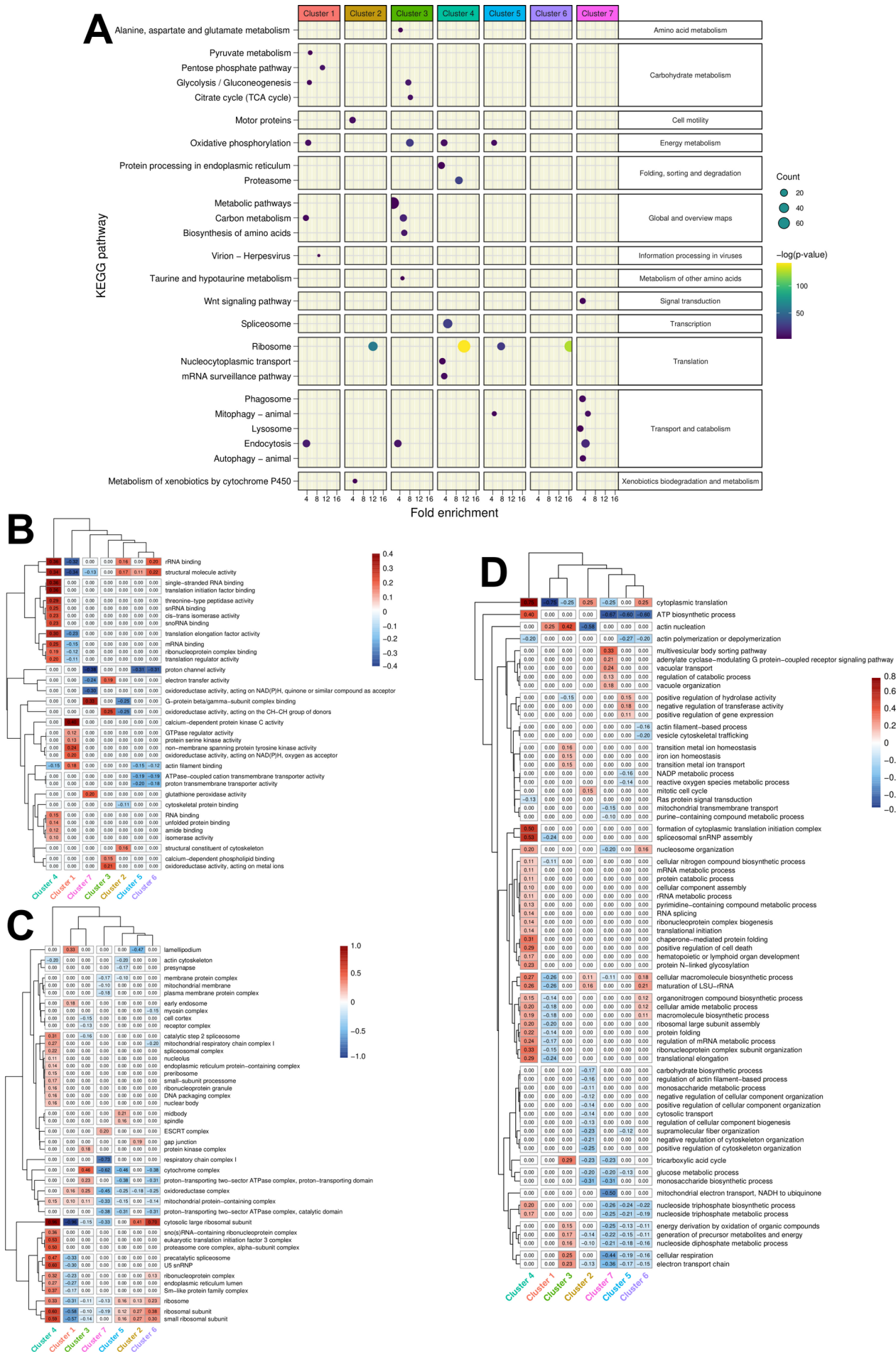
**Fig. 1. scRNA-seq analysis of *C. gigas* circulating hemocytes reveals 7 transcriptomic cell clusters** (A) Schematic of the scRNA-seq 10X Genomics Chromium microfluidic technology and bioinformatics processing workflow used. Dissociated hemocytes were collected from a pathogen-free oyster and encapsulated in droplets for processing. After sequencing, the data were processed bioinformatically. (B) Uniform Manifold Approximation and Projection (UMAP) plot for dimensional reduction of the data set and summary of cells and the number of Differentially Expressed Genes (DEGs) in each cluster. The table shows the characteristics (number of cells, percentage of total cells and number of Differentially Expressed Genes in each cluster) of the seven clusters identified. (C) Heatmap showing the top 10 overexpressed genes in each cell per cluster as determined by FindAllMarkers() function in Seurat, corresponding to clusters in UMAP plots from **Fig. 1B**, ranked by log2FC. (D) Dot plot representing the ten most enriched DEGs per cluster based on average expression (avg\_log2FC). The color gradient of the dot represents the expression level, while the size represents the percentage of cells expressing each gene per cluster.

Gene	log2FC	Pct 1	Pct 2	Description	Cluster
G29121	4.31	1	0.099	L-galactono-gamma-lactone oxidase	Cluster 1
G32145	4.18	0.992	0.079	Angiotensin-1 receptor	
G32146	3.90	0.985	0.063	Angiotensin-1 receptor	
G8841	3.13	0.988	0.049	C-type lectin domain-containing protein	
G32297	2.88	0.99	0.232	PNPLA domain-containing protein	
G30522	2.76	1	0.515	Cytochrome P450 2D8	
G35372	2.69	1	0.483	60S ribosomal protein L26	
G32147	2.45	0.969	0.131	Angiotensin-1 receptor	
G14414	2.45	1	0.343	DUF4773 domain-containing protein	
G12549	2.38	0.974	0.132	Glutaredoxin domain-containing protein	
G30255	5.31	1	0.15	Metalloproteinase inhibitor 3	Cluster 2
G26854	4.24	1	0.16	Stanniocalcin	
G17965	4.21	0.94	0.04	Complement C1q-like protein 2	
G15010	3.99	0.99	0.17	Kyphoscoliosis peptidase	
G2459	3.82	1	0.21	Uncharacterized protein	
G23243	3.51	0.98	0.07	Fibronectin type-III domain-containing protein	
G30256	3.23	1	0.69	NTR domain-containing protein	
G17277	3.21	0.99	0.04	Putative modulator of levamisole receptor-1	
G11154	3.16	0.68	0.05	Collagen alpha-1(XII) chain	
G25104	3.1	0.84	0.08	C1q domain-containing protein	
G21091	5.51	0.99	0.13	X-box binding protein-like protein	Cluster 3
G12733	5.42	1	0.12	Galectin	
G21093	5.17	0.94	0.05	IrfG	
G31109	4.36	0.77	0.03	Cystatin domain-containing protein	
G30939	4.25	1	0.86	Peptide ABC transporter permease	
G30622	3.95	1	0.21	BHLH domain-containing protein	
G35473	3.86	0.96	0.04	Uncharacterized protein	
G27614	3.78	0.96	0.13	G Protein Receptor F1-2 domain-containing protein	
G31119	3.77	0.99	0.42	Arrestin C domain-containing protein	
G990	3.6	0.93	0.03	LITAF domain-containing protein	
G29172	3.09	0.78	0.49	G Protein Receptor F1-2 domain-containing protein	Cluster 4
G24596	2.37	0.78	0.25	High mobility group protein DSP1	
G5854	1.8	0.64	0.38	N-acetyltransferase domain-containing protein	
G20685	1.59	0.88	0.53	Ribosomal protein L22	
G11965	1.57	0.92	0.7	Ribosome L4 associated C domain-containing protein	
G5652	1.54	0.97	0.75	60S ribosomal protein L30	
G15477	1.52	0.55	0.3	Metalloendopeptidase	
G4972	1.49	0.96	0.8	60S ribosomal protein L12	
G35227	1.46	0.55	0.06	ML domain-containing protein	
G21588	1.42	0.94	0.69	60S ribosomal protein L13	
G31601	5.24	0.38	0.06	Polyribonucleotide nucleotidyltransferase	Cluster 5
G12472	3.99	0.85	0.03	Galectin	
G21444	3.58	1	0.41	Uncharacterized protein	
G26307	3.42	0.6	0.02	Uncharacterized protein	
G14441	3.35	0.84	0.08	CLUBN	
G29323	3.21	0.9	0.01	GTPase IMAP family member 4	
G24298	3.18	0.29	0.05	Putative Transmembrane protease serine 9	
G34885	3.13	0.59	0.01	Gliding motility-associated C-terminal domain-containing protein	
G7003	2.55	0.99	0.16	ETS-related transcription factor Elf-4	
G22986	2.37	0.94	0.1	Metalloendopeptidase	
G3767	6.45	0.74	0.02	Natterin-1	Cluster 6
G24301	6.16	0.99	0.02	Aspartate-semialdehyde dehydrogenase	
G3335	5.19	1	0.16	ncRNA	
G24293	4.57	0.59	0.01	Uncharacterized protein	
G31054	3.56	1	0.13	GATA-binding factor 3	
G14292	3.29	0.87	0.03	Membrane-associated guanylate kinase inverted 3	
G29512	3.2	0.99	0	Caveolin	
G5551	2.93	0.97	0.02	Neuronal acetylcholine receptor subunit alpha-6	
G13115	2.73	0.83	0.26	ncRNA	
G16883	2.62	1	0.2	Allograft inflammatory factor	
G10637	4.21	0.99	0.03	BZIP domain-containing protein	Cluster 7
G4005	4.07	0.99	0.21	Ras-related protein Rab-20	
G32859	3.9	0.93	0.04	G Protein Receptor F1-2 domain-containing protein	
G1906	3.82	0.5	0.02	Glutathione peroxidase	
G28618	3.53	0.91	0.09	Sporozoite and liver stage asparagine-rich protein	
G21981	3.5	0.94	0.05	ncRNA	
G6991	3.4	0.85	0.01	Rho-GAP domain-containing protein	
G13140	3.12	1	0.03	DBH-like monoxygenase protein 1	
G2123	3.05	0.94	0.02	SPRY domain-containing SOCS box protein 3	
G1560	3.02	0.95	0.17	Glutathione peroxidase	

**Table 1. Top 10 overexpressed genes identified in each transcriptomic cluster.** The first column indicates the gene number according to the annotation. ‘log2FC’ represents the log2 fold change of the gene in the cluster compared to all other cells. ‘Pct1’ is the percentage of cells expressing the gene in the cluster and ‘Pct2’ is the fraction of cells expressing the gene in all other clusters. The description is the annotation of the expressed gene.

# **KEGG pathways and GO-terms analyses reveal functional diversity in *C. gigas* hemocytes.**

The ScRNA-seq data demonstrated that specific functions are carried out by the hemocyte cell types that comprise the seven transcriptomic clusters. A preliminary overview of the functions over-represented in each cluster was obtained through a KEGG pathway analysis on the overrepresented transcripts (Log2FC > 0.25) using the *C. gigas* annotation provided by the DAVID consortium (26), thereby identifying pathways specifically enriched in each cluster. Cluster 1 demonstrated enrichment in viral processing and endocytosis (**Figure 2A and Supplemental Table S2**). Additionally, it demonstrated enrichment in pyruvate metabolism, glycolysis/gluconeogenesis and the pentose phosphate pathways. Clusters 1 and 3 exhibited a distinctive enrichment in carbohydrate metabolism and endocytosis. In particular, cluster 3 exhibited enriched transcripts associated with glycolysis/gluconeogenesis and TCA cycle activities. Cluster 4 was enriched in protein synthesis, including transcription (spliceosome, ribosome, nucleocytoplasmic transport and mRNA surveillance pathway) folding, sorting and degradation of protein pathways. Clusters 2, 5 and 6 exhibited a shared signature of enrichment in ribosome-related genes. Cluster 2 demonstrated a specific enrichment in motor protein-coding genes responsible for cell motility and xenobiotic metabolism. Remarkably, clusters 1, 3, 4 and 5 exhibited enriched oxidative phosphorylation transcripts, whereas the transcripts of cluster 7 were enriched in vesicular trafficking and endo-lysosomal pathways (endocytosis, endosome, phagosome, lysosome, auto and mitophagy).



**Fig. 2. KEGG and Gene Ontology analysis of the gene signature in each cluster. (A)** A synthetic representation of the KEGG pathway analysis is shown. Colored columns represent the 7 transcriptomic clusters. Each row is a KEGG pathway, the colored dot represents the  $-\log(p\text{-value})$  and the dot size represents the number (count) of enriched genes in each pathway category. The fold enrichment is shown on the x-axis. Panels **(B)**, **(C)** and **(D)** show the results of Gene Ontology terms (GO-terms) for Biological Processes (BP), Cellular Components (CC) and Molecular Functions (MF) respectively, obtained with the overexpressed genes of each cluster (Absolute value  $\text{Log}_2\text{FC} > 0.25$  and significant  $p\text{-value} < 0.001$ ) using RBGOA analysis ( $p\text{-value} \leq 0.001$ ) for three different ontology universes. Each panel corresponds to one ontology universe, and the analysis highlights over- and under-enriched terms. The number in the heatmap and the scale indicate the proportion of significant positive GO-terms. Positive values indicate over-enrichment and negative values indicate under-enrichment of the respective BP, CC and MF ontologies.

For a more detailed functional characterization of each transcriptomic cluster, we performed a re-annotation of the *C. gigas* genome using a combination of tools to enhance the GO term richness of the existing annotation prior to functional GO term analysis. To this end, we used the *C. gigas* genome (Genbank reference GCA\_902806645.1) and the associated gff3 annotation file from the Roslin Institute (24). These files were used to extract and process the longest CDSs for GO-term annotation using the Orson pipeline (see Materials and Methods). Of the 30,724 extracted CDSs, 22,462 were annotated (GO-terms and sequence description), yielding an annotation percentage of 73.1 %. Of the 30,724 CDSs, 22,391 were annotated with Molecular Functions (MF), Biological Processes (BP), and Cellular Components (CC) GO-terms (**Supp. Fig. S2 and Supp. Data 2**). Using the GO-term annotation and the  $\text{Log}_2\text{FC}$  of genes calculated after scRNA-seq processing in each cluster, GO enrichment analysis was performed by rank-based gene ontology analysis (RBGOA) (27). RBGOA analysis was

performed on the GO-terms identified in each cluster (**Supp. Data 3**). The results are presented in **Figures 2B, C and D**.

The scRNA-seq-based analysis identified seven distinct transcriptomic profiles for each cell cluster, thereby shedding light on greater heterogeneity and functional diversity of *C. gigas* hemocytes than previously described. Cluster 1, comprising 27.6 % of cells, is characterized by its morphology and capacity to remodel actin cytoskeleton, as well as oxidoreductase activity. This is evidenced by an enrichment in oxidoreductase activity acting on NAD(P)H and kinase activities BP, actin nucleation MF and lamellipodium and early endosomes CC. Cluster 2, comprising 23.1 % of cells, has an increased translation activity, as indicated by enrichment in BP, MF and CC terms related to rRNA binding, cytoplasmic translation, maturation of LSU-rRNA, and ribosomes respectively. Cluster 3, representing 17.8 % of cells, shows enrichment in cellular oxidation and actin nucleation, as evidenced by an enrichment in BP and MF related to oxidoreductase activities acting on metal ions, electron transfer, cellular oxidation and actin nucleation, and CC related to cytochromes, oxidoreductase complex, and mitochondrial protein-containing complexes. Cluster 4, representing 16.9 % of cells, has transcriptomic signatures reminiscent of spliceosome assembly, rRNA maturation, and ATP biosynthesis (BP related to RNA binding, translation, and proteasome, MF related to spliceosome assembly, rRNA maturation, and ATP biosynthesis, and CC related to nuclear bodies and ribosomes). Cluster 5, comprising 7 % of cells, demonstrates enrichment in structural molecule activity, hydrolase regulation, and gene expression, as evidenced by its BP, MF, and CC related to ribosome, midbody, and spindle localization. Cluster 6, comprising 4.6 % of cells, is characterized by a specific nucleosome organization, translation, and biosynthetic processes, with related BP, MF, and CC tied to ribosomes, ribonucleoprotein complex and rRNA binding. Cluster 7, representing a mere 3 % of cells, is characterized by multivesicular body sorting, vacuolar transport, vacuole organization, and G protein-coupled receptor signaling, with related BP, MF, and CC of the ESCRT machinery.

## 244 **Seven morphologically distinct immune cell types are identified by quantitative cytology** 245 **and transcriptomic markers.**

246 Cytological studies were conducted using MCDH staining to better characterize the diversity  
247 of circulating hemocytes in *C. gigas*. Seven distinct hemocyte morphotypes were identified: 3  
248 non-granular (acidophilic blasts, basophilic blasts and hyalinocytes) and 4 granular (small  
249 granule cells, big granule cells, vesicular cells and macrophage-like) (**Fig. 3 and Supp. Fig.**  
250 **S3**). Hyalinocytes (30 % of total hemocytes) are large cells with an irregular spreading  
251 membrane. They contain an azurophilic cytoplasm without granulations and their irregular  
252 nucleus varies in size (**Fig. 3A panel H**). Macrophage-like cells (19 % of the hemocytes)  
253 present an irregular membrane punctuated by rare pseudopodia with a polylobed nucleus and a  
254 basophilic cytoplasm where polychromatic inclusions of various sizes could be observed (**Fig**  
255 **3A. panel ML**). Basophilic blasts characterized by a basophil cytoplasm (**Fig. 3A panel BBL**)  
256 and acidophilic blasts with acidophil cytoplasm (**Fig. 3A panel ABL**) accounted for 18 % and  
257 15 %, respectively, of the total hemocytes. They are rounded cells without granulation, with a  
258 uniform and regular dense nucleus and a high nucleo-cytoplasmic ratio. Small granule cells  
259 (12 % of total hemocytes) have an irregular membrane punctuated by rare pseudopodia, an  
260 acidophilic cytoplasm, and numerous homogeneous purple granules (**Fig. 3A panel SGC**).  
261 Big granule cells (4 % of total hemocytes) are rounded cells with a basophilic cytoplasm  
262 containing large, slightly dark purple to black vesicles of heterogeneous size (**Fig. 3A panel**  
263 **BGC**). Vesicular cells (2 % of total hemocytes) are rounded cells with acidophilic cytoplasm,  
264 rich in homogeneous transparent and fluorescent to UV light vesicles with an irregular nucleus  
265 (**Fig. 3A panel VC**). Only blasts (BBL, ABL) have a nucleo-cytoplasmic ratio less than 1,  
266 whereas the other hemocytes described have a nucleo-cytoplasmic ratio greater than 1.  
267 To gain further insight into the functional characterization of these diverse hemocyte  
268 morphotypes and to associate them with the transcriptomic clusters identified in the scRNA-  
269 seq, we enhanced an existing hemocyte fractionation approach using an isopycnic Percoll  
270 density gradient to sort the hemocytes (28). Cell sorting was performed using a discontinuous

bioRxiv preprint doi: <https://doi.org/10.1101/2024.07.19.604245>; this version posted August 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

271 Percoll gradient with densities ranging from 1.0647 to 1.1049. Seven distinct density fractions

272 were established (**Fig. 3B**) and the hemocyte composition was then characterized by

273 cytological analysis of each fraction (**Fig. 3A**; see also **Supp. Fig. S4** for statistical

274 significance). The uneven distribution of hemocyte morphotypes along the density gradient

275 enabled a relative separation of the different cell populations (**Supp. Fig S4**). In summary,

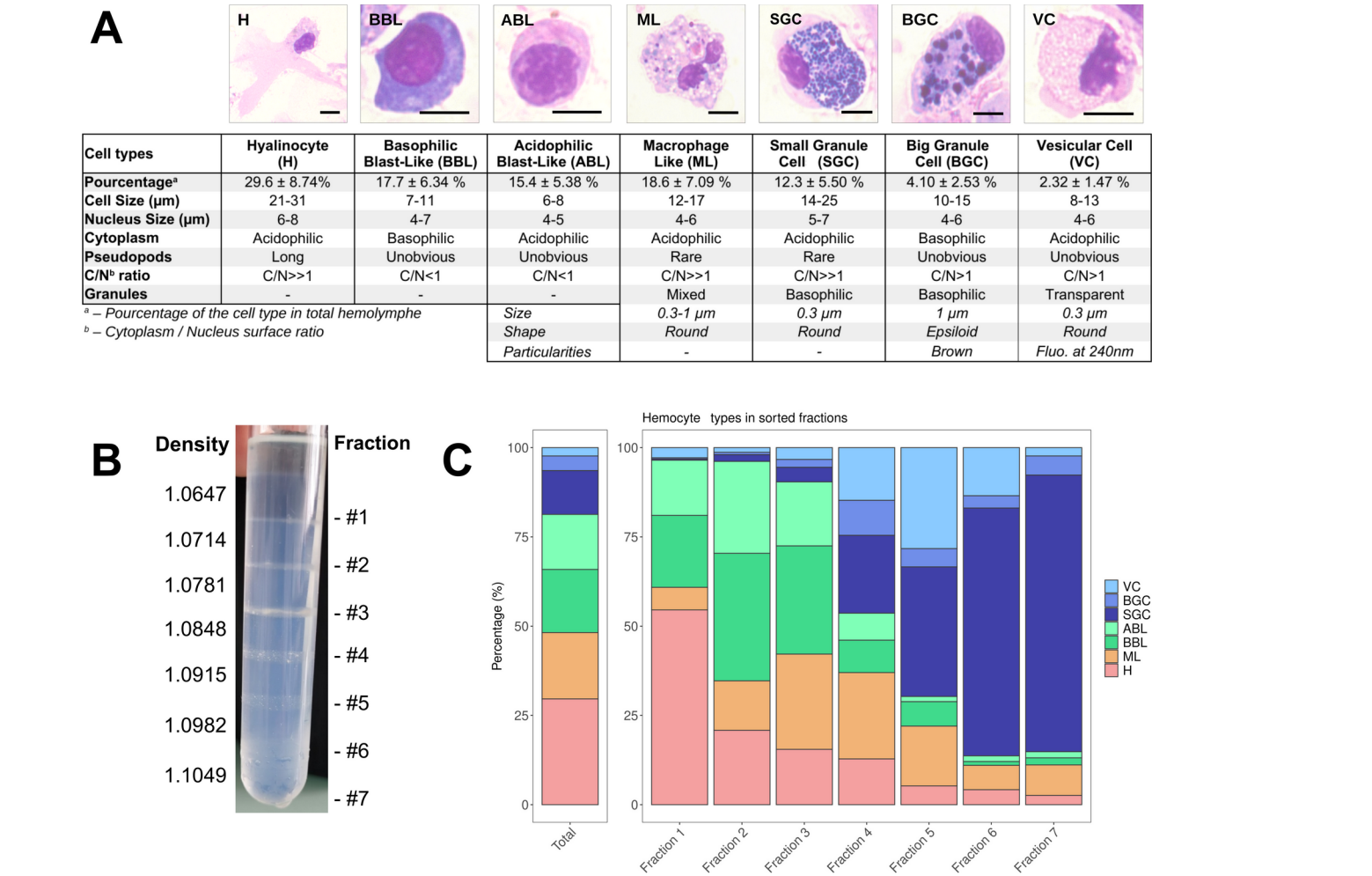
276 hyalinocytes were significantly enriched in the first fraction, while macrophage-like cells were

277 significantly enriched in fraction 3 and fraction 4. Acidophilic blasts were significantly

278 enriched in fraction 2. Basophilic blasts were significantly enriched in fractions 2 and 3.

279 Vesicular cells were enriched in fraction 5, big granule cells were enriched in fraction 4 and

280 small granule cells in fractions 6 and 7 (**Fig. 3C and Supp. Fig. S4**).



282 **Fig 3. *C. gigas* naive hemocyte formula and Percoll gradient hemolymph fractionation**

283 **(A)** Morphology, percentages and characteristics of the 7 cell types identified by MCDH

284 staining. **H** : Hyalinocyte, **ML** : Macrophage Like, **BBL** : Basophilic Blast Like cell, **ABL** :

285 Acidophilic Blast Like cell, **SGC** : Small Granule Cell, **BGC** : Big Granule Cell, **VC** :  
286 Vesicular Cell. Scale bar : 5  $\mu$ m. Hyalinocytes (54 %) and blast cells (ABL & BBL) (35 %)  
287 were predominant in the first fraction. The second fraction showed an increase in blasts (ABL  
288 26 %, BBL 36 %) and a decrease in hyalinocytes (21 %). Fraction 3 had a mixed content with  
289 a decrease in hyalinocytes and blasts (15 %), and a majority of blasts and macrophage-like  
290 cells. Fraction 4 had an increase in granular cells (SGC 22 %, BGC 10 %, VC 15 %) and a  
291 decrease in blasts (ABL 8 %, BBL 9 %). Fraction 5 showed an increase in small granule cells  
292 (36 %) and vesicular cells (28 %), and a decrease in big granule cells ( 5%). Fraction 6 had  
293 fewer hyalinocytes, macrophage-like cells, and blast cells compared to small granule cells and  
294 vesicular cells. The last fraction consisted mainly of small granule cells (81 %). **(B)** Sorting of  
295 hemocytes on a discontinuous Percoll gradient. 7 fractions were identified along the gradient  
296 at the top of each density cushion (from d=1.0647 at #1 to d= 1.1049 at #7). **(C)**  
297 Representation of the average values (from 5 different fractionation experiments) of the  
298 different hemocyte types in the seven percoll gradient fractions compared to the average  
299 hemolymph composition of a naive oyster (Total). **VC** : Vesicular Cells, **BGC** : Big Granule  
300 Cells, **SGC** : Small Granule Cells, **ABL** : Acidophilic Blast Like cells, **BBL** : Basophilic Blast  
301 Like cells, **ML** : Macrophage Like cells and **H** : Hyalinocytes respectively. **(Supp. Fig. S4** for  
302 statistics).

303 To identify the transcriptomic cell clusters corresponding to the different hemocyte  
304 morphotypes, we used RT-qPCR to detect the expression of different cluster-specific marker  
305 genes in the different hemocytes in the Percoll density fractions. The marker genes were  
306 selected without any a priori assumptions based solely on their expression levels (Log2FC)  
307 and their percentage of expression (Pct1 / Pct2 ratio) in the cluster of interest relative to all  
308 other cell clusters (**Table 2**). The expression level of each marker in the seven different  
309 clusters confirmed that the 14 selected marker genes were differentially expressed in each  
310 scRNA-seq transcriptomic cluster (**Supp. Fig. S5**).

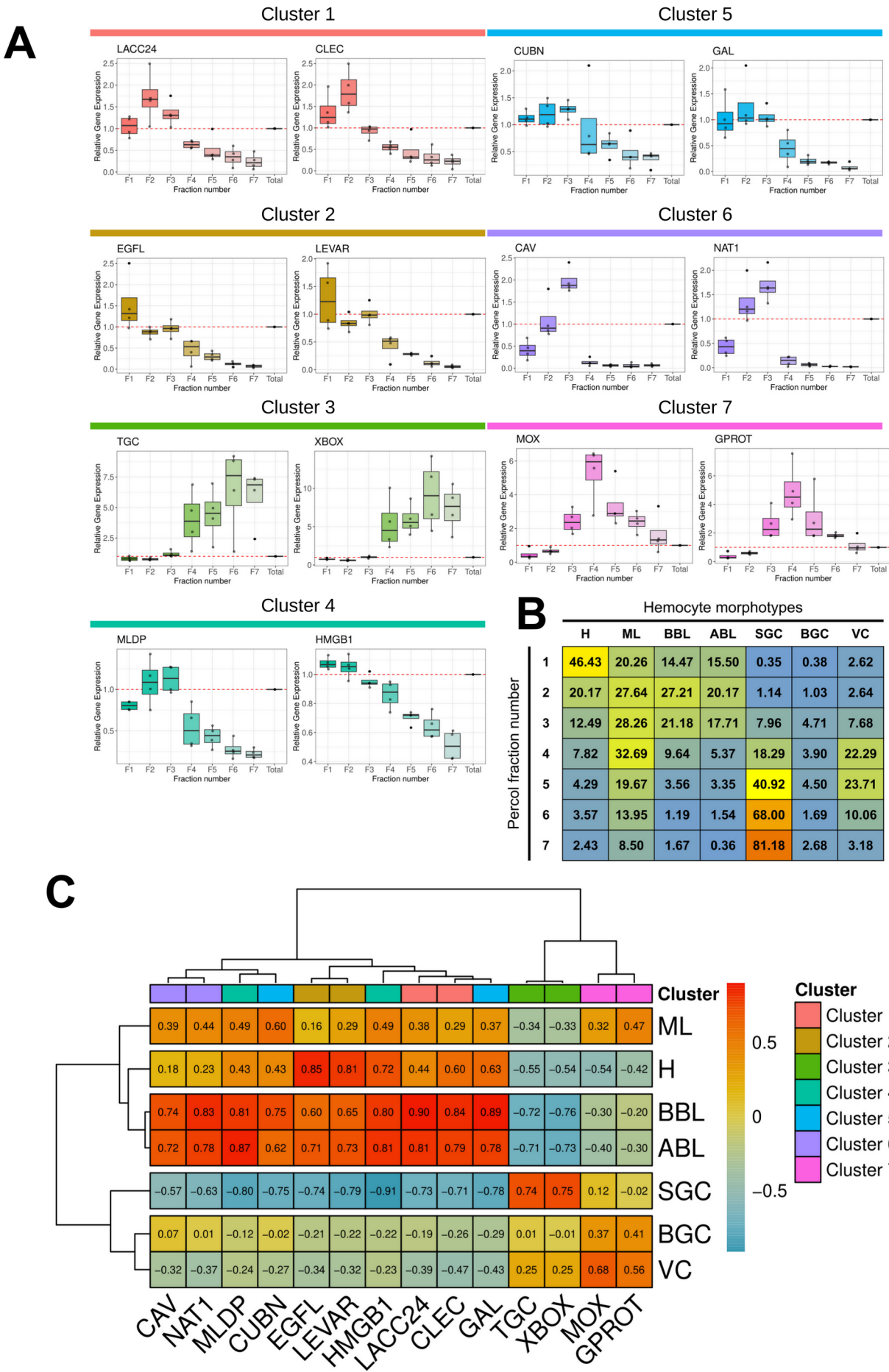
Gene Number	Avg. Log2FC	Pct.1/Pct.2 Ratio	Cluster	Description	Marker Name
G8994	2.05	38.20	1	Laccase-24	LACC24
G8841	3.13	20.16		C-type lectin domain-containing protein	CLEC
G24846	2.31	31.52	2	EGF-like domain-containing protein 8	EGFL
G17277	3.21	23.00		Putative modulator of levamisole receptor-1	LEVAR
G22387	3.16	16.76	3	TGc domain-containing protein	TGC
G21091	5.51	7.69		X-box binding protein-like protein	XBOX
G35227	1.46	9.98	4	ML domain-containing protein	MLDP
G31074	0.94	6.51		High mobility group protein B1	HMGB1
G14441	3.35	11.01	5	Cubilin	CUBN
G12472	3.99	24.91		Galectin	GAL
G29512	3.20	246.25	6	Caveolin	CAV
G3767	6.45	41.11		Natterin-1	NAT1
G13140	3.12	40.00	7	DBH-like monooxygenase protein 1	MOX
G32859	3.90	23.82		G protein receptor F1-2 domain-containing protein	GPROT

**Table 2. Table of the 14 marker genes specific to the different transcriptomic clusters.**

Gene number, average Log2FC, pct1/pct2 ratio (percentage of cells expressing this transcript in the cluster divided by the percentage of all other cells expressing this transcript) and cluster number are reported. The description is taken from our annotation and the marker name is derived from the description.

RT-qPCR expression profiles in the hemocyte fractions obtained from the Percoll density gradient revealed distinct patterns according to the different transcriptomic cluster marker genes (**Fig. 4A**). Cluster 1 markers (LACC24 and CLEC) were overexpressed in fractions 1, 2, 3 and underexpressed in the remaining fractions relative to total hemocytes. Cluster 2 markers (EGFL and LEVAR) showed a decreasing pattern of expression from fraction 1 to fraction 7. Cluster 3 markers (TGC and XBOX) showed a significant increase in expression in fractions 4 to 7. Cluster 4 markers (MLDP and HMGB1) showed a gradually decreasing expression pattern from fraction 2 to fraction 7. Cluster 5 marker genes (GAL and CUBN) were underexpressed in fractions 4 to 7, but not differentially expressed in fractions 1 to 3 compared to total hemocytes. Cluster 6 marker genes (CAV and NAT1) were overexpressed in fraction 3, expressed similarly to total hemocytes in fraction 2, and underexpressed in fractions 1, and 4 to 7. Finally, cluster 7 marker genes (MOX and GPROT) were overexpressed only in fractions 3 to 7. Correlation analysis and statistical validation

330 demonstrated a clear association between hemocyte morphotypes and 14 gene markers (**Fig.**  
 331 **4B, Fig. 4C and Supp. Table S4**). Cluster 3 marker genes (TGC and XBOX) correlated  
 332 positively ( $r = 0.74$  &  $0.75$ ) and specifically with small granule cells (SGC). Cluster 7 markers  
 333 (MOX and GPROT) correlated ( $r = 0.68$  and  $r = 0.56$ ) with vesicular cells (VC). Cluster 2  
 334 markers (EGFL and LEVAR) correlated positively with hyalinocytes (H) ( $r=0.85$  and  $r=0.81$ ).  
 335 No specific markers could be identified for blasts and macrophage-like cells.



337 **Fig 4. Characterization of molecular markers specific to the different hemocyte**

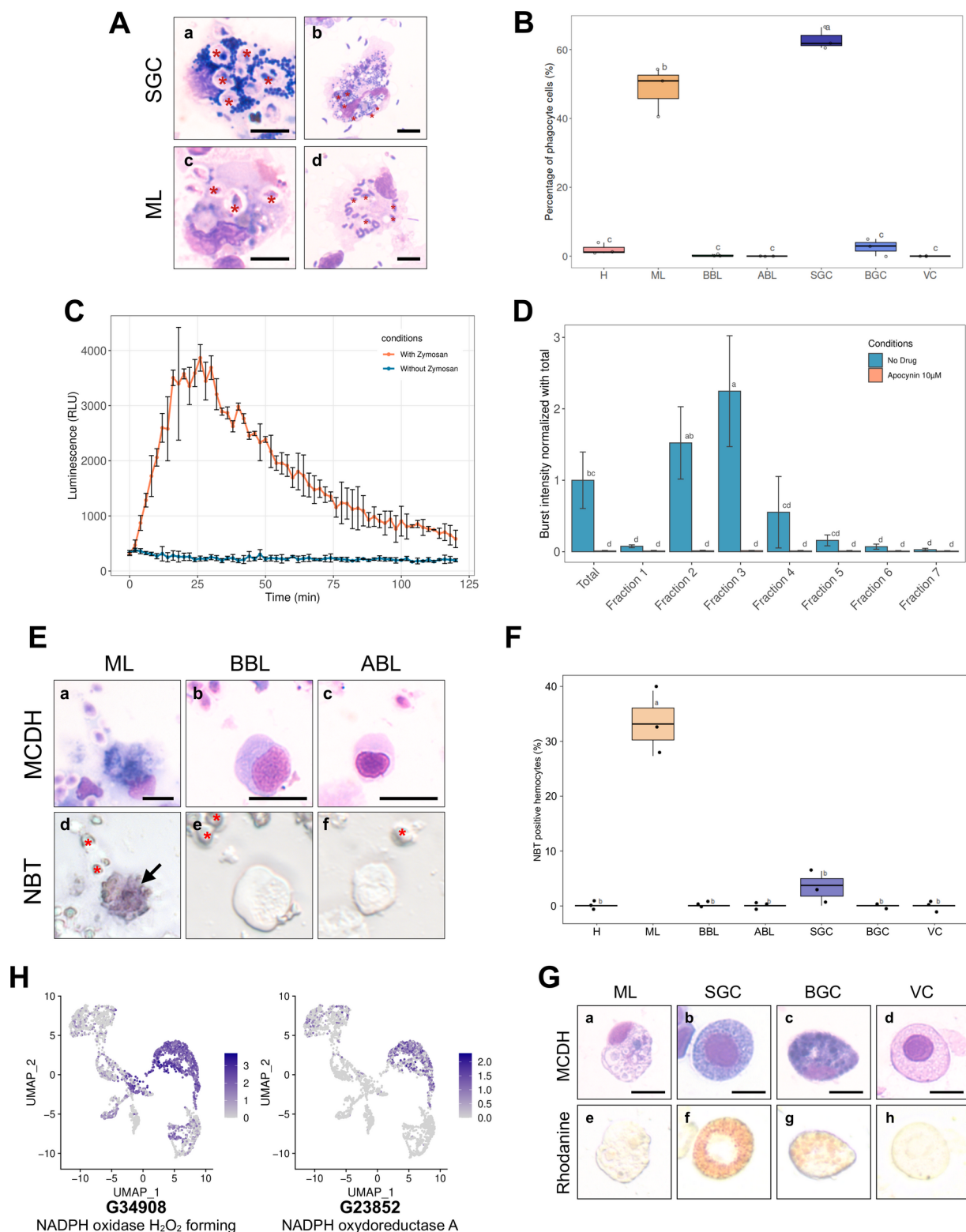
338 **morphotypes. (A)** Relative expression level of the 14 markers in the various fractions after

339 gradient density sorting. The graphs show the relative expression of genes compared to their  
 340 expression in total hemocytes in the various fractions (red dotted line). Standard deviations  
 341 were calculated based on four independent experiments. **(B)** Average percentage of each  
 342 hemocyte type in the 7 Percoll gradient fractions used to quantify marker gene expression by  
 343 qPCR. **(C)** Correlation matrix between the relative gene expression of each marker gene in  
 344 each fraction and the percentage of each hemocyte type in the same fractions. Values and  
 345 color scale represent the Pearson correlation coefficient (r) ranging from -1 (inverse  
 346 correlation) to +1 (full correlation). **H** : Hyalinocyte, **ML** : Macrophage Like, **BBL** :  
 347 Basophilic Blast Like cell, **ABL** : Acidophilic Blast Like cell, **SGC** : Small Granule Cell,  
 348 **BGC** : Big Granule Cell, **VC** : Vesicular Cell. **LACC24** : Laccase 24, **CLEC** : C-type lectin  
 349 domain-containing protein, **EGFL** : EGF-like domain-containing protein 8, **LEVAR** :  
 350 Putative regulator of levamisole receptor-1, **TGC** : TGc domain-containing protein, **XBOX** :  
 351 X-box binding protein-like protein, **MLDP** : ML domain containing protein, **HMGB1** : High  
 352 mobility group protein B1, **CUBN** : Cubilin, **GAL** : Galectin, **CAV** : Caveolin, **NAT1** :  
 353 Natterin-1, **MOX** : DBH-like monooxygenase protein 1, **GPROT** : G protein receptor F1-2  
 354 domain-containing protein.

355 By employing RT-qPCR, cell sorting on Percoll gradients, and scRNA-seq analysis, we could  
 356 identify cell types corresponding to three transcriptomic clusters. Cluster 3 corresponds to  
 357 small granule cells (SGC), cluster 2 to hyalinocytes (H), and cluster 7 to vesicular cells (VC).  
 358 Cluster 3, cluster 2 and cluster 7 represent 17.8 %, 23 % and 3 % of the total cells analyzed by  
 359 scRNA-seq, respectively. This is consistent with cytological data, which indicated the  
 360 presence of 12.5 % +/- 5 % Small Granule Cells, 30 % +/- 9.3 % Hyalinocytes and 2.3 % +/-  
 361 1.6 % Vesicular Cells (**Fig. 1B and Fig. 3A**). Furthermore, the molecular and cellular  
 362 functions derived from GO-term and KEGG analyses (**Fig. 2**) were in alignment with the  
 363 anticipated functions for these three cell types.

364 **Only macrophage-like and small granule cells behave as professional phagocytes.**

365 The hemocytes separated on the Percoll gradient were characterized functionally to gain  
366 further insight into their functional specialization. Two hemocyte functions, namely  
367 phagocytosis and production of Reactive Oxygen Species (ROS), are known to carry out  
368 major cellular antimicrobial activities. Granular cells have been suggested to be the  
369 professional phagocytes specialized for these functions (29). Phagocytosis and oxidative burst  
370 were studied using cell response toward zymosan particles (30).



**Fig 5. Phagocytosis, Reactive Oxygen Species production capacity and copper storage of hemocytes.** (A) Images of small granule cells (SGC) and macrophage-like (ML) cells with phagocytosed zymosan particles (panels a & c - red stars) and *Vibrio tasmaniensis* LMG20012<sup>T</sup> bacteria (panels b & d - red stars) from whole hemolymph sample. Scale bar : 5 μm. (B) Quantification of the phagocytic activity of zymosan particles by each cell type. The

graph shows the result of 3 independent experiments. **(C)** Luminescence recording to detect the production of Reactive Oxygen Species (ROS). In orange, a biphasic curve was obtained on naive oyster hemolymph after zymosan addition at  $t = 0$  min. In blue, the control condition corresponds to hemocytes without zymosan addition. **(D)** Graph showing the intensity of ROS production in each Percoll fraction. Normalized burst intensity was calculated from the luminescence peak obtained from each fraction. In blue, no drug was added to the experiment, in orange, ROS production was impaired by the addition of apocynin. **(E)** NBT (NitroBlueTetrazolium) staining of hemocytes exposed to zymosan particles. Hemocytes morphology after MCDH staining: Macrophage Like **(a)**, Basophilic **(b)** and Acidophilic **(c)** Blast cells. NBT staining of the different types of hemocytes **(d-f)**. Red stars show zymosan and bacteria particles. Black arrows indicate Macrophage-Like cells. Scale bar : 10  $\mu\text{m}$  **(F)** Quantification of NBT-positive cells present in the total hemolymph of oysters exposed to zymosan. **(H)** UMAP plots showing cells expressing NADPH oxidase found in the scRNA-seq dataset and their expression level. **(G)** Labeling of intracellular copper stores in *C.gigas* hemocytes. MCDH (upper panels) and rhodanine (lower panels) staining of oyster hemocytes to reveal copper accumulation. Scale bar : 10 $\mu\text{m}$ . For panels **(B)**, **(D)** and **(F)** the alphabetic characters displayed above the data points in each plot represent statistically significant differences among the groups, as determined by Tukey's test following ANOVA. Groups denoted by different letters differ significantly at the  $p < 0.05$  level of statistical significance.

**H** : Hyalinocytes, **ABL** : Acidophilic Blast-Like cells, **BBL** : Basophilic Blast-Like cells, **ML** : Macrophage-Like cells, **SGC** : Small Granule Cells, **VC** : Vesicular Cells and **BGC** : Big Granule Cells.

The phagocytic activity of hemocytes was first tested on a sample of total oyster hemolymph. Cells were incubated for 1 hour with either zymosan particles or the bacterial strain *Vibrio tasmaniensis* LMG20012<sup>T</sup>. Only the small granule cells and the macrophage-like cells exhibited efficient phagocytosis for both zymosan or vibrios, as observed after MCDH staining (**Fig. 5A**, **Supp. Fig. 6** and **Supp. Fig. 7A**). Macrophage-like cells and small granule

cells showed a phagocytic activity of 49 % and 55 %, respectively, and a phagocytosis index of 3.5 and 5.2 particles per cell respectively (**Fig. 5B and Supp. Fig. 7B**), as confirmed in 3 independent experiments. Very limited phagocytic activity was observed for hyalinocytes (1.7 %), basophilic blasts (0.18 %), and big granule cells (2.7 %) with a phagocytosis index of 2, 1, and 2.5 particles per cell, respectively (**Fig. 5B and Supp. Fig. 7B**). These results confirmed that only small granule cells and macrophage-like cells behave as professional phagocytes demonstrating robust phagocytic activity.

# **Only macrophage-like cells produce Reactive Oxygen Species.**

The next step was to assess the capacity of *C. gigas* hemocytes in each Percoll fraction to produce Reactive Oxygen Species (ROS) upon stimulation by zymosan, utilizing the Luminol oxidation assay. Luminol luminescence peaked 25 minutes after exposure of hemocytes (isolated from total hemolymph) to zymosan, indicating a robust oxidative burst (**Fig. 5C**). The production of ROS was dependent on a NADPH oxidase, as it was completely inhibited by the NADPH oxidase-specific inhibitor apocynin (**Fig. 5C**). To ascertain which hemocyte types were involved in ROS production we tested Percoll density-sorted hemocytes for oxidative burst activity. Fractions 2 and 3 displayed higher oxidative burst activity than the other fractions. The burst intensity of fraction 3 was twice that of total hemocytes, and fraction 2 also exhibited significantly increased burst activity. In contrast, fraction 4 showed a significant decrease in oxidative burst, and no activity was observed for fractions 1, 6 and 7 (**Fig. 5D**). These results indicate that NADPH-dependent oxidative burst activity is carried out by fractions enriched in macrophage-like (ML) cells and blast-like cells (ABL and BBL). Previous studies have shown that granular cells can produce ROS (29). In contrast, we found that small granule cells collected in fraction 7 could not produce ROS through an oxidative burst. To identify which type of hemocyte from blast-like cells and macrophage-like cells produces ROS within a few minutes after exposure to zymosan, ROS production was investigated using NitroBlueTetrazolium (NBT) reduction to stain hemocytes directly.

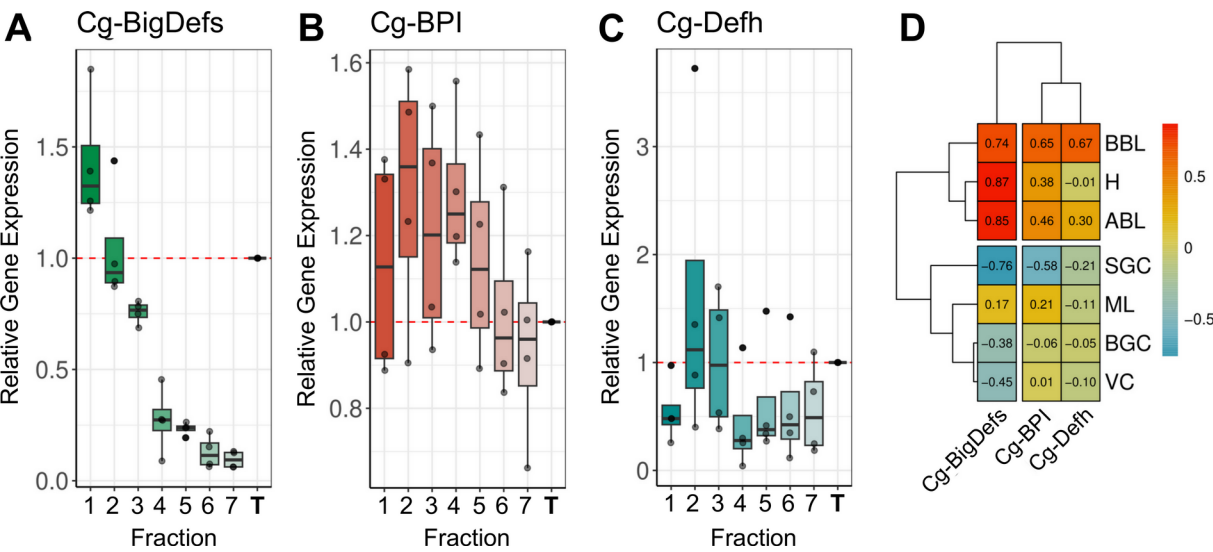
Correlative microscopic analysis using MCDH staining can then be conducted after the NBT reduction reaction. Macrophage-like cells were strongly and significantly stained by NBT reduction (33.2 %) (**Fig. 5E, panels a & d**). In contrast, some small granule cells were lightly stained by NBT reduction (**Fig. 5F**). Blast-like cells, big granule cells, vesicular cells, and hyalinocytes were never NBT stained, confirming that these cell types were not involved in ROS production (**Supp. Fig. S7**). Taken together, these observations demonstrate that small granule cells and macrophage-like cells are the two professional phagocytes among hemocytes. However, only macrophage-like cells are capable of oxidative burst upon exposure to zymosan. In light of these new functional data, we further analyzed the expression level of NADPH-oxidase-related enzymes in the scRNA-seq dataset. The cells in cluster 1 predominantly expressed two NADPH oxidase isoforms (gene numbers G34908 and G23852) compared to other clusters (**Fig. 5H**). Furthermore, this cluster expresses macrophage-specific genes such as the angiopoietin receptor (**Table 1**), as well as maturation factors for dual oxidase, an enzyme involved in peroxide formation (**Supp. Fig. S8**). Collectively, these data indicated that cells of transcriptomic cluster 1 corresponded to macrophage-like cells.

#### **Small granule cells and big granule cells accumulate intracellular copper.**

The effects of copper on oyster hemocytes have been studied extensively due to its abundance in polluted marine environments (19). In addition, several studies have shown that *Vibrio* species pathogenic for *C. gigas* possess copper resistance genes, which are crucial for their survival within hemocytes (9). This prompted us to investigate which hemocyte types are involved in copper metabolism. To this end, total hemocytes were isolated from naive oysters and stained with rhodanine to reveal copper storage in cells. Rhodanine staining revealed that 33 % +/- 2 % of small granule cells and 30 % +/- 10 % of big granule cells exhibited a specific reddish/brown staining indicative of a high concentration of copper in their granules (**Fig. 5G and Supp. Fig. S9**). These results provide functional proof that SGC are specialized in metal homeostasis in addition to phagocytosis, as suggested by the scRNAseq data (Cluster 3).

#### **Antimicrobial peptides are expressed by agranular cells, blasts and hyalinocytes.**

Antimicrobial peptides (AMPs) have long been studied for their role in the invertebrate humoral immune response. They are expressed by hemocytes, including those in *C. gigas* (31). The aim was to ascertain whether different hemocyte cell types expressed distinct AMPs. However, due to the limited sequencing depth, the scRNA-seq analysis was not sensitive enough to reveal AMP expression. This limitation was addressed by investigating AMP expression through RT-qPCR on Percoll density-fractionated hemocytes. The results indicated that Cg-Bigdefs 1-2, Cg-BPI and hemocyte defensin were predominantly expressed by agranular cells, blasts and hyalinocytes (ABL, BBL and H) (Fig. 6A, B and C). The expression of these AMPs was associated with Blasts abundance, while the expression of Cg-BigDefs 1-2 was only associated with hyalinocytes (Fig. 6D). Granular cells (VC, ML, BGC and SGC) did not seem to express any of the analyzed AMPs. These data suggest that some of the agranular cells are specialized in the production of humoral effectors.

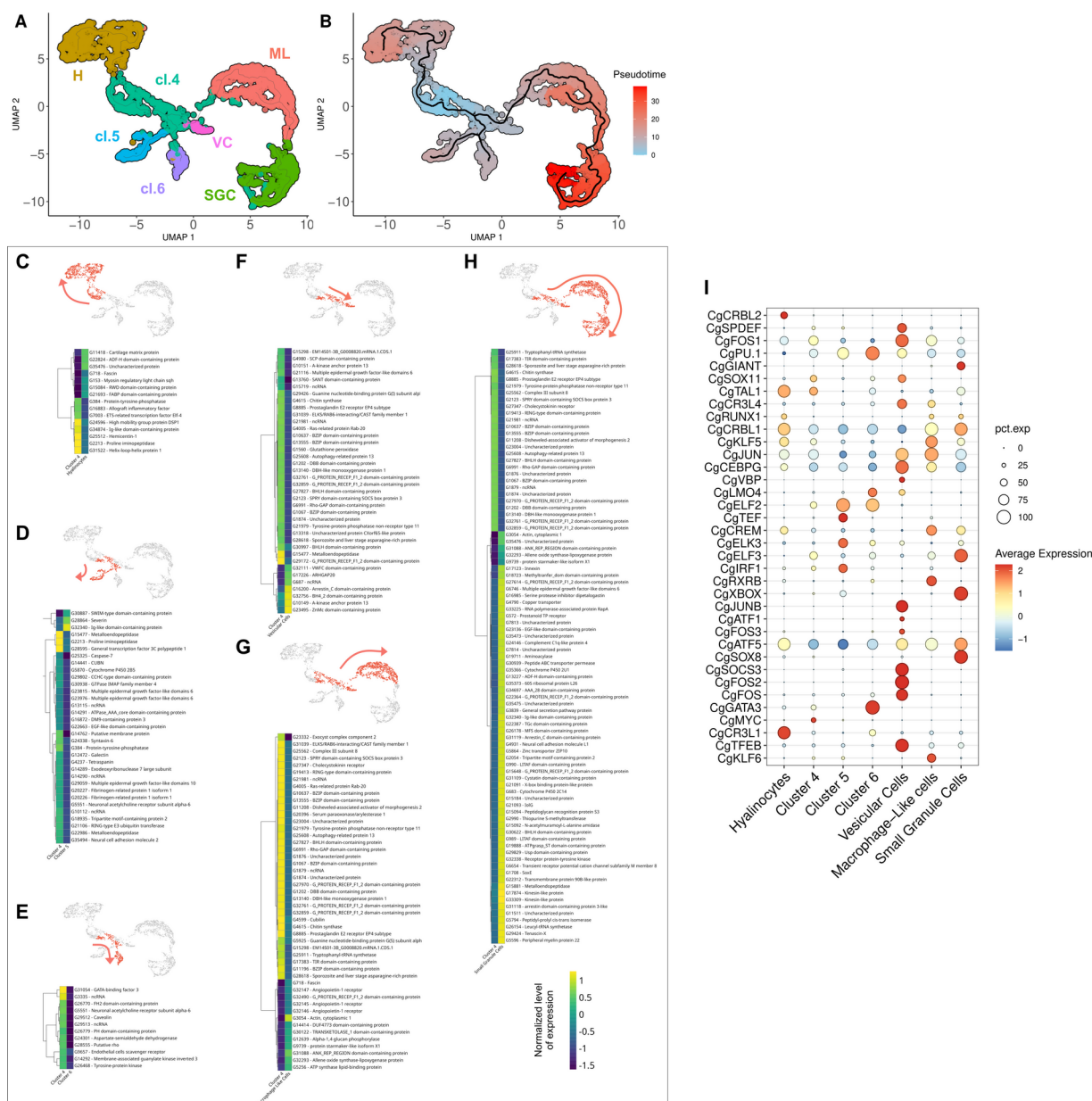


**Fig 6. Hemocyte expression profiles of some antimicrobial peptides. (A) (B) and (C)** Relative Gene Expression in the 7 Percoll hemocyte fractions of Big-Defensin1 & 2 (Cg-BigDefs), BPI (Cg-BPI) and hemocyte defensin (Cg-Defh), respectively, in comparison to the gene expression level in unfractionated hemolymph. **(D)** Correlation matrix between the relative gene expression of BigDefensin1 & 2, BPI and hemocyte defensin gene in each fraction and the percentage of each hemocyte type in each fraction (**H** : Hyalinocytes, **ABL** :

476 Acidophilic Blast Like, **BBL** : Basophilic Blast Like, **SGC** : Small Granule Cell, **ML** :  
 477 Macrophage Like, **BGC** : Big Granule Cell, **VC** : Vesicular Cell. Values and color scale  
 478 represent the Pearson correlation coefficient (r) ranging from -1 (inverse correlation) to +1  
 479 (full correlation).

480 **Tentative model of hemocyte lineages and differentiation pathways in *C. gigas*.**

481 The ontogeny, lineage and differentiation pathways of bivalves remain largely unknown (32).  
 482 However, there are some indications of circulating and proliferating hemocyte progenitors in  
 483 the hemolymph of *C. gigas* (33). GO-terms analysis of the 7 transcriptomic clusters revealed  
 484 different functional signatures, including the transcriptomic signature of cluster 4, which  
 485 showed a high expression level of ribosomal proteins (**Supp. Fig. S1J**) . This particularity has  
 486 been observed in hematopoietic stem cells in vertebrates (34–36). Furthermore, scRNA-seq  
 487 approaches can now be used to deduce differentiation pathways from mRNA splicing variant  
 488 analysis using bioinformatic tools like Monocle3 (37). This revealed an overexpression of  
 489 genes involved in the splicing, transcription and translation continuum in the same fourth  
 490 cluster (**Fig. 2**), thereby reinforcing the hypothesis of a pool of quiescent or immature cells  
 491 that can differentiate upon stimulation.



**Figure 7. Pseudotime ordering of cells revealed 6 potential differentiation pathways of hemocytes.** (A) UMAP plot of scRNA-seq analysis showing the 7 transcriptomic clusters used for pseudotime analysis. 4 clusters were identified cytologically (SGC for small granule cells - cluster 3, H for hyalinocytes - cluster 2, ML for Macrophage Like - cluster 1 and VC for vesicular cells - cluster 7), cl.4, cl.5, and cl.6 represent clusters 4, 5, and 6, respectively. (B) Graphical representation (UMAP projection) of the Monocle 3 pseudo-time order of the clustered cells. Cluster 4 (cl.4) was used as the origin for the pseudotime analysis. (C) (D) (E) (F) (G) and (H) show the gene expression level of selected marker genes obtained from the monocle3 trajectory analysis at the beginning and end of the modeled differentiation

pathways (in red on the UMAP plot) from cluster 4 to hyalinocytes, to cluster 5 cells, to cluster 6 cells, to Vesicular Cells (VC), to Macrophage-Like cells (ML) and to Small Granule Cells (SGC) respectively. The color scale represents the normalized expression level of each gene. **(I)** Dot plot showing the average expression and the percentage of cells expressing identified transcripts encoding for transcription factors in the scRNA-seq dataset. Average expression is expressed in Log2FC.

Cluster 4 was chosen to enroot the pseudotime analysis to deduce differentiation pathways and cell lineages using Monocle3. **(Fig. 7A)**. By temporally ordering the 2817 cells analyzed by scRNA-seq, 5 cell lineages could be defined **(Fig. 7B)**. Differentiation pathway 1 leads to hyalinocytes (H) **(Fig. 7C)**. This transition is characterized by the downregulation of 8 genes, two of which are transcription factors (G7003 and G31522). The hyalinocyte cluster was also characterized by the overexpression of genes involved in cell contractility (G22824, G153 and G11418). Differentiation pathway 2 leads to cells of cluster 5 and is characterized by the downregulation of about 30 genes, and 3 genes related to Zn finger protein, actin cleavage and Ig-like domain proteins were specifically overexpressed (G30887, G28864 and G32340) **(Fig. 7D)**. Differentiation pathway 3 leads to cells of cluster 6 and is characterized by the down-expression of 12 genes, and the upregulation of a GATA family transcription factor (G31054) **(Fig. 7E)**. The pathways between clusters 4, 5 and 6 found by Monocle3 analysis were pseudo temporally short and few specific markers were identified, suggesting that they were transcriptionally close. Differentiation pathway 4 leads to vesicular cells (VC) **(Fig. 7F)** where a large number of genes were silenced to give rise to these cells. Seven genes were specifically overexpressed in VC (G32111, G17226, G687, G16200, G32756, G10149 and G23495). Interestingly, 6 genes encoding potential transcription factors were downregulated in this lineage (G30997, G10637, G1067, G13555, G2123 and G27827). Differentiation pathway 5 leads to macrophage-like cells (ML), and is characterized by the underexpression of 35 genes and the overexpression of 13 genes. Among the 35 genes, 6 are putative transcription factors (G2123, G10637, G13555, G27827, G1067 and G11196) **(Fig. 7G)**. Finally, we can outline a

lineage ending in small granule cells (SGC) (**Fig. 7H**). This pathway involved cluster 4 (immature cells), VC, ML and SGC and was characterized by the downregulation of 27 genes, including 5 potential transcription factors (G10637, G13555, G27827, G1067 and G2123). SGC showed a distinct transcriptomic signature with 61 overexpressed genes, including 3 transcription factors (G1708, G21091 and G30622). Based on these findings, we postulate that immature cluster 4 cells possess pluripotent potential and can give rise to four terminally differentiated cell types : cluster 5 and 6 cells, hyalinocytes and small granule cells, and two other transient hemocyte cell types : vesicular cells and macrophage-like cells. All data from the Monocle3 analyses are available as Supplementary Data (**Supp. Fig. S10**).

Differentiation pathway analysis thus revealed the over- or under-expression of various transcription factors in the identified pathways. Given the established role of transcription factors as master regulators of cell differentiation and their utility in delineating cell lineages, we investigated the combinatorial expression patterns of transcription factors among the different transcriptomic cell clusters. Based on GO-terms annotation, 28 different sequences corresponding to transcription factors were isolated from the scRNA-seq dataset. The transcription factor function was confirmed by manual annotation (**Supp. Table S5**) and **Figure 7I** shows the average expression profiles of these factors in the different clusters. **Supplementary Figure S11** illustrates their expression levels in single cells. Two transcription factors, CgATF5 and CgCRBL1, exhibited a contrasting expression profile with an increased average expression in macrophage-like cells, hyalinocytes and small granule cells versus a low expression profile in cells in clusters 4, 5, 6 and vesicular cells. We also identified transcription factors that were specifically expressed in the different transcriptomic clusters : CgSPDEF, CgSOCS3, CgFOS and CgTFEB were specific for vesicular cells, CgSOX8, CgXBOX and CgELF3 were specific for small granule cells, CgCR3L1 and CgTAL1 for hyalinocytes, and CgJUN, CgKLF5, CgKLF6 and CgCREM for macrophage-like cells. Eight additional transcription factors were specifically identified in cluster 6 (CgGATA3, CgPU.1 and CgELF2), cluster 5 (CgELF2, CgELK3 and CgIRF1) and cluster 4

(CgTAL1 and CgFOS1). These data potentially define four distinct hematopoietic lineages originating from one type of immature blast cells and give rise to hyalinocytes, SGC (via VC and ML), or two distinct differentiated blast-like cells. We also identified a combination of transcription factors specific to lineages and cell types that are potential master regulators of cell fate during hematopoiesis.

## Discussion

The findings of our study represent a significant advancement in our understanding of the functional diversity and lineages of *C. gigas* hemocytes. Single-cell RNA-seq and cytology were combined to identify seven distinct hemocyte transcriptomic cell clusters and an equivalent number of morphotypes. These include four granular cell types (big granule cells, macrophage-like cells, small granule cells and vesicular cells), two distinct blast-like cells (basophilic and acidophilic blast-like cells), and one agranular epithelial-like cell type (hyalinocytes). A significant challenge was to identify correlations between transcriptomic and cytological data to fully define each cytological cell type. This challenge was overcome by combining multiple approaches, including isopycnic Percoll density gradient fractionation combined with the analysis of transcriptomic markers expression, and functional assays including phagocytosis, oxidative burst, copper accumulation, AMP expression and finally pseudotime analysis of gene expression. These results confirmed the historical classification of the 3 main cell groups : blasts, hyalinocytes, and granular cells (15) and deepened our understanding of the functional specificities of poorly characterized cell types. In particular, we identified distinct transcriptional and functional subtypes among blasts and granular cells with complementary immune specialization and lineage relationships between cells.

One significant outcome of the present study is the identification of cell types involved in antimicrobial activities, including phagocytosis, intracellular copper accumulation, an oxidative burst and antimicrobial peptide production. These cell types have been extensively studied for their role in antibacterial and antiparasitic defenses, as they are found in the large

majority of invertebrates (17). Nevertheless, there has been considerable debate surrounding the cell types specialized for these critical immune functions in bivalves, particularly oysters. Moreover, the involvement of the different hemocyte subpopulations in immune functions is not yet fully understood.

Our findings reveal that the macrophage-like (ML) cells and small granule cells (SGC) are the sole hemocyte cell types that function as professional phagocytes, as demonstrated against Zymosan or *Vibrio*. These two distinct cell types could be distinguished functionally. First, two distinct transcriptomic clusters were identified for each cell type (Cluster 1 for ML / BGC and Cluster 3 for SGC). Secondly, only ML induced a measurable oxidative burst. Thirdly, only SGC accumulated intracellular copper in specific granules. The two types of professional phagocytes belong to the same granular cell lineage, as determined by pseudotime analysis. The notion that ML could serve as a precursor for SGC may seem counterintuitive. However, this is not an isolated phenomenon among invertebrate hemocytes. For instance, in *Drosophila* larvae (38), some populations of professional phagocytes, the sessile plasmatocytes, give rise to crystal cells or lamellocytes that are morphologically and functionally distinct from plasmatocytes (39). Similarly, the existence of multiple professional phagocytes in the oyster is reminiscent of vertebrate macrophages, polynuclear neutrophils and dendritic cells, which possess distinct functional specializations, including efferocytosis (40), oxidative burst, NETosis, or antigen presentation (41).

The characterization of professional phagocytes in oysters is of particular importance for a deeper understanding of oyster-*Vibrio* interactions during pathogenesis. Some of the most extensively studied oyster pathogens, including strains of *V. tasmaniensis* and *V. aestuarianus francensis*, harbor virulence traits that enable them to disrupt the phagocytic activity of hemocytes during pathogenesis. For instance, *V. tasmaniensis* behaves as a facultative intracellular pathogen with phagocytosis-dependent cytotoxicity (11). Additionally, both *V.*

611 *tasmaniensis* and *V. aestuarianus* display resistance to copper toxicity through CopA and  
612 CusA/B transporters. This trait is essential for the survival and virulence of these pathogens in  
613 oysters (9, 42). Since SGCs have been demonstrated to be professional phagocytes with  
614 copper-rich granules, the cellular interactions between SGCs and these vibrios are likely to be  
615 critical during the antibacterial host response and pathogenesis. ML are phagocytes that  
616 possess a very potent NADPH-oxidase-dependent oxidative burst. The oxidative burst is a  
617 rapid and potent antimicrobial response observed in professional phagocytes, such as  
618 polynuclear neutrophils in mammals. It is worth noting that NADPH-dependent ETosis has  
619 been observed in *C. gigas* in a manner analogous to that observed in human neutrophils (43).  
620 This cell death is characterized by the projection of DNA extracellular traps that capture and  
621 kill some pathogens like *Vibrio* (43). Therefore, it is reasonable to hypothesize that ML may  
622 also be involved in ETosis.

623

624 The specialized functions of the two other types of granular cells, the BGCs and the VCs,  
625 remain unclear. Despite the difficulty in identifying a specific scRNAseq transcriptomic  
626 cluster for BGCs, the level of expression of laccase 24 was found to be higher in a particular  
627 subcluster among ML (**Supp. Fig. S8**) and pseudo-time analysis highlighted the same  
628 subcluster as an alternative differentiation state among ML (**Fig. 7B**). The enrichment of  
629 transcripts involved in oxido-reduction pathways, particularly laccase 24, aligns with their  
630 potential role in melanization and response to oxidative stress (44). While melanin-like  
631 deposits have been observed in some cases of infestation by the parasites *Martelia* or *Bonamia*  
632 in the Sydney rock oyster (45), this mechanism is not as robust as that described in arthropods,  
633 which perform melanization through a prophenoloxidase activation cascade (46, 47). In many  
634 marine invertebrates (44), a type of hemocyte known as Brown Cells could be related to the  
635 BGCs described here. When observed without any staining (as in **Fig. 5F**), their big granules  
636 with a yellow to dark brown content appear to align with the historical description of brown  
637 cells that often infiltrate tissues (like gills) in animals exposed to polluted waters (44). It has

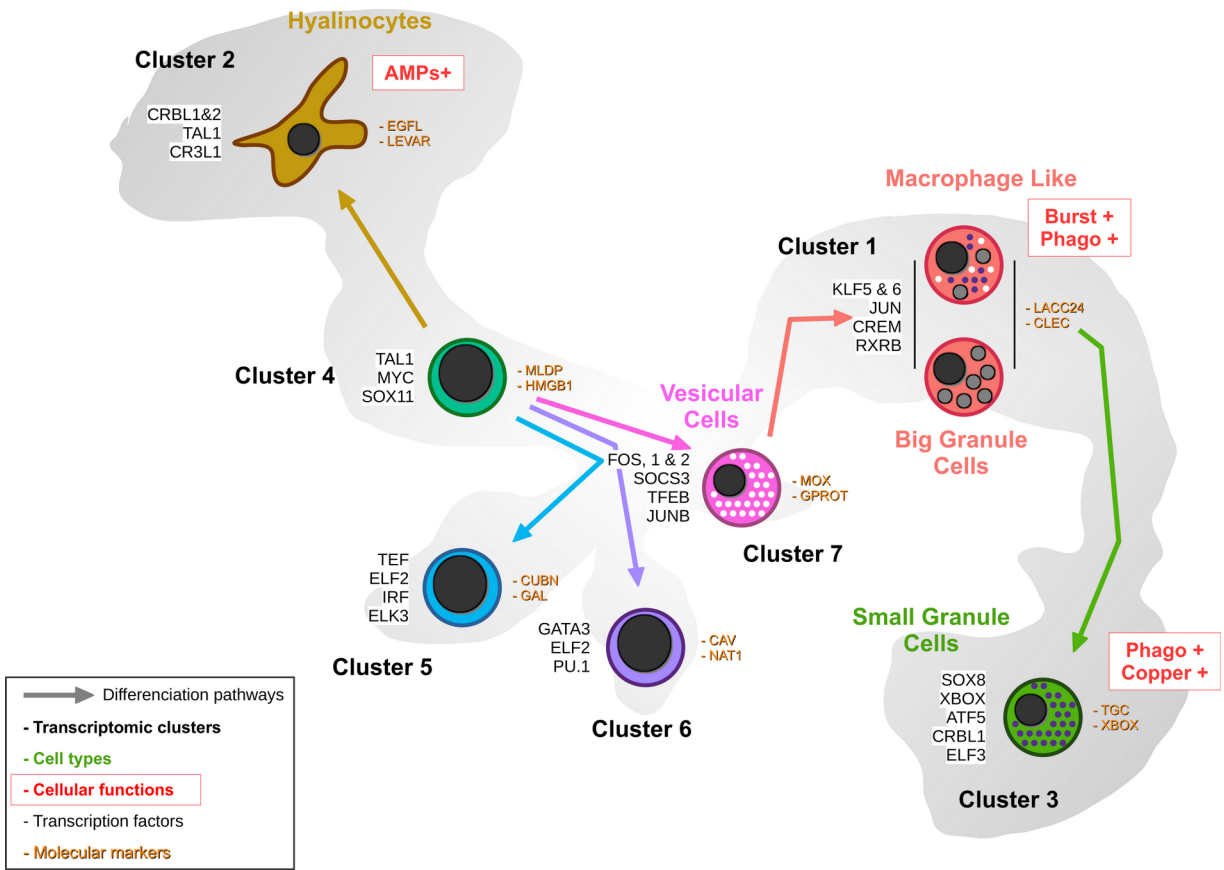
therefore been theorized that these cells are involved in detoxification processes. Our pseudotime analysis indicates that they likely originate from ML, with limited phagocytosis activity and a specialized role in melanization and potentially heavy metal detoxification, as evidenced by rhodanine staining showing copper accumulation in some of their granules (**Fig. 5G**). Further studies are recommended to clarify the role of BGCs, particularly in the context of parasitic infestation or exposure to toxic stresses. Lastly, VC are granular cells that remain to be functionally characterized. Their transcriptomic profile suggests strong intracellular vesicular trafficking and autophagy activity. However, our functional assays did not reveal any particular immune-related function. Their clear granules appear auto-fluorescent when illuminated with UV light (**Supp. Fig. S11**) but the biochemical nature of the content of these granules remains to be characterized. As they also appear as cell intermediates along the granular cell differentiation pathway in pseudotime analysis, they could represent functionally immature precursors of the other three granular cell types, much like promyelocytes which possess specific azurophilic granules but are functionally immature precursors during granulocyte differentiation in humans. However, the enrichment of autophagy-related transcripts in VC calls for further investigation into a potential role in antiviral immunity, as autophagy has been suggested to play a role in the response to OsHV-1 virus (48).

Hyalinocytes are a homogenous cell type, with only one morphotype matching one transcriptomic cluster. It can be deduced from pseudotime analysis that they originate from a specific and very different differentiation pathway than the granular cells. According to the literature, hyalinocytes are involved in the early stages of inflammation and can infiltrate wounds and interact with foreign particles. In the flat oyster *Ostrea edulis*, they contribute to shell production and wound healing (49). In the Sydney rock oyster they play a role in cell aggregation (50), while in *C. virginica* (51) they contribute to encapsulation, reminiscent of lamellocytes in *Drosophila*. Our results suggest an important role of AMP expression in the immune response. Indeed, *Cg-BigDefs*, which participate in the control of oyster microbiota

665 (52), were found to be expressed in both hyalinocytes and blast-like cells. Moreover,  
666 hyalinocytes from the oyster *O. edulis* have been shown to express the AMP Myticin C (53),  
667 which lends further support to this immune function. Among the AMPs, we also found that  
668 Cg-BPI and Cg-Defh, are more expressed in BBL, ABL than in other cell types. These results  
669 are somewhat unexpected, given the prevailing assumption that AMPs are stored in granules  
670 of granular cells, rather than agranular cells (16). These results highlight the necessity to  
671 reassess the role of specific agranular cell types in the active production of humoral immune  
672 effectors. Our findings suggest that hyalinocytes and/or the blast-like cells may be a cellular  
673 target of the OsHV-1 virus, the causal agent of POMS, which dampens the expression of  
674 certain AMPs (7), thereby inducing bacterial dysbiosis. It is still unclear whether this is due to  
675 a decreased expression of AMPs and/or inhibition of immature blast cell differentiation  
676 involved in the renewal of agranular cell types.

677

678 It should be noted that the complexity of blast-like cells could not be fully elucidated in this  
679 study, as 3 clusters and only 2 morphotypes were identified. Cell fractionation using Percoll  
680 gradient failed to yield pure blast-enriched fractions (ABL and BBL), preventing precise  
681 functional characterization. The enrichment in transcripts of the transcription synthesis  
682 degradation continuum aligns with the definition of undifferentiated blast-type cells and with  
683 the basophilic staining obtained in MCDH for BBL cells, as immature blast cells are  
684 characterized by a basophilic cytoplasmic staining in humans. However, our results show that  
685 certain blast populations can produce AMPs, suggesting that these cells may also play a role in  
686 the production of humoral effectors. Ultimately, it remains unclear whether these circulating  
687 immature cells are indeed the stem cells from which all hemocytes originate.



689 **Figure 8. Proposed hemocyte ontology in *Crassostrea gigas* based on the transcriptomic,**  
690 **cytological and functional results obtained.** Cells are colored according to the same color  
691 code as the transcriptomic clusters. Cluster numbers and cell types are indicated. To the left of  
692 the cells are the overexpressed transcription factors and to the right are the identified marker  
693 genes in each cluster. Functional characteristics of hyalinocytes, macrophage-like cells and  
694 small granule cells are marked in red. (**AMP** : AntiMicrobial Peptide, **Burst** : ROS  
695 production, **Phago** : phagocytosis)

696 In the animal kingdom, the innate immune system relies on specialized cells derived from  
697 pluripotent precursors through hematopoiesis. Transcription factors in particular have been  
698 found to exhibit a high degree of conservation throughout the animal kingdom, from  
699 invertebrates to vertebrates. However, the mechanisms underlying the functional  
700 differentiation of bivalves are only partially available or understood (54). Recent research has  
701 indicated the potential existence of hemocyte progenitors, also known as blast-like cells in  
702 several bivalve species. These include clams, mussels, scallops, marine mussels, freshwater  
703 mussels, oysters, pearl oysters, and wing shells (54). However, various models for

hematopoiesis in bivalves have been proposed and extensively debated without a clear consensus or definitive proof. Single-cell RNA sequencing and pseudotime analysis have enabled us to propose a refined model of hematopoiesis at an unprecedented level of detail. In this model, the different types of hemocytes are likely produced through four differentiation pathways that originate from a common progenitor. One pathway results in the formation of agranular hyalinocytes, while another independent pathway gives rise to the granular cells, including VC, ML, BGC, and SGC. Furthermore, two additional differentiation pathways have been identified that may lead to terminally differentiated blast-like cells. The differentiation of mature hemocytes involves the establishment of lineage-specific gene expression profiles, which rely on transcription factors to modulate the expression of their target genes. Our study identified a combination of transcription factors that are differentially expressed during *C. gigas* hematopoiesis and are specific to differentiation stages and cell lineages, including many well-known hematopoietic transcription factors such as GATA, PU-1, TAL1 and SOX factors, which are positioned in this model of *C. gigas* hematopoiesis (55–57) (**Fig. 8**).

This study significantly advances our understanding of bivalve immunity, especially in comparison to arthropods. By introducing a standardized reference hemocytogram for oysters using MCDH staining, defining cell type-specific markers and key transcription factors likely involved in cell fate determination, as well as clarifying the functions of different hemocytes, we have paved the way for future in-depth studies. This will facilitate further studies of the oyster's immune response to various biotic and abiotic stress at the cellular level. Improved comprehension of antiviral and antibacterial responses in bivalves, along with an enhanced understanding of immune priming and immune memory at the cellular level in bivalves, will benefit health and population management practices for sustainable aquaculture production. These findings will also contribute to the broader field of evolutionary immunology by enabling comparative studies and elucidating the diversification of immune cells and immunity-related genes in a protostome.

731

732

## 733 **Materials and Methods**

734

### 735 **Conservation of oysters**

736 The work described here was performed using two different sources of oysters of the same  
 737 species *Crassostrea (Magallana) gigas*. ISA (Ifremer Standardized Oysters - La tremblade -  
 738 France) oysters for the scRNA-seq experiment and oysters provided by a local supplier  
 739 (<https://www.huitres-bouzigues.com>). Animals were washed and kept in 10 L tanks containing  
 740 seawater and a bubbler to oxygenate and homogenize the water. Water was changed daily and  
 741 oyster health was monitored. All animals used in this study were 18 months of age.

### 742 **Hemocyte collection and processing**

743 *Crassostrea gigas* hemocytes were collected from live animals by puncture of the adductor  
 744 muscle. The oyster shell was incised on the posterior side with forceps. Hemolymph was  
 745 collected using a 23Gx1" needle mounted on a 5 mL syringe prefilled with 2 mL of ice-cold  
 746 Alsever modified medium (20.8 g glucose – 8 g trisodium citrate - 22.5 g sodium chloride -  
 747 0.4 g BSA - pH=7.5). Samples were centrifuged for 4 minutes at 200 g – 4 °C and the  
 748 supernatant was removed and replaced with 1 mL of fresh Alsever modified medium. Each  
 749 hemocyte sample was thoroughly checked for quality, counted under a microscope using  
 750 KOVA (Kova International, USA) slides, and stored on ice prior to processing. For scRNA-  
 751 seq analysis, resuspended hemocytes were filtered on 30 µm filters, counted, the solution was  
 752 adjusted to 1.10<sup>6</sup> cells per mL and stored on ice prior to 10X genomic library preparation.

### 753 ***C. gigas* genome annotation**

754 The *C. gigas* genome (Genbank GCA\_905397895.1) (24) was used as a reference. Prior to  
 755 annotation, the longest CDS sequences were extracted from the gff3 file. Annotation was  
 756 realized using the ORSON script (<https://gitlab.ifremer.fr/bioinfo/workflows/orson>). ORSON  
 757 combines cutting-edge tools for annotation processes within a Nextflow pipeline. The ORSON  
 758 script performs a sequence similarity search with BLAST (58) against the Uniprot-Swissprot  
 759 and Uniprot-trEMBL databases, and functional prediction with InterProScan (59) and eggNOG

(60) orthogroup annotation. Interproscan analysis was performed against Pfam, Prosite, CDD, TIGR, SMART, SuperFamily, PRINTS and Hamap databases. Results were collected and processed using Blast2GO (61) for annotation mapping and validation.

### **Drop Seq-based scRNA-seq library generation**

The 10X Genomics protocol, Single Cell 3' Reagent Kits v2 User Guide from the manufacturer (10X Genomics, USA) was followed to prepare gel in emulsion beads (GEM) containing single cells, hydrogel beads, and reverse transcription reagents, perform barcoded cDNA synthesis, and generate sequencing libraries from pooled cDNAs. The concentration of single-cell suspensions was approximately 1000 cells /  $\mu$ L, as estimated by manual counting, and cells were loaded according to the 10X protocol to capture approximately 3000 cells per reaction. Library construction (after GEM digestion) was performed using 10X reagents according to the manufacturer's instructions. Libraries (paired-end reads 75 bp) were sequenced on an Illumina NovaSeq (Illumina, USA) using two sequencing lanes per sample.

### **scRNA-seq analysis**

Reads were aligned to the *C. gigas* reference genome (Genbank GCA\_905397895.1) (27) using STAR solo software (v 2.7.10) (23). Unique molecular identifiers (UMIs) were extracted and counted for each cell, and an expression matrix was generated for further analysis. Single-cell RNA sequencing (scRNA-seq) data analysis was performed using the R programming language (version 4.2.1) (R Core Team, 2018) and the Seurat package (version 4.3.0) (62). The data were then pre-processed to remove unwanted sources of variation and to normalize gene expression. Cells with small library sizes and a high proportion of mitochondrial genes were excluded. Data normalization was performed using the SCTransform method. After normalization, highly variable genes were identified using the FindVariableFeatures function and the top variable genes were selected for downstream analyses. Dimensionality reduction was performed using Principal Component Analysis (PCA), followed by Uniform Manifold Approximation and Projection (UMAP) to visualize the data. Cell clustering was performed using the 'FindClusters' function, using the previously identified significant principal

787 components (dims = 6) and a resolution parameter ( $r = 0.1$ ) to define cluster granularity.  
788 Differential expression analysis was performed using the 'FindAllMarkers' function (pct.min  
789 = 0.25) to identify genes differentially expressed between clusters, with statistical significance  
790 determined using the Wilcoxon rank sum test. Functional enrichment analysis of differentially  
791 expressed genes was performed using gene set enrichment analysis (RBGOA) (27).

## 792 **KEGG pathway analysis**

793 KEGG analysis was performed using DAVID Bioinformatics Resources (NIAID/NIH) (26).  
794 Gene lists of specifically overexpressed genes in each cluster were obtained after scRNA-seq  
795 processing (genes with  $\text{Log}_2\text{FC} > 0.25$  and significant p-value  $< 0.001$ ) and used for KEGG  
796 annotation. The *C. gigas* reference genome from the DAVID bioinformatics resource was  
797 used for this analysis, with thresholds of 2 for counts and 0.05 for EASE value (p-value).  
798 KEGG annotation results were post-processed and presented as a heatmap showing the KEGG  
799 pathway, fold enrichment, p-value significance and number of positive terms.

## 800 **Rank-Based Gene Ontology Analysis**

801 RBGOA first clusters GOs according to their representative genes to identify the most  
802 significant GOs, and then ranks the identified biological processes according to their average  
803 expression levels (over all representative genes). Finally, biological processes, molecular  
804 functions and cellular components significantly enriched in DEGs are identified by a Mann-  
805 Whitney rank test with a strict FDR correction. The files generated by the GO-MWU scripts  
806 ([https://github.com/z0on/GO\\_MWU](https://github.com/z0on/GO_MWU)) were post-processed to extract the category names and  
807 the fraction indicating the number of "good candidates" relative to the total number of genes  
808 belonging to that category. The "good candidates" were the genes that exceeded an arbitrary  
809 'absValue' cutoff (defined as 0.001) in their significance measure. The results were presented  
810 as a heatmap.

## 811 **Percoll Density Gradient Separation of Hemocytes**

812 A concentration series of Percoll (Cytiva, Sweden) diluted in Alsever modified medium was  
813 prepared as follows: 10, 20, 30, 40, 50, 60 and 70 % (vol/vol). Discontinuous density gradients

(from 10 % Percoll with a density of 1.0647 to 70 % Percoll with  $d=1.1049$ ) were made using 1.5 mL of each concentration and loaded with 1 mL of the hemocyte suspension corresponding to approximately  $2.10^7$  cells. Centrifugation was performed (30 min, 800 g, 4 °C) in a swinging bucket on a Beckman Coulter JS-13.1 rotor (Beckman Coulter, USA). Hemocytes concentrated at each density interface were collected separately with a 70 mm long 20Gx2.75" needle mounted on a 1 mL syringe. The hemocytes were then washed from the Percoll by adding 10 mL of ice-cold Alsever modified medium, pelleted by centrifugation (10 min, 200 g, 4°C) and resuspended in Alsever modified medium or filtered seawater.

### **Cytological description of the hemocyte populations**

200,000 fresh hemocytes were seeded onto a slide using a Cytospin 4 centrifuge (Thermo Scientific, USA). The samples were then stained using the panoptic MCDH (Micro Chromatic Detection for Hematology) (Cellavision, Sweden) staining protocol. This protocol produces purple hues typical of Romanowsky-Giemsa staining results. After staining, the samples were observed using a LEICA DMR (Leica AG, Germany) transmitted light microscope with a 40x magnification objective. Each slide was imaged and the hemocytes were counted and characterized based on their morphology.

### **Real Time-quantitative Polymerase Chain Reaction (RT-qPCR)**

Total RNA was extracted using the RNeasy kit (Qiagen, the Netherlands) and cDNA was synthesized from 1 µg total RNA using the Superscript IV kit (ThermoFisher Scientific, USA) with oligo(dT) primers. RT-PCR was performed on LightCycler© 480 thermocycler using the SYBR Green 1 Master kit (Roche, Switzerland). Primers were used at 200 nM. Primer sequences are listed in **Supplementary Table S3**. Expression was normalized to *Cg-rps6* reference gene. The standard cycling conditions were 95°C pre-incubation for 3 minutes followed by 40 cycles of 95°C for 10 seconds, 60°C for 20 seconds, and 72°C for 25 seconds. Each pair of primers was first tested and validated on a total hemocyte RNA preparation to control melting curves and establish calibration lines.

### **Oxidative Burst Assay**

841 The production of reactive oxygen species was quantified by luminescence assay. Briefly,  
842 hemocytes from hemolymph puncture or Percoll density gradient were washed once with  
843 filtered sterile water. 50  $\mu$ L of hemocytes were plated in triplicate on a 96-well plate ( $3.10^5$   
844 cells/cm<sup>2</sup>). 50  $\mu$ L of 40 mM luminol (Merck and Co, USA) was added to each well. After 45  
845 minutes of incubation at room temperature, the oxidative burst was induced by adding 100  $\mu$ L  
846 of zymosan (Merck and Co, USA) at a multiplicity of infection (MOI) of 50:1. The plate was  
847 immediately placed in a Berthold Centro XS3 LB 960 luminescence microplate reader  
848 (Berthold GmbH, Germany) to measure luminescence emission every 2 minutes for 2 hours.

#### 849 **Phagocytosis assay**

850 400  $\mu$ L of filtered, sterile water-washed hemocytes were seeded in a 24-well plate at a  
851 concentration of  $3.10^5$  cells/cm<sup>2</sup>. After 15 minutes, 50  $\mu$ L of zymosan was added to the  
852 fractions at an MOI of 20:1. For LMG20012<sup>T</sup> Vibrio, 50  $\mu$ L of bacteria was added to a total  
853 hemolymph at an MOI of 5:1. After 1 hour of contact at room temperature, the cells were  
854 resuspended and 200  $\mu$ L of the suspension was applied to a microscope slide using a Cytospin  
855 centrifuge. The slides were then stained with MCDH and observed under a LEICA DMR  
856 transmitted light microscope with a 40x magnification objective (Leica AG, Germany) to  
857 count phagocytic cells.

#### 858 **NitroBlueTetrazolium (NBT) staining**

859 ROS production was measured using NitroBlue Tetrazolium reduction after zymosan  
860 stimulation. Briefly, 1 mL ( $1.10^6$  cells) of hemocyte solution was mixed with 50  $\mu$ L of filtered  
861 NBT solution (15 mg/mL in water). Zymosan was added at a 4:1 MOI and the mixture was  
862 incubated at room temperature for 10 minutes on a rocking shaker. Then, 50  $\mu$ L of each  
863 sample was plated onto glass coverslips and observed under a transmitted light microscope  
864 LEICA DMR with a 40x magnification objective (Leica AG, Germany) to count NBT-positive  
865 cells. The positions of positive NBT cells were recorded prior to MCDH staining to identify  
866 hemocyte types.

#### 867 **Rhodanine copper staining of hemocytes**

868 The storage of copper by hemocytes was examined by Rhodanine staining. Briefly,  $1.10^5$   
 869 hemocytes were plated on Superfrost slides using a cytopsin and circled with a hydrophobic  
 870 pen to retain the staining solution. The Copper Stain Kit (SkyTek, USA) was used to stain the  
 871 hemocytes. As described by the kit manufacturer, one drop of Rhodanine solution was added  
 872 to 9 drops of acetate to form the working solution. Five drops were placed on the cytopsin  
 873 cells and a 5 mL Eppendorf tube with the cap removed was placed over the cells to prevent  
 874 evaporation of the working solution. The slide and the balanced Eppendorf tube were placed  
 875 in a beaker of boiling distilled water for 20 minutes. The slide was then washed with 5 drops  
 876 of acetate and 3 drops of hematoxylin were placed on the slide for 1 minute at room  
 877 temperature. The slides were then washed a final time with acetate and observed under a  
 878 LEICA DMR transmitted light microscope with a 40x magnification objective (Leica AG,  
 879 Germany) to count rhodanine-positive cells. The positions of rhodanine-positive cells were  
 880 recorded prior to MCDH staining to identify hemocyte types.

### 881 **Pseudotemporal ordering of cells with Monocle3**

882 Cells from the *C. gigas* dataset were analyzed using Monocle3 ([https://github.com/cole-](https://github.com/cole-trapnell-lab/monocle3)  
 883 [trapnell-lab/monocle3](https://github.com/cole-trapnell-lab/monocle3)) (37). The Monocle3 analysis was performed on the Seurat object  
 884 following the aforementioned processing steps. Clustering information (features, genes,  
 885 partitions, clusters and UMAP coordinates) was transferred to a CDS object. The cell  
 886 trajectory was calculated using the *learn\_graph* function. The *choose\_graph\_segments*  
 887 function was used to select three lineages. The gene expression along pseudotime data was  
 888 extracted from the result. Then, the data were used to plot genes along pseudotime in three  
 889 lineages using ggplot2 v3.4.4 R package and the heatmap was generated using the pheatmap  
 890 v1.0.12 R package.

### 891 **Statistical Analysis**

892 To evaluate differences between samples, a statistical analysis was performed using (version  
 893 4.2.1) (R Core Team, 2018) and appropriate packages. All data were examined for normality,  
 894 and statistical tests were selected accordingly. One-way analysis of variance (ANOVA) was

used for normally distributed data. Seven different hemolymph samples were used for cytological analysis. Oysters were provided by our local supplier) and approximately 300 hemocytes were counted per sample. Seven independent experiments were performed for Percoll density gradient separation. A Tukey test was used to evaluate the statistical difference between the proportions of hemocyte types. The phagocytic capacity of hemocytes was tested in three independent experiments and statistical differences were evaluated using the Tukey test for both the phagocytic capacity between hemocyte types and the number of particles per phagocyte. Finally, oxidative burst capacity was tested 3 times on Percoll-separated hemocytes. The Tukey test was also used to assess statistical differences between conditions. The null hypothesis was rejected at a significance level of  $p = 0.05$ .

905  
906  
907  
908

## References

1. Y. M. Bar-On, R. Phillips, R. Milo, The biomass distribution on Earth. *Proceedings of the National Academy of Sciences* **115**, 6506–6511 (2018).
2. B. Leicester Bayne, *Biology of Oysters* (Academic Press, 1st Edition., 2017)vol. 41.
3. G. Zhang, L. Li, J. Meng, H. Qi, T. Qu, F. Xu, L. Zhang, Molecular Basis for Adaptation of Oysters to Stressful Marine Intertidal Environments. *Annu. Rev. Anim. Biosci.* **4**, 357–381 (2016).
4. FAO, Global aquaculture production. Fisheries and aquaculture division (2023).
5. F. Pernet, C. Lupo, C. Bacher, R. J. Whittington, Infectious diseases in oyster aquaculture require a new integrated approach. *Philosophical Transactions of the Royal Society B: Biological Sciences* **371**, 20150213 (2016).
6. A. Alfjorden, M. Areskog, D. Bruno, R. Carnegie, D. Cheslett, S. Feist, S. Ford, S. Jones, A. Lillehaug, L. Madsen, T. Renault, N. Ruane, P. Vennerstrom, New Trends in Important Diseases Affecting the Culture of Fish and Molluscs in the ICES Area 2002– 2015. *ICES cooperative research report*, doi: 10.17895/ices.pub.2800 (2017).
7. J. de Lorgeril, A. Lucasson, B. Petton, E. Toulza, C. Montagnani, C. Clerissi, J. Vidal-Dupiol, C. Chaparro, R. Galinier, J.-M. Escoubas, P. Haffner, L. Dégremont, G. M. Charrière, M. Lafont, A. Delort, A. Vergnes, M. Chiarello, N. Faury, T. Rubio, M. A. Leroy, A. Pérignon, D. Régler, B. Morga, M. Alunno-Bruscia, P. Boudry, F. Le Roux, D. Destoumieux-Garzón, Y. Gueguen, G. Mitta, Immune-suppression by OsHV-1 viral infection causes fatal bacteraemia in Pacific oysters. *Nat Commun* **9**, 4215 (2018).
8. N. M. Coyle, C. O'Toole, J. C. L. Thomas, D. Ryder, E. J. Feil, M. Geary, T. P. Bean, A. W. Joseph, A. Waine, D. Cheslett, D. W. Verner-Jeffreys, *Vibrio aestuarianus* clade A and clade B isolates are associated with Pacific oyster (*Magallana gigas*) disease outbreaks across Ireland. *Microbial Genomics* **9**, 001078 (2023).
9. A. Mesnil, M. Jacquot, C. Garcia, D. Tourbiez, L. Canier, A. Bidois, L. Dégremont, D. Cheslett, M. Geary, A. Vetri, A. Roque, D. Furones, A. Garden, P. Orozova, I. Arzul, M. Sicard, G. M. Charrière, D. Destoumieux-Garzón, M.-A. Travers, Emergence and clonal expansion of *Vibrio aestuarianus* lineages pathogenic for oysters in Europe. *Molecular Ecology* **32**, 2869–2883 (2023).

10. D. Oyanedel, A. Lagorce, M. Bruto, P. Haffner, A. Morot, Y. Labreuche, Y. Dorant, S. de La Forest Divonne, F. Delavat, N. Inguibert, C. Montagnani, B. Morga, E. Toulza, C. Chaparro, J.-M. Escoubas, Y. Gueguen, J. Vidal-Dupiol, J. de Lorgeril, B. Petton, L. Degremont, D. Tourbiez, L.-L. Pimparé, M. Leroy, O. Romatif, J. Pouzadoux, G. Mitta, F. Le Roux, G. M. Charrière, M.-A. Travers, D. Destoumieux-Garzón, Cooperation and cheating orchestrate *Vibrio* assemblages and polymicrobial synergy in oysters infected with OsHV-1 virus. *Proceedings of the National Academy of Sciences* **120**, e2305195120 (2023).
11. T. Rubio, D. Oyanedel, Y. Labreuche, E. Toulza, X. Luo, M. Bruto, C. Chaparro, M. Torres, J. de Lorgeril, P. Haffner, J. Vidal-Dupiol, A. Lagorce, B. Petton, G. Mitta, A. Jacq, F. Le Roux, G. M. Charrière, D. Destoumieux-Garzón, Species-specific mechanisms of cytotoxicity toward immune cells determine the successful outcome of *Vibrio* infections. *Proceedings of the National Academy of Sciences* **116**, 14238–14247 (2019).
12. Y. Labreuche, F. Le Roux, J. Henry, C. Zatylny, A. Huvet, C. Lambert, P. Soudant, D. Mazel, J.-L. Nicolas, *Vibrio aestuarianus* zinc metalloprotease causes lethality in the Pacific oyster *Crassostrea gigas* and impairs the host cellular immune defenses. *Fish Shellfish Immunol* **29**, 753–758 (2010).
13. M. Lafont, A. Vergnes, J. Vidal-Dupiol, J. de Lorgeril, Y. Gueguen, P. Haffner, B. Petton, C. Chaparro, C. Barrachina, D. Destoumieux-Garzón, G. Mitta, B. Gourbal, C. Montagnani, A Sustained Immune Response Supports Long-Term Antiviral Immune Priming in the Pacific Oyster, *Crassostrea gigas*. *mbio* **11**, 10.1128/mbio.02777-19 (2020).
14. M. Fallet, C. Montagnani, B. Petton, L. Dantan, J. de Lorgeril, S. Comarmond, C. Chaparro, E. Toulza, S. Boitard, J.-M. Escoubas, A. Vergnes, J. Le Grand, I. Bulla, Y. Gueguen, J. Vidal-Dupiol, C. Grunau, G. Mitta, C. Cosseau, Early life microbial exposures shape the *Crassostrea gigas* immune system for lifelong and intergenerational disease protection. *Microbiome* **10**, 85 (2022).
15. N. R. de la Ballina, F. Maresca, A. Cao, A. Villalba, Bivalve Haemocyte Subpopulations: A Review. *Front. Immunol.* **13**, 826255 (2022).
16. E. Bachère, R. D. Rosa, P. Schmitt, A. C. Poirier, N. Merou, G. M. Charrière, D. Destoumieux-Garzón, The new insights into the oyster antimicrobial defense: Cellular, molecular and genetic view. *Fish & Shellfish Immunology* **46**, 50–64 (2015).
17. J.-M. Escoubas, B. Gourbal, D. Duval, T. J. Green, G. M. Charrière, D. Destoumieux-Garzón, C. Montagnani, “Immunity in Molluscs” in *Encyclopedia of Immunobiology*, M. J. H. Ratcliffe, Ed. (Academic Press, Oxford, 2016; <https://www.sciencedirect.com/science/article/pii/B9780123742797120041>), pp. 417–436.
18. A. S. Mount, A. P. Wheeler, R. P. Paradkar, D. Snider, Hemocyte-Mediated Shell Mineralization in the Eastern Oyster. *Science* **304**, 297–300 (2004).
19. W. S. S. Fischer, “Environmental influence on bivalve hemocyte function” in *Am Fish Soc Symp* (1988)vol. 18, pp. 225–237.
20. T. V. Nguyen, A. C. Alfaro, F. Merien, Omics approaches to investigate host–pathogen interactions in mass mortality outbreaks of *Crassostrea gigas*. *Reviews in Aquaculture* **11**, 1308–1324 (2019).
21. J. Meng, G. Zhang, W.-X. Wang, Functional heterogeneity of immune defenses in molluscan oysters *Crassostrea hongkongensis* revealed by high-throughput single-cell transcriptome. *Fish & Shellfish Immunology* **120**, 202–213 (2022).
22. R. D. Rosa, P. Alonso, A. Santini, A. Vergnes, E. Bachère, High polymorphism in big defensin gene expression reveals presence–absence gene variability (PAV) in the oyster *Crassostrea gigas*. *Developmental & Comparative Immunology* **49**, 231–238 (2015).
23. B. Kaminow, D. Yunusov, A. Dobin, STARsolo: accurate, fast and versatile mapping/quantification of single-cell and single-nucleus RNA-seq data. *bioRxiv* [Preprint] (2021). <https://doi.org/10.1101/2021.05.05.442755>.
24. C. Peñaloza, A. P. Gutierrez, L. Eöry, S. Wang, X. Guo, A. L. Archibald, T. P. Bean, R. D. Houston, A chromosome-level genome assembly for the Pacific oyster *Crassostrea gigas*. *GigaScience* **10** (2021).

25. R. Satija, J. A. Farrell, D. Gennert, A. F. Schier, A. Regev, Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* **33**, 495–502 (2015).
26. B. T. Sherman, M. Hao, J. Qiu, X. Jiao, M. W. Baseler, H. C. Lane, T. Imamichi, W. Chang, DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Research* **50**, W216–W221 (2022).
27. R. M. Wright, G. V. Aglyamova, E. Meyer, M. V. Matz, Gene expression associated with white syndromes in a reef building coral, *Acropora hyacinthus*. *BMC Genomics* **16**, 371 (2015).
28. E. Bachère, D. Chagot, H. Grizel, Separation of *Crassostrea gigas* hemocytes by density gradient centrifugation and counterflow centrifugal elutriation. *Developmental & Comparative Immunology* **12**, 549–559 (1988).
29. W. Wang, M. Li, L. Wang, H. Chen, Z. Liu, Z. Jia, L. Qiu, L. Song, The granulocytes are the main immunocompetent hemocytes in *Crassostrea gigas*. *Developmental & Comparative Immunology* **67**, 221–228 (2017).
30. C. Lambert, P. Soudant, G. Choquet, C. Paillard, Measurement of *Crassostrea gigas* hemocyte oxidative metabolism by flow cytometry and the inhibiting capacity of pathogenic vibrios. *Fish & Shellfish Immunology* **15**, 225–240 (2003).
31. P. Schmitt, R. Rosa, M. Duperthuy, J. de Lorgeril, E. Bachère, D. Destoumieux-Garzon, The Antimicrobial Defense of the Pacific Oyster, *Crassostrea gigas*. How Diversity may Compensate for Scarcity in the Regulation of Resident/Pathogenic Microflora. *Frontiers in Microbiology* **3** (2012).
32. G. Vogt, Hidden Treasures in Stem Cells of Indeterminately Growing Bilateral Invertebrates. *Stem Cell Rev and Rep* **8**, 305–317 (2012).
33. M. Jemaà, N. Morin, P. Cavellier, J. Cau, J.-M. Strub, C. Delsert, Adult somatic progenitor cells and hematopoiesis in oyster. *Journal of Experimental Biology*, jeb.106575 (2014).
34. S. P. Fahl, M. Wang, Y. Zhang, A.-C. E. Duc, D. L. Wiest, Regulatory Roles of Rpl22 in Hematopoiesis: An Old Dog with New Tricks. *Crit Rev Immunol* **35**, 379–400 (2015).
35. M. Rehn, A. Wenzel, A.-K. Frank, M. B. Schuster, S. Pundhir, N. Jørgensen, K. Vitting-Seerup, Y. Ge, J. Jendholm, M. Michaut, E. M. Schoof, T. L. Jensen, N. Rapin, R. T. Sapio, K. L. Andersen, A. H. Lund, M. Solimena, M. Holzenberger, D. G. Pestov, B. T. Porse, PTBP1 promotes hematopoietic stem cell maintenance and red blood cell development by ensuring sufficient availability of ribosomal constituents. *Cell Reports* **39**, 110793 (2022).
36. E. I. Athanasiadis, J. G. Botthof, H. Andres, L. Ferreira, P. Lio, A. Cvejic, Single-cell RNA-sequencing uncovers transcriptional states and fate decisions in haematopoiesis. *Nat Commun* **8**, 2045 (2017).
37. J. Cao, M. Spielmann, X. Qiu, X. Huang, D. M. Ibrahim, A. J. Hill, F. Zhang, S. Mundlos, L. Christiansen, F. J. Steemers, C. Trapnell, J. Shendure, The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).
38. D. Hultmark, I. Andó, Hematopoietic plasticity mapped in *Drosophila* and other insects. *eLife* **11**, e78906 (2022).
39. U. Banerjee, J. R. Girard, L. M. Goins, C. M. Spratford, *Drosophila* as a Genetic Model for Hematopoiesis. *Genetics* **211**, 367–417 (2019).
40. E. Boada-Romero, J. Martinez, B. L. Heckmann, D. R. Green, The clearance of dead cells by efferocytosis. *Nat Rev Mol Cell Biol* **21**, 398–414 (2020).
41. D. C. Dale, L. Boxer, W. C. Liles, The phagocytes: neutrophils and monocytes. *Blood* **112**, 935–945 (2008).
42. A. S. Vanhove, T. P. Rubio, A. N. Nguyen, A. Lemire, D. Roche, J. Nicod, A. Vergnes, A. C. Poirier, E. Disconzi, E. Bachère, F. Le Roux, A. Jacq, G. M. Charrière, D. Destoumieux-Garzon, Copper homeostasis

- at the host vibrio interface: lessons from intracellular vibrio transcriptomics. *Environmental Microbiology* **18**, 875–888 (2016).
43. A. C. Poirier, P. Schmitt, R. D. Rosa, A. S. Vanhove, S. Kieffer-Jaquinod, T. P. Rubio, G. M. Charrière, D. Destoumieux-Garzón, Antimicrobial Histones and DNA Traps in Invertebrate Immunity. *Journal of Biological Chemistry* **289**, 24821–24831 (2014).
44. G. Zaroogian, P. Yevich, Cytology and biochemistry of brown cells in *Crassostrea virginica* collected at clean and contaminated stations. *Environmental Pollution* **79**, 191–197 (1993).
45. B. Allam, D. Raftos, Immune responses to infectious diseases in bivalves. *Journal of Invertebrate Pathology* **131**, 121–136 (2015).
46. J. Nakhleh, L. El Moussawi, M. A. Osta, “Chapter Three - The Melanization Response in Insect Immunity” in *Advances in Insect Physiology*, P. Ligoxygakis, Ed. (Academic Press, 2017)vol. 52 of *Insect Immunity*, pp. 83–109.
47. P. Amparyup, W. Charoensapsri, A. Tassanakajon, Prophenoloxidase system and its role in shrimp immune responses against major pathogens. *Fish & Shellfish Immunology* **34**, 990–1001 (2013).
48. S. Picot, N. Faury, C. Pelletier, I. Arzul, B. Chollet, L. Dégremont, T. Renault, B. Morga, Monitoring Autophagy at Cellular and Molecular Level in *Crassostrea gigas* During an Experimental Ostreid Herpesvirus 1 (OsHV-1) Infection. *Front. Cell. Infect. Microbiol.* **12** (2022).
49. N. R. de la Ballina, A. Villalba, A. Cao, Differences in proteomic profile between two haemocyte types, granulocytes and hyalinocytes, of the flat oyster *Ostrea edulis*. *Fish & Shellfish Immunology* **100**, 456–466 (2020).
50. S. Aladaileh, S. V. Nair, D. Birch, D. A. Raftos, Sydney rock oyster (*Saccostrea glomerata*) hemocytes: Morphology and function. *Journal of Invertebrate Pathology* **96**, 48–63 (2007).
51. E. Rifkin, T. C. Cheng, H. R. Hohl, An electron-microscope study of the constituents of encapsulating cysts in the American oyster, *Crassostrea virginica*, formed in response to *Tylocephalum* metacestodes. *Journal of Invertebrate Pathology* **14**, 211–226 (1969).
52. N. De San Nicolas, A. Asokan, R. D. Rosa, S. N. Voisin, M.-A. Travers, G. Rocha, L. Dantan, Y. Dorant, G. Mitta, B. Petton, G. M. Charrière, J.-M. Escoubas, V. Boulo, J. Pouzadoux, H. Meudal, K. Loth, V. Aucagne, A. F. Delmas, P. Bulet, C. Montagnani, D. Destoumieux-Garzón, Functional Diversification of Oyster Big Defensins Generates Antimicrobial Specificity and Synergy against Members of the Microbiota. *Marine Drugs* **20**, 745 (2022).
53. N. R. de la Ballina, A. Villalba, A. Cao, Shotgun analysis to identify differences in protein expression between granulocytes and hyalinocytes of the European flat oyster *Ostrea edulis*. *Fish Shellfish Immunol* **119**, 678–691 (2021).
54. E. A. Pila, J. T. Sullivan, X. Z. Wu, J. Fang, S. P. Rudko, M. A. Gordy, P. C. Hanington, Haematopoiesis in molluscs: A review of haemocyte development and function in gastropods, cephalopods and bivalves. *Developmental & Comparative Immunology* **58**, 119–128 (2016).
55. M. de Bruijn, E. Dzierzak, Runx transcription factors in the development and function of the definitive hematopoietic system. *Blood* **129**, 2061–2069 (2017).
56. E. W. Scott, M. C. Simon, J. Anastasi, H. Singh, Requirement of transcription factor PU.1 in the development of multiple hematopoietic lineages. *Science* **265**, 1573–1577 (1994).
57. I. Bergiers, T. Andrews, Ö. Vargel Bölükbaşı, A. Buness, E. Janosz, N. Lopez-Anguila, K. Ganter, K. Kosim, C. Celen, G. Itr Perçin, P. Collier, B. Baying, V. Benes, M. Hemberg, C. Lancrin, Single-cell transcriptomics reveals a new dynamical function of transcription factors during embryonic hematopoiesis. *eLife* **7**, e29312 (2018).
58. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403–410 (1990).

59. P. Jones, D. Binns, H.-Y. Chang, M. Fraser, W. Li, C. McAnulla, H. McWilliam, J. Maslen, A. Mitchell, G. Nuka, S. Pesseat, A. F. Quinn, A. Sangrador-Vegas, M. Scheremetjew, S.-Y. Yong, R. Lopez, S. Hunter, InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
60. C. P. Cantalapiedra, A. Hernández-Plaza, I. Letunic, P. Bork, J. Huerta-Cepas, eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution* **38**, 5825–5829 (2021).
61. S. Götz, J. M. García-Gómez, J. Terol, T. D. Williams, S. H. Nagaraj, M. J. Nueda, M. Robles, M. Talón, J. Dopazo, A. Conesa, High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research* **36**, 3420–3435 (2008).
62. Y. Hao, S. Hao, E. Andersen-Nissen, W. M. Mauck, S. Zheng, A. Butler, M. J. Lee, A. J. Wilk, C. Darby, M. Zager, P. Hoffman, M. Stoeckius, E. Papalexi, E. P. Mimitou, J. Jain, A. Srivastava, T. Stuart, L. M. Fleming, B. Yeung, A. J. Rogers, J. M. McElrath, C. A. Blish, R. Gottardo, P. Smibert, R. Satija, Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).

## Acknowledgments

We are grateful to the staff of the Ifremer platform of “La Tremblade” for technical support in animal housing. scRNA-seq data generated and used in this work were produced through the MGX platform (University of Montpellier, CNRS ,INSERM). We thank the bioinformatic service of Ifremer (SEBIMER) for their help in bioinformatics and the qPCR platform GenSeq (University of Montpellier). We would like to thank Viviane Boulo and Danielle Mello for their enriching discussions, as well as all the members of the 2MAP laboratory team for the fruitful discussions throughout this project.

## Ethics

This work did not require ethical approval from a human subject or animal welfare committee.

## Funding

This work was funded by the Agence Nationale de la Recherche (MOSAR-DEF, ANR-19-CE18-0025; DECICOMP, ANR-19-CE20-0004 and TRANSCAN ANR-18-CE35-0009), the University of Montpellier, iSite MUSE and the “GT-Huître” initiative from Ifremer. This study falls within the framework of the “Laboratoires d’Excellence (LABEX)” Tulip (ANR-10-LABX-41). Sébastien De La Forest Divonne was awarded a PhD grant from the Region

930 Occitanie (HemoFight project) and the University of Perpignan Via Domitia graduate school  
931 ED305.

932

### 933 **Author contributions**

934 Conceptualization: G.M., D.D.-G., B.G., G.M.C. and E.V.

935 Methodology: S.d.L.F.D., J.P., G.M., D.D.-G., B.G., G.M.C. and E.V.

936 Investigation: S.d.L.F.D., J.P., G.M.C. and E.V.

937 Supervision: G.M.C. and E.V.

938 Writing (original draft): S.d.L.F.D. and E.V.

939 Writing (review and editing): all authors.

940

### 941 **Competing interests**

942 The authors declare that they have no competing interests.

943

### 944 **Declaration of AI use**

945 We have not used AI-assisted technologies in creating this article.

946

### 947 **Data and materials availability**

948 All data needed to evaluate the conclusions in the paper are present in the paper and/or the

949 Supplementary Materials. Raw reads are available at ENA under the project accession number

950 PRJEB74031.

# Supplementary Materials for

## **Diversity and functional specialization of oyster immune cells uncovered by integrative single cell level investigations**

Sébastien de la Forest Divonne *et al.*

\*Corresponding author. Email: [emmanuel.vignal@umontpellier.fr](mailto:emmanuel.vignal@umontpellier.fr)

### **This PDF file includes:**

Data S1 to S3

Tables S1 to S5

Figs. S1 to S12

**Data S1. Single-cell RNA-seq analysis result file.** CSV file containing the scRNA-seq analysis results with cluster number (<cluster>), gene number (<gene>), chromosome reference where the gene is located (<chromosome>), average expression level (<avg\_log2FC>), percentage in the cluster (<pct1>) and in other clusters (<pct2>), pct1 / pct2 ratio (<pct. ratio>), adjusted p-val and p-val values (<p-val> and <p\_val\_adj>), and gene description (<description>).

**Data S2. Annotation file of CDS extracted from the *Crassostrea gigas* genome file.** CSV file containing the annotation results collected and processed using Blast2GO for annotation mapping and validation.

**Data S3. Compilation file for RBGOA results.** CSV file containing concatenated results of all RBGOA tests used to draw gene ontology analysis heatmaps with cluster number (<cluster\_number>), GO term universe (<goterm\_universe>), number of good gene candidates (<number\_of\_good\_candidates>), total number of genes in this GOTerm category (<total\_number\_of\_genes\_of\_this\_category>), GO term name (<goterm\_name>), adjusted p-value (<pval-adj>) and GO term name variation (<variation>).

Number of Reads	127,959,215
Reads With Valid Barcodes	97.03 %
Sequencing Saturation	75.63 %
Q30 Bases in CB+UMI	96.14 %
Q30 Bases in RNA read	93.62 %
Reads Mapped to Genome: Unique+Multiple	89.22 %
Reads Mapped to Genome: Unique	82.01 %
Reads Mapped to Gene: Unique+Multiple Gene	72.26 %
Reads Mapped to Gene: Unique Gene	72.26 %
Estimated Number of Cells	2937
Unique Reads in Cells Mapped to Gene	67,950,796
Fraction of Unique Reads in Cells	73.48 %
Mean Reads per Cell	23,136
Median Reads per Cell	18,145
UMIs in Cells	16,306,374
Mean UMI per Cell	5552
Median UMI per Cell	4412
Mean Gene per Cell	1578
Median Gene per Cell	1434
Total Gene Detected	23,841

**Table S1. STARsolo summary metrics report.** Metrics dashboard obtained after the STARsolo step, describing the quality of the sequencing and the various characteristics of the cells detected after aligning the reads to the *C. gigas* genome from the Roslin Institute.

Cluster	KEGG Prefix	KEGG Number	KEGG Pathway	Fold Enrichment	P-value (<0.05)	Genes
1	crg	04144	Endocytosis	3.75	1.77E-06	G32465, G28245, G29411, G28554, G23342, G27546, G4677, G4248, G23299, G4679, G25017, G27834, G23833, G20650, G1765, G2151, G15227, G11863, G17437, G30122, G34645, G20982, G31268, G25073, G11958, G17273
1	crg	00030	Pentose phosphate pathway	10.19	4.25E-05	G20983, G16716, COX2, COX1, ND1, CYTB, G2226, G3525, G3658, ND5
1	crg	00190	Oxidative phosphorylation	4.45	3.29E-04	G30122, G18442, G34645, G20982, G31268, G16805, G25073, G11958, G17273, G7839
1	crg	01200	Carbon metabolism	3.54	1.77E-03	G34645, G20982, G11958, G23780, G7839, G13028
1	crg	00010	Glycolysis / Gluconeogenesis	4.94	6.58E-03	G18442, G22426, G22427, G23780, G7839
1	crg	00620	Pyruvate metabolism	5.26	1.38E-02	G22034, G32778, G32842
1	crg	03266	Virion - Herpesvirus	8.74	4.39E-02	G5571, G25192, G5077, G3494, G33210, G26985, G26901, G23358, G1703, G1900, G1543, G2172, G12279, G5185, G26391, G10525, G4452, G11817, G25420, G2853, G4635, G3304, G25620, G5649, G3526, G740, G24712, G26976, G25626, G15236, G34618, G34615, G18702
2	crg	03010	Ribosome	11.87	5.45E-27	G32475, G24506, G4572, G3088, G1531, G15286, G1299, G153, G22991, G1471, G10110, G1470
2	crg	04814	Motor proteins	3.64	3.62E-04	G5106, G26659, G18623, G18624, G14458
2	crg	00980	Metabolism of xenobiotics by cytochrome P450	4.61	2.17E-02	G31970, G3073, G5034, G3298, G5653, G27276, G25266, G650, G331, G23645, G1779, G16296, G23701, G31393, G27816, G10181, G15131, G21371, G2133, G34736
3	crg	00190	Oxidative phosphorylation	8.03	1.47E-12	G26194, G5034, G4037, G3500, G20764, G16296, G23701, G14096, G22912, G15171, G27816, G21794, G19054, G15078, G16455, G34649, G9692
3	crg	01200	Carbon metabolism	5.42	4.82E-08	G26194, G14096, G22912, G21794, G19054, G15078, G4037, G3500, G34649, G12763
3	crg	00010	Glycolysis / Gluconeogenesis	7.41	5.01E-06	G26194, G14096, G22912, G21794, G4037, G25300, G34649, G7042, G25304, G9692, G20764
3	crg	01230	Biosynthesis of amino acids	5.77	1.41E-05	G10612, G34830, G4232, G31680, G28015, G28554, G2650, G35288, G23530, G22892, G3216, G2800, G22524, G25627, G21676, G20683, G1765, G11863
3	crg	04144	Endocytosis	3.2	3.22E-05	G16296, G23701, G5034, G15171, G27816, G19054, G16455
3	crg	00020	Citrate cycle (TCA cycle)	8.23	1.48E-04	G13605, G5034, G3012, G3298, G650, G331, G22961, G5758, G14096, G27816, G13000, G2033, G4037, G3500, G34394, G23387, G25763, G20764, G1779, G23701, G31393, G21533, G15171, G19054, G15131, G1375, G21371, G12581, G9692, G3073, G8883, G31970, G30520, G25177, G27276, G5653, G16296, G2133, G16455, G34649, G7042, G9342, G26194, G23482, G6556, G23164, G5106, G25300, G25266, G25304, G23645, G25904, G22912, G21794, G15078, G10181, G22919, G16488, G34736, G12763, G2280
3	crg	01100	Metabolic pathways	1.27	1.48E-02	G23164, G25300, G25763, G25304, G20764
3	crg	00250	Alanine, aspartate and glutamate metabolism	4.26	2.79E-02	G23482, G34394, G25763, G9342
3	crg	00430	Taurine and hypotaurine metabolism	5.05	4.23E-02	G32202, G25192, G4582, G5077, G3494, G33210, G33255, G5115, G5555, G11948, G31071, G23694, G4502, G732, G1703, G1900, G21588, G18212, G1543, G28190, G12279, G13687, G28274, G10525, G7023, G27586, G3387, G11817, G26894, G4635, G3304, G2853, G24036, G5649, G24712, G26976, G1172, G12546, G27592, G11965, G5571, G4085, G5375, G5652, G23471, G32582, G26985, G26901, G23358, G18432, G16773, G21360, G10473, G2172, G35349, G5185, G26391, G4452, G4972, G3885, G25420, G25620, G3526, G740, G2636, G25626, G16242, G15236, G34618, G1350, G34615, G18702
4	crg	03050	Proteasome	9.09	3.16E-14	G11447, G11879, G9178, G28086, G6441, G27078, G6503, G27105, G4954, G21626, G1667, G22517, G11085, G13469, G32646, G1140, G29381, G12565, G30766, G34613, G30069, G23072, G3035, G27237, G4644, G67, G3357, G28826, G21225, G22448, G21463, G20152, G31757, G12256, G16599, G2099, G16114, G11265, G19821, G5284, G4177, G23063, G3388, G35242, G23100, G2896, G5803, G21555, G22119, G18742, G13575, G14401, G10466
4	crg	03040	Spliceosome	4.6	1.62E-13	G31970, G3077, G3298, G26024, G34551, G11926, G27610, G650, G2799, G17080, G1779, G1500, ND3, G31998
4	crg	00190	Oxidative phosphorylation	3.14	3.88E-04	G21555, G1768, G21245, G23072, G3035, G1892, G27237, G22452, G12561, G11265, G19821, G15119
4	crg	03015	mRNA surveillance pathway	3.18	1.10E-03	G21555, G1768, G23072, G4818, G3035, G2485, G31807, G27237, G12561, G11265, G19821, G27008
4	crg	03013	Nucleocytoplasmic transport	2.44	9.12E-03	G11932, G29430, G10911, G22082, G26491, G29962, G2451, G2694, G6732, G4985, G20713, G204, G20332, G2609, G10994
4	crg	04141	Protein processing in endoplasmic reticulum	2.06	1.28E-02	G27592, G32202, G5185, G5375, G33210, G27586, G11948, G26985, G3885, G26894, G3304, G2853, G18212, G18432, G16773, G1543, G16242, G21360, G13687, G2172, G12279
5	crg	03010	Ribosome	7.41	1.68E-12	G3430, COX2, COX1, ND1, CYTB, ND3, ND2, ND5, ND4
5	crg	00190	Oxidative phosphorylation	4.59	6.31E-04	G30487, G25784, G29415, G3710, G22696, G12500
5	crg	04137	Mitophagy - animal	4.65	8.54E-03	G32202, G25192, G4582, G3494, G5115, G5555, G11948, G4502, G732, G1703, G21588, G18212, G1543, G28190, G12279, G13687, G10525, G7023, G27586, G3387, G11817, G26894, G4635, G3304, G2853, G24036, G5649, G24712, G26976, G27592, G5571, G4085, G5652, G26985, G26901, G23358, G18432, G16773, G21360, G10473, G2172, G5185, G26391, G3885, G25420, G25620, G740, G2636, G25626, G16242, G15236, G34618, G34615, G18702
6	crg	03010	Ribosome	16.48	2.26E-55	G10612, G23299, G27538, G22524, G12273, G16951, G2151, G3185, G6211, G13317, G2450, G31680, G4232, G28015, G2650, G35288, G23342, G20567, G17254, G18762, G19412, G21991, G1774, G1993, G30759
7	crg	04144	Endocytosis	4.04	4.91E-09	G18587, G25784, G11088, G2334, G30759, G3710, G27737, G30997, G12500
7	crg	04137	Mitophagy - animal	4.99	3.50E-04	G3185, G22181, G3430, G28015, G24310, G17795, G14056, G2026, G1780, G30759, G32778, G32842, G90
7	crg	04145	Phagosome	2.82	1.88E-03	G20344, G5397, G25608, G25784, G11385, G2024, G11088, G30759, G10907, G21965, G12500
7	crg	04140	Autophagy - animal	2.97	3.47E-03	G24405, G23180, G11208, G35235, G16231, G2024, G14136, G2334, G1780, G2797
7	crg	04310	Wnt signaling pathway	2.9	6.82E-03	G11645, G5397, G22181, G3430, G2026, G27524, G18705, G26324, G16202, G2976, G27526, G23337
7	crg	04142	Lysosome	1.92	4.51E-02	

**Table S2. Table presenting the result of KEGG analysis performed using DAVID Bioinformatics Resources.** GO term enrichment analysis was conducted on specifically overexpressed genes in each cluster obtained after scRNA-seq processing (genes with Log2FC > 0.25 and significant p-value < 0.001) to highlight the most relevant GO terms associated with a given gene list. The visualization of the different pathways can be obtained from the KEGG website using the KEGG prefix and KEGG number (<https://www.genome.jp/kegg/pathway.html>)

Target	Primer name	Sequence
Laccase 24	LACC24-F	CCT-TGA-TTC-TTC-TTG-CCA-TCC-G
	LACC24-R	AAA-GCT-TGC-GAT-CTT-TGG-CAA
C-type lectin domain-containing protein	CLEC-F	ATC-GGC-TTC-TAC-ATG-GAC-TGA-C
	CLEC-R	GTG-TCT-AAA-GCT-GCG-CCG-AT
Putative modulator of levamisole receptor-1	LEVAR-F	GTG-ACA-GAC-TTC-CCT-CAC-CCT
	LEVAR-R	GCA-CTG-AGT-CGA-GTC-GTA-TGT
EGF-like domain-containing protein 8	EGFL-F	GAG-TGT-TTG-ACA-GGA-CGA-AGC
	EGFL-R	CAT-CAT-CGT-TTC-CAA-CTG-AGG-C
X-Box binding protein-like domain	XBOX-F	GGG-TCA-ACA-GTG-CTA-GGC-AAT
	XBOX-R	GTA-AGC-CAC-CAT-CCC-TAC-CAC
Transglutaminase-like domain containing protein	TGC-F	CTA-CAA-GCT-GGA-CAC-CAC-CAA
	TGC-R	GCA-TTG-ACC-AGT-GAC-ACA-GTC
MD-2-related lipid-recognition domain containing protein	MLDP-F	CTT-GGA-CCT-CGT-TAT-CTT-CGC
	MLDP-R	CTC-CCT-CTG-GTC-CAC-AAA-CAA
High mobility group protein B1	HMGB1	GCC-CAC-GCT-GAA-CTA-TAC-AAG
	HMGB1	CAC-CTT-GTA-GTC-CCT-GAG-TGG
Galectin	GAL-F	CCA-CAG-TAT-CAA-CGA-CCC-TCC
	GAL-R	TCA-CTA-CCG-TCA-TAG-GGA-CCG
Cubilin	CUBN-F	TAA-GTT-CAC-TCT-GGC-CCA-AGG
	CUBN-R	GCT-CAT-GAT-CGT-AGT-GGT-GCT
Natterin-1	NAT1-F	CCG-TAC-GAT-GGT-GAG-GAG-AAA
	NAT1-R	CCG-TCC-CAC-ATA-CTT-GTC-GTT
Caveolin	CAV-F	TCC-AAA-TGA-CCA-TGA-CCC-AGA
	CAV-R	ACT-CTA-TTC-TTG-GTC-GCC-TGG
G Protein receptor F1-2 domain-containing protein	GPROT-F	CGA-ACG-CCT-GCT-TCT-GAT-ATG
	GPROT-R	TCC-ACA-TCG-AAT-GCT-CTG-TCT
DBH-like monooxygenase protein 1	MOX-F	CCT-CCG-CAG-CAA-GAA-GAA-GTA
	MOX-R	TTC-TGT-TTC-GTC-CTC-TCC-ACG
Ribosomal protein S6	RPS6-F	CAG-AAG-TGC-CAG-CTG-ACA-GTC
	RPS6-R	AGA-AGC-AAT-CTC-ACA-CGG-AC
Big Defensin 1&2	BigDef1-2-F	TTC-GCC-TGC-TTC-CAT-ACT-GG
	BigDef1-2-R	GTC-ATG-GTC-ACT-CCT-TAT-TC
Hemocyte defensin	HemDef-F	CTA-CCA-GTT-GTT-CAT-ACA-GAG
	HemDef-R	TCT-TGG-TCA-GAT-TCA-GTC-TGG
Bactericidal Permeability Increasing Protein	BPI	GGA-GGC-GGA-AAT-GGA-TTA-CT
	BPI	TGG-TTG-ACA-TCG-TTG-CTG-AC

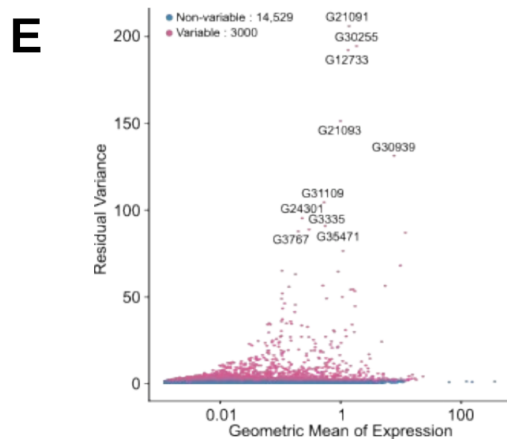
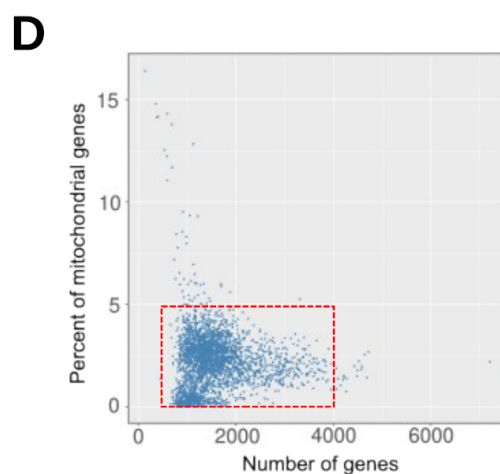
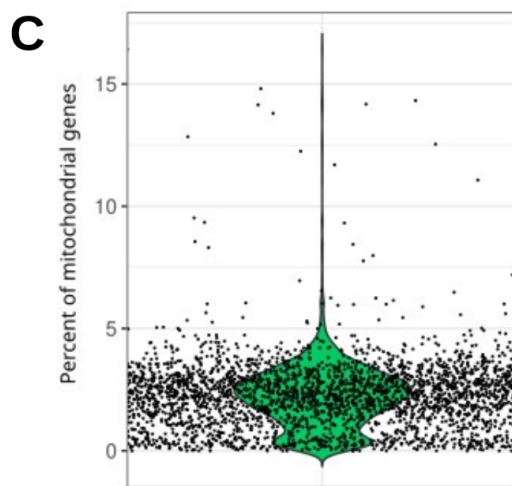
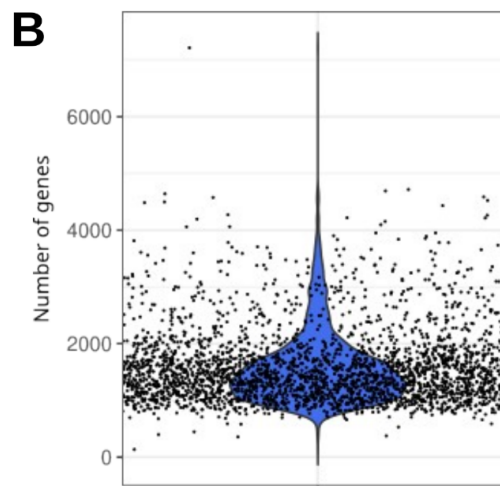
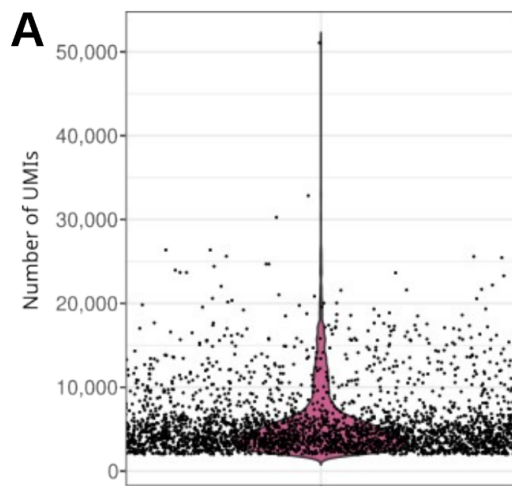
**Table S3. Sequences of primers used in this study.** Name of the transcript targeted by the primer pair, name of the primer used in this study and nucleotide sequence of each primer.

Fraction	Cells	Average	Std.Dev.	Std.Error	Min	Max
1	H	46.43	9.47	4.73	37.50	59.63
1	ML	20.26	9.09	4.55	7.34	28.38
1	BBL	14.47	2.29	1.15	12.16	16.50
1	ABL	15.50	4.94	2.47	8.11	18.35
1	G	0.35	0.44	0.22	0.00	0.92
1	BGC	0.38	0.75	0.38	0.00	1.50
1	VC	2.62	1.97	0.98	0.92	5.41
Fraction	Cells	Average	Std.Dev.	Std.Error	Min	Max
2	H	20.17	5.26	2.63	14.10	26.56
2	ML	27.64	4.48	2.24	22.79	33.33
2	BBL	27.21	9.94	4.97	16.22	39.71
2	ABL	20.17	2.36	1.18	16.91	22.52
2	G	1.14	0.79	0.39	0.00	1.80
2	BGC	1.03	1.26	0.63	0.00	2.56
2	VC	2.64	1.56	0.78	0.74	4.50
Fraction	Cells	Average	Std.Dev.	Std.Error	Min	Max
3	H	12.49	3.78	1.89	10.08	18.04
3	ML	28.26	8.90	4.45	17.12	38.66
3	BBL	21.18	3.11	1.55	16.89	24.32
3	ABL	17.71	5.78	2.89	11.86	25.21
3	G	7.96	5.30	2.65	3.09	14.41
3	BGC	4.71	4.16	2.08	0.00	10.14
3	VC	7.68	2.87	1.43	4.20	10.81
Fraction	Cells	Average	Std.Dev.	Std.Error	Min	Max
4	H	7.82	5.97	2.99	2.20	14.29
4	ML	32.69	13.33	6.66	19.34	50.82
4	BBL	9.64	3.40	1.70	4.76	12.30
4	ABL	5.37	0.97	0.48	4.42	6.67
4	G	18.29	10.39	5.19	5.74	30.94
4	BGC	3.90	1.51	0.76	2.20	5.71
4	VC	22.29	8.25	4.12	11.48	30.77
Fraction	Cells	Average	Std.Dev.	Std.Error	Min	Max
5	H	4.29	4.41	2.21	0.76	10.45
5	ML	19.67	12.04	6.02	3.03	30.88
5	BBL	3.56	1.91	0.95	1.49	6.06
5	ABL	3.35	2.12	1.06	0.75	5.88
5	G	40.92	7.31	3.66	33.82	49.24
5	BGC	4.50	2.98	1.49	1.47	8.33
5	VC	23.71	4.48	2.24	17.91	28.79
Fraction	Cells	Average	Std.Dev.	Std.Error	Min	Max
6	H	3.57	3.82	1.91	0.65	9.09
6	ML	13.95	4.91	2.45	7.10	18.18
6	BBL	1.19	1.38	0.69	0.00	2.58
6	ABL	1.54	0.65	0.33	0.65	2.17
6	G	68.00	10.11	5.06	58.70	78.46
6	BGC	1.69	1.99	1.00	0.00	3.87
6	VC	10.06	5.30	2.65	3.08	15.94
Fraction	Cells	Average	Std.Dev.	Std.Error	Min	Max
7	H	2.43	3.04	1.52	0.00	6.31
7	ML	8.50	6.35	3.18	1.93	17.05
7	BBL	1.67	1.51	0.75	0.00	3.38
7	ABL	0.36	0.72	0.36	0.00	1.45
7	G	81.18	8.99	4.49	68.18	88.41
7	BGC	2.68	1.50	0.75	0.90	4.55
7	VC	3.18	2.92	1.46	0.00	6.82

**Table S4. Hemocyte composition of the 7 Percoll fractions used for qPCR analysis.** For each cell type in each fraction, the table presents the average percentage, standard deviation, standard error, minimum, maximum, and median count values.

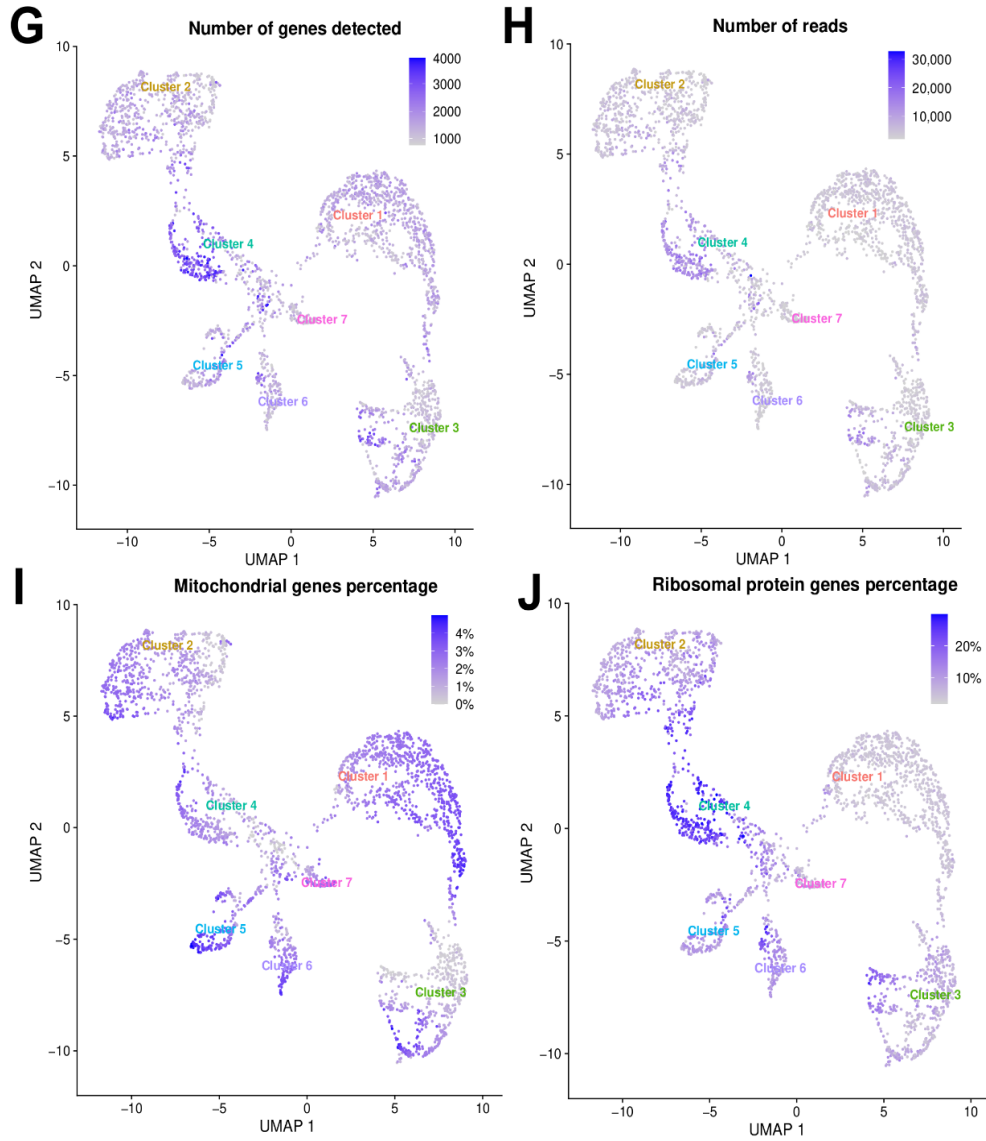
Gene Number	Annotation Name	TF Name	Blast alignment results						Homo sapiens Homologs	
			Best Hit Blast Name	Specie	Accession number	Bit Score	Percentage Identity	Coverage	Name	Uniprot Number
G29966	Krüppel-like factor 15	CgKLF8	Krüppel-like factor 6	C.gigas	XP_011428991	659.448	100	100	HsKLF8	O95600
G30997	BHLH domain-containing protein	CgTFEB	Transcription factor EC isoform X1	C.gigas	XP_011447558	1089.72	100	100	HsTFEB	P19484
G17147	BZIP domain-containing protein	CgCR3L1	cAMP-responsive element-binding protein 3-like protein 1 isoform 1	H.sapiens	NP_443086	1060.06	100	100	HsCR3L2	Q70SY1
G3043	Transcriptional regulator	CgMYC	Transcriptional regulator Myc-A-like	C.angulata	XP_052688275	548.128	100	100	HsMYC	P01108
G31054	GATA-binding factor 3	CgGATA3	Transcription factor GATA-3 isoform X1	C.gigas	XP_011412780	966.066	100	100	HsGATA2	P23769
G11196	BZIP domain-containing protein	CgFOS	Proto-oncogene c-Fos	C.gigas	XP_011446794	664.84	100	100	HsATF3	P18847
G10637	BZIP domain-containing protein	CgFOS2	Fos-related antigen 2	C.gigas	XP_011439871	271.937	100	100	HsATF4	P18848
G2123	SPRY domain-containing SOCS box protein 3	CgSOCS3	SPRY domain-containing SOCS box protein 3	C.gigas	XP_011413771	488.804	100	100	HsSPSB4	Q96A44
G1708	SoxE	CgSOX8	Transcription factor Sox-8-like	C.gigas	NP_001295801	947.192	99.8	100	HsSOX9	P48436
G3506	BZIP domain-containing protein	CgATF5	cAMP-dependent transcription factor ATF-4	C.gigas	XP_011439092	566.614	85.9	100	HsATF5	Q9Y2D1
G11198	BZIP domain-containing protein	CgFOS3	Fos-related antigen 1	C.gigas	XP_034323853	456.062	100	99.42	none	none
G10636	BZIP domain-containing protein	CgATF1	Basic leucine zipper transcriptional factor ATF-like isoform X1	C.gigas	XP_011439867	270.011	100	99.41	none	none
G13555	BZIP domain-containing protein	CgJUNB	Transcription factor jun-B isoform X1	C.gigas	XP_011440881	224.557	100	99.33	none	none
G21091	X-box binding protein like protein	CgXBOX	X-box binding protein-like protein	C.ariakensis	AEF33390	375.491	80.5	99.12	none	none
G12164	Estrogen receptor	CgRXRB	Estrogen receptor beta	C.gigas	XP_011424805	1395.56	100	98.79	HsRXRA	P19793
G748	Interferon regulatory factor	CgIRF1	Interferon regulatory factor 1 isoform X1	C.gigas	XP_011446992	932.742	100	98.51	HsIRF1	P10914
G5204	ETS-related transcription factor Elk-3	CgELF3	ETS homologous factor isoform X1	C.gigas	XP_034310128	883.248	100	96.26	HsELF3	P78545
G28398	ETS domain-containing protein	CgELK3	ETS domain-containing protein Elk-3 isoform X1	C.gigas	XP_011421195	753.051	100	93.91	HsELK3	P41970
G35512	cAMP-responsive element modulator	CgCREM	cAMP-responsive element modulator isoform X2	C.gigas	XP_011448206	572.778	100	92.79	HsCREB1	P16220
G11013	BZIP domain-containing protein	CgTEF	Thyrotroph embryonic factor-like	C.virginica	XP_022332233	570.466	86.4	85.1	HsTEF	Q10587
G7003	ETS-related transcription factor Elk-4	CgELF2	ETS-related transcription factor Elk-2	C.gigas	XP_011447491	523.857	99.6	84.54	HsELF1	P32519
G27920	LIM domain transcription factor LMO4.1	CgLMO4	LIM domain transcription factor LMO4.1 isoform X2	C.gigas	XP_011443581	319.316	98.2	81.86	HsLMO4	P61968
G10865	BZIP domain-containing protein	CgVBP	Transcription factor VBP	C.gigas	XP_011457057	438.343	100	77.95	HsHLF	Q16534
G35439	BZIP domain-containing protein	CgCEBPB	CCAAT/enhancer-binding protein beta	C.gigas	XP_011452433	442.195	100	75.61	HsCEBPB	P53567
G2334	Proto-oncogene c-Jun	CgJUN	Transcription factor AP-1	C.gigas	XP_034313954	429.098	100	58.31	HsJUN	P06627
G26127	Krüppel-like factor 5	CgKLF5	Krüppel-like factor 5 isoform X1	C.gigas	XP_011421316	935.975	100	97.26	HsKLF7	O75840
G20682	cAMP-responsive element-binding protein-like 2	CgCRBL1	cAMP responsive element binding protein-like, partial	C.gigas	AAU93879	318.546	99.4	56.83	HsCRBL2	O60519
G17294	Run1-related transcription factor 1	CgRUNX1	Run1-related transcription factor 1-like isoform X5	C.angulata	XP_052711215	1085.86	100	56.68	HsRUNX3	Q13761
G10077	BZIP domain-containing protein	CgCR3L4	Clumping factor A isoform X1	C.gigas	XP_011425092	775.393	100	56.11	HsCR3L3	Q88CJ9
G2021	BHLH domain-containing protein	CgTAL1	T-cell acute lymphocytic leukemia protein 1-like isoform X1	C.gigas	XP_011439324	394.045	100	49.23	HsTAL2	Q16559
G2933	Putative transcription factor SOX-14	CgSOX11	Transcription factor Sox-11	C.gigas	XP_011445203	424.091	100	43.97	HsSOX12	O15370
G11358	BZIP domain-containing protein	CgGIANT	Protein giant	C.gigas	XP_011417277	486.108	100	38.41	HsTEF	Q10587
G6354	ETS domain-containing protein	CgPU.1	ETS-related transcription factor Elk-3	C.gigas	XP_03420185	189.119	100	36	HsPU.1	P17947
G10633	BZIP domain-containing protein	CgFOS1	Fos-related antigen 1	C.gigas	XP_019918193	371.719	99.5	31.4	HsFOSL1	P15407
G6394	ETS domain-containing protein	CgPDEF	SAM pointed domain-containing Ets transcription factor-like isoform X2	C.gigas	XP_052697855	356.681	99.4	30.12	HsPDEF	O95238
G32716	cAMP-responsive element-binding protein-like 2	CgCRBL2	cAMP-responsive element-binding protein-like 2 isoform X1	C.gigas	XP_034309665	297.745	83.4	26.7	HsCRBL2	O60519

**Table S5. Transcription factors identified in the scRNA-seq dataset of *Crassostrea gigas* hemocytes.** Transcription factors identified in the scRNA-seq dataset and their homology with human proteins, as indicated by the blast alignment results. Each entry includes the gene identifier, the protein it represents in *C.gigas*, and its human counterpart. The bit score and coverage indicate the strength and extent of the alignment, respectively.



**F**

Number of cells before filtering	2937
Average number of genes detected by cell	1578 $\pm$ 643
Average number of UMIs detected by cell	5552 $\pm$ 3848
Filtering parameters	
Cells with gene between 750 and 4000	
Cells with less than 5 % of mitochondrial genes	
Number of cells after filtering	2817
Average number of genes detected by cell	1568 $\pm$ 590
Average number of UMIs detected by cell	5432 $\pm$ 3462

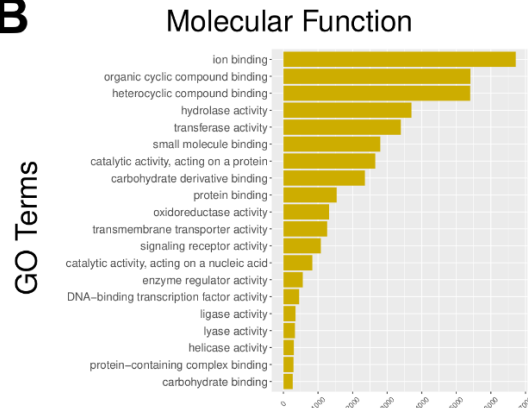
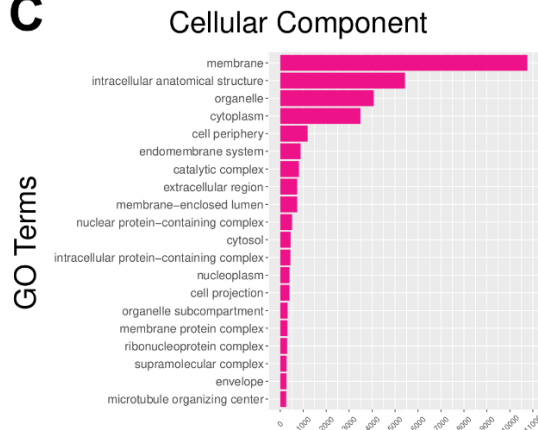
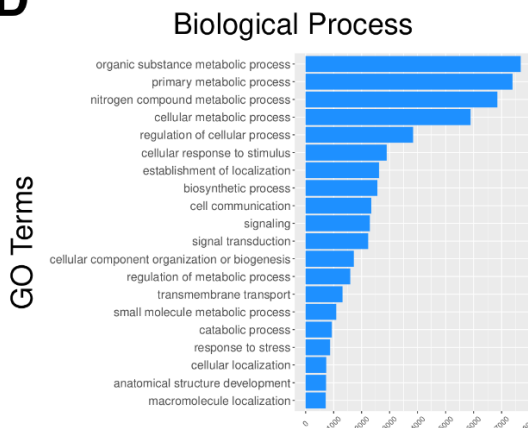


**Figure S1. ScRNA-seq quality control metrics.** Distribution of **A)** unique molecular identifiers (UMIs), **B)** genes and **C)** percentage of mitochondrial genes detected per cell. Each dot represents one cell. **D)** Plot of the percentage of mitochondrial genes versus the number of genes detected in each cell. The red box represents the cells selected for further analysis (number of genes detected between 750 and 4000 and with a percentage of mitochondrial genes less than 5%) **E)** The *FindVariableFeatures()* function was used to identify features with high cell-to-cell variation in the dataset to highlight the biological signal in the single cell dataset. **F)** Table

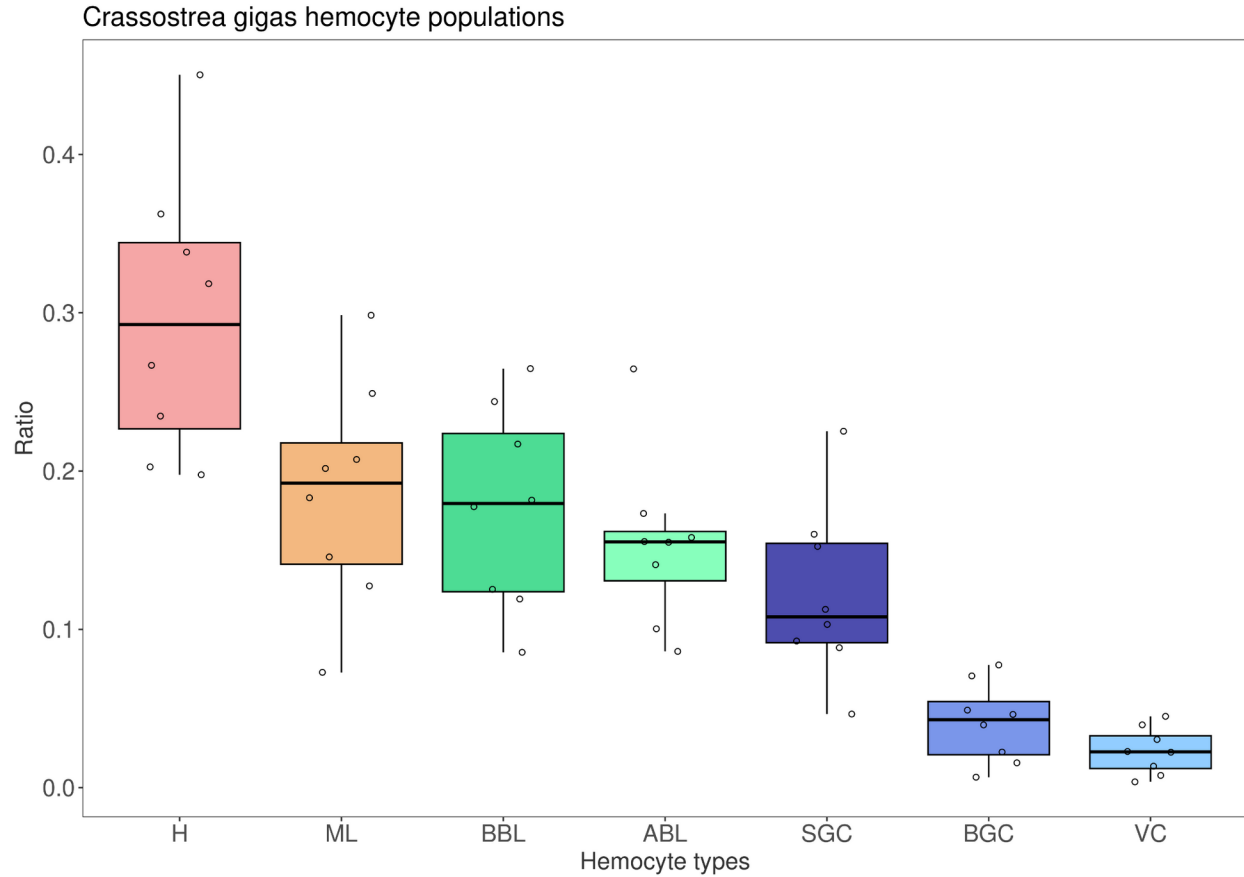
summarizing some quality control metrics. The table shows the thresholds to remove poor quality cells (doublets or empty droplets). The number of cells, UMIs and genes before and after filtering are shown. **G)** Uniform Manifold Approximation and Projection (UMAP) plot of the cells with the number of expressed genes. **H)** UMAP plot of the number of reads per cell. **I)** UMAP plot of the percentage of mitochondrial genes in each cell. **J)** UMAP graph of the percentage of ribosomal protein transcripts in each cell.

**A**

Stage	Number of CDS	% of CDS
Input	30,724	100
<b>Database querying</b>		
Blast annotated	30,511	99.3
Interpro annotated	30,718	99.9
EggNogg annotated	12,630	41.1
<b>Blast2Go</b>		
Mapped	28,570	92.9
Annotated	22,462	73.1

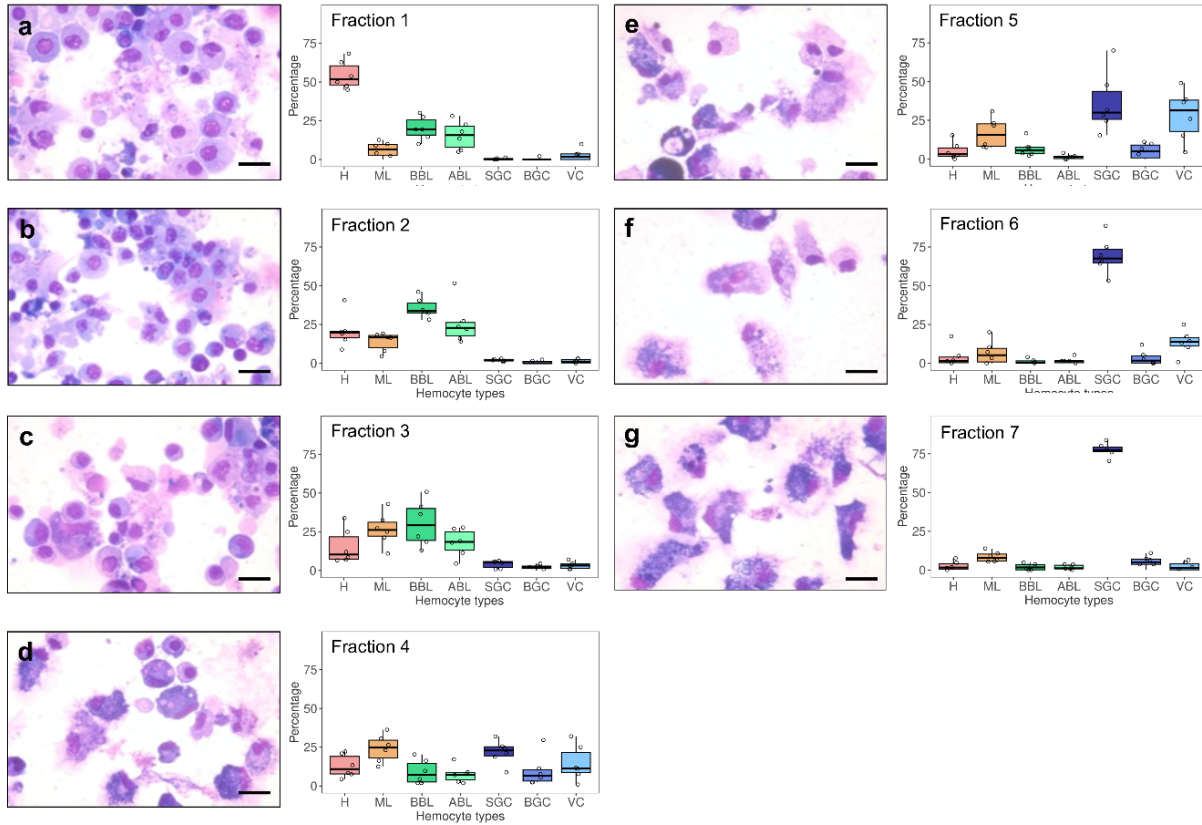
**B****C****D**

**Figure S2. Results of the *C. gigas* genome re-annotation.** **A)** The number and percentage of Coding DNA Sequences (CDS) with valid annotation after each annotation step are shown. A BLAST query was performed against the TrEMBL/Uniprot database and InterproScan annotation against Pfam, PrositeSiteProfiles, CDD, TIGRFAM, PRINTS, SMART, SUPERFAMILY and Hamap databases. Blast and InterProScan results were compiled and processed using Blast2Go. A first mapping step was used to enrich the Blast result with GO-terms, and the annotation step was used to optimize and validate the GO-terms annotations. **B)** **C)** and **D)** show the distribution of the various categories of GO-terms across the three primary domains of Gene Ontology : Molecular Function, Cellular Component and Biological Process, respectively.

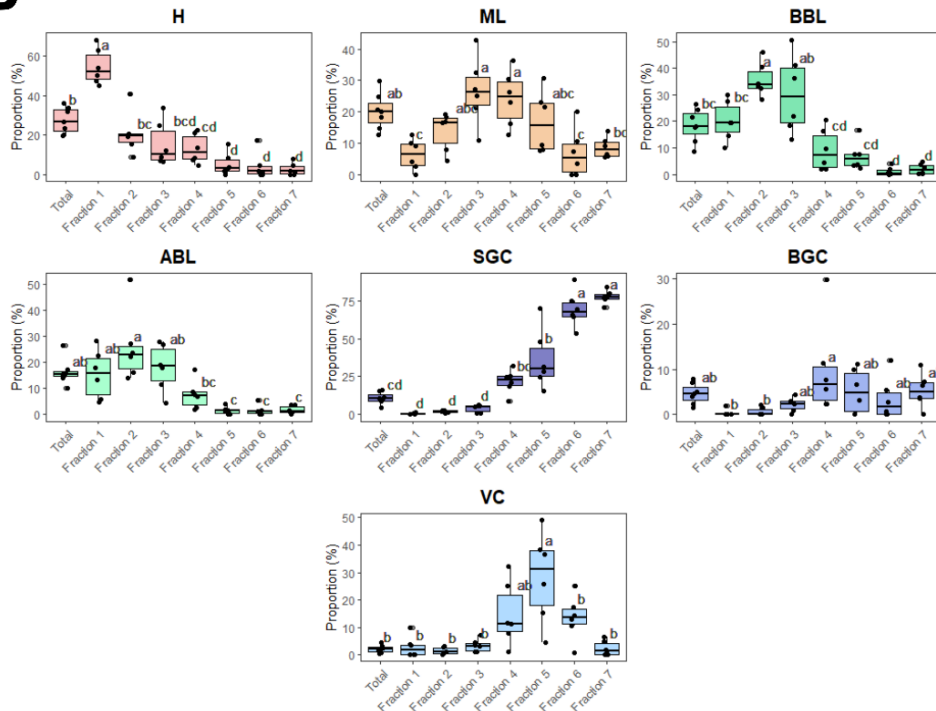


**Figure S3. Whisker plot showing the distribution of counts of different hemocyte populations.** Hemolymph was collected from eight different oysters and hemocytes were plated on a slide using cytopsin centrifugation. The proportion of each hemocyte type in these hemolymphs was calculated after MCDH staining. **H** : Hyalinocytes, **ML** : Macrophage-Like cells, **BBL** : Basophilic Blast-Like cells, **ABL** : Acidophilic Blast-Like cells, **SGC** : Small Granule Cells, **BGC** : Big Granule Cells, and **VC** : Vesicular Cells. The calculation of the proportion of each cell type in these hemolymphs reveals a heterogeneous composition of hemocytes between individuals as well as between cell types.

# A

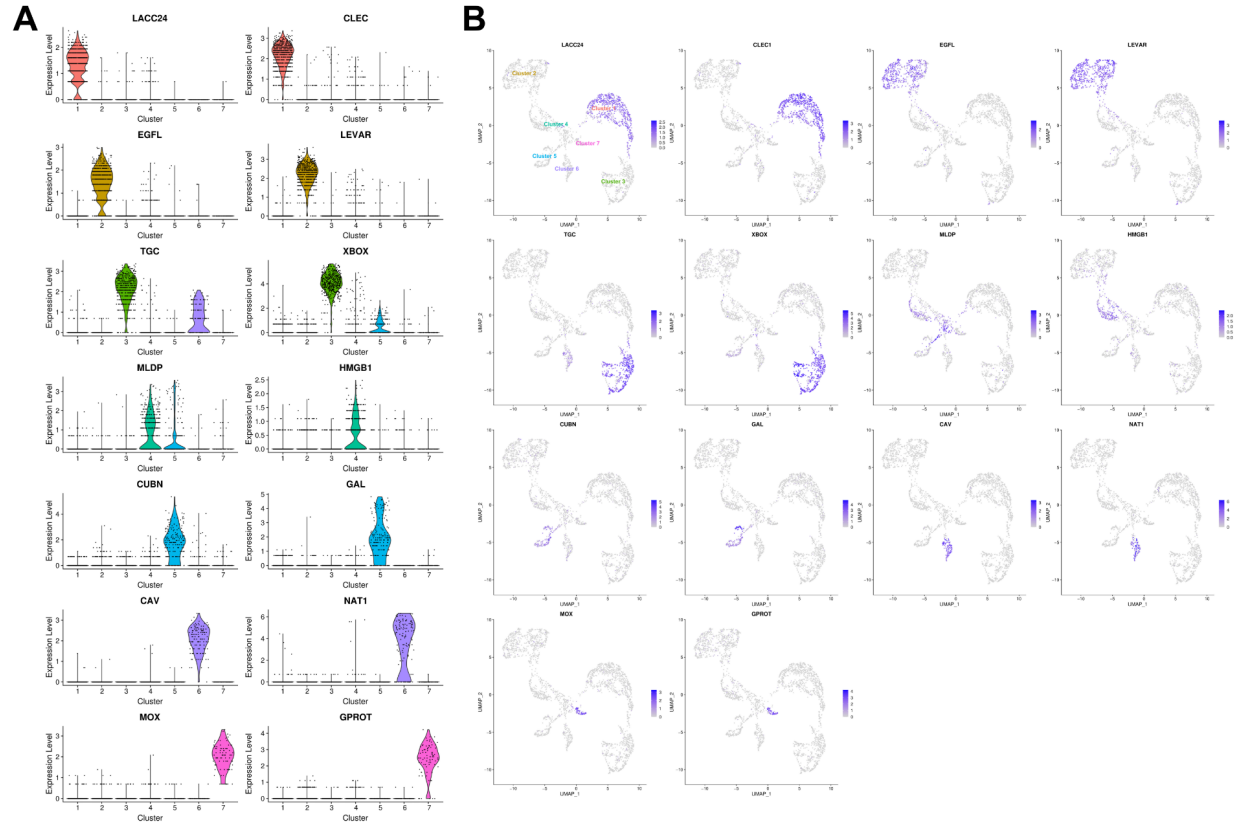


# B



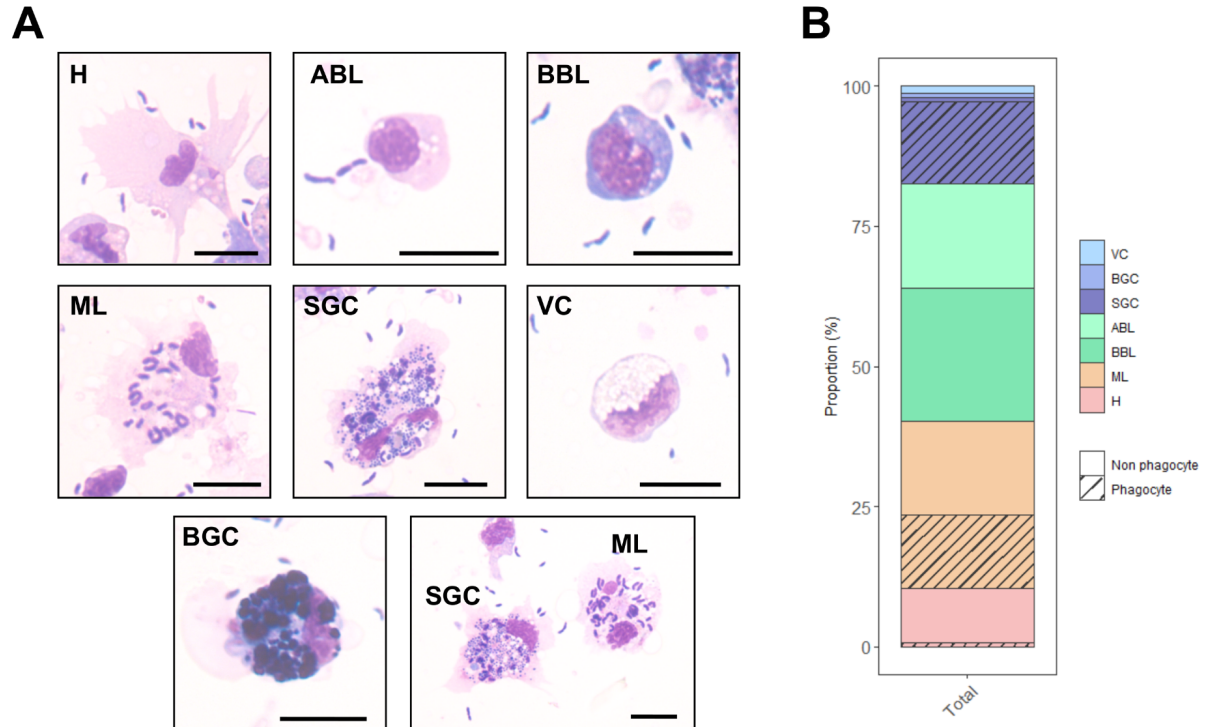
**Figure S4. Statistical significance of enrichment in different hemocyte types in Percoll gradient fractions.** (A) **a,b,c,d,e,f, and g** : MCDH staining of cells from the 7 fractions isolated from the percoll gradient in (B). Scale bar : 10  $\mu$ m. **Fraction 1 to 7** : Quantification of the different types of hemocytes found in each of the 7 fractions from 5 independent fractionation experiments. (B) Statistical significance of enrichment of different hemocyte types in Percoll gradient fractions. Results are from six independent experiments. Statistical significance is indicated by letters, as different letters indicate a significant difference between enrichments of cell types within the Percoll density gradient fractions (ANOVA, Tukey's test, p-value <0.05). Hyalinocytes (**H**) were significantly enriched in the first fraction compared to the other fractions and compared to unsorted hemocytes. However, they were significantly depleted in fractions 4, 5, 6 and 7 compared to unsorted hemocytes. Macrophage-like cells (**ML**) were significantly enriched in fractions 3 and 4 compared to fractions 1, 6 and 7. They were depleted in fraction 1 compared to unsorted hemocytes. Acidophilic blasts (**ABL**) were significantly depleted in fractions 4, 5, 6, and 7 compared to unsorted hemocytes. Basophilic blasts (**BBL**) were significantly enriched in fractions 2 and 3 compared to fractions 4, 5, 6, and 7 and in fraction 1 compared to fractions 6 and 7. Compared to unsorted hemocytes, basophilic blasts (**BBL**) were significantly enriched in fraction 2 and depleted in fractions 6 to 7. Small granule cells (**SGC**) were significantly depleted in fractions 1, 2, and 3 compared to fractions 4, 5, 6, and 7, and also significantly depleted in fractions 4 and 5 compared to fractions 6 and 7. In addition, small granule cells (**SGC**) were significantly enriched in fractions 5, 6, and 7 compared to unsorted hemocytes. The distribution of the big granule cells (**BGC**) showed a significant enrichment in fraction 4, compared to fractions 1 and 2, but no significant changes were observed with unsorted hemocytes. Vesicular cells (**VC**) were enriched in fraction 5 compared to fractions 1, 2,

3, 6, 7 and unsorted hemocytes in fractions 1, 2, and 3 compared to fraction 5 and enriched in fraction 5 compared to fraction 7 and unsorted hemocytes.

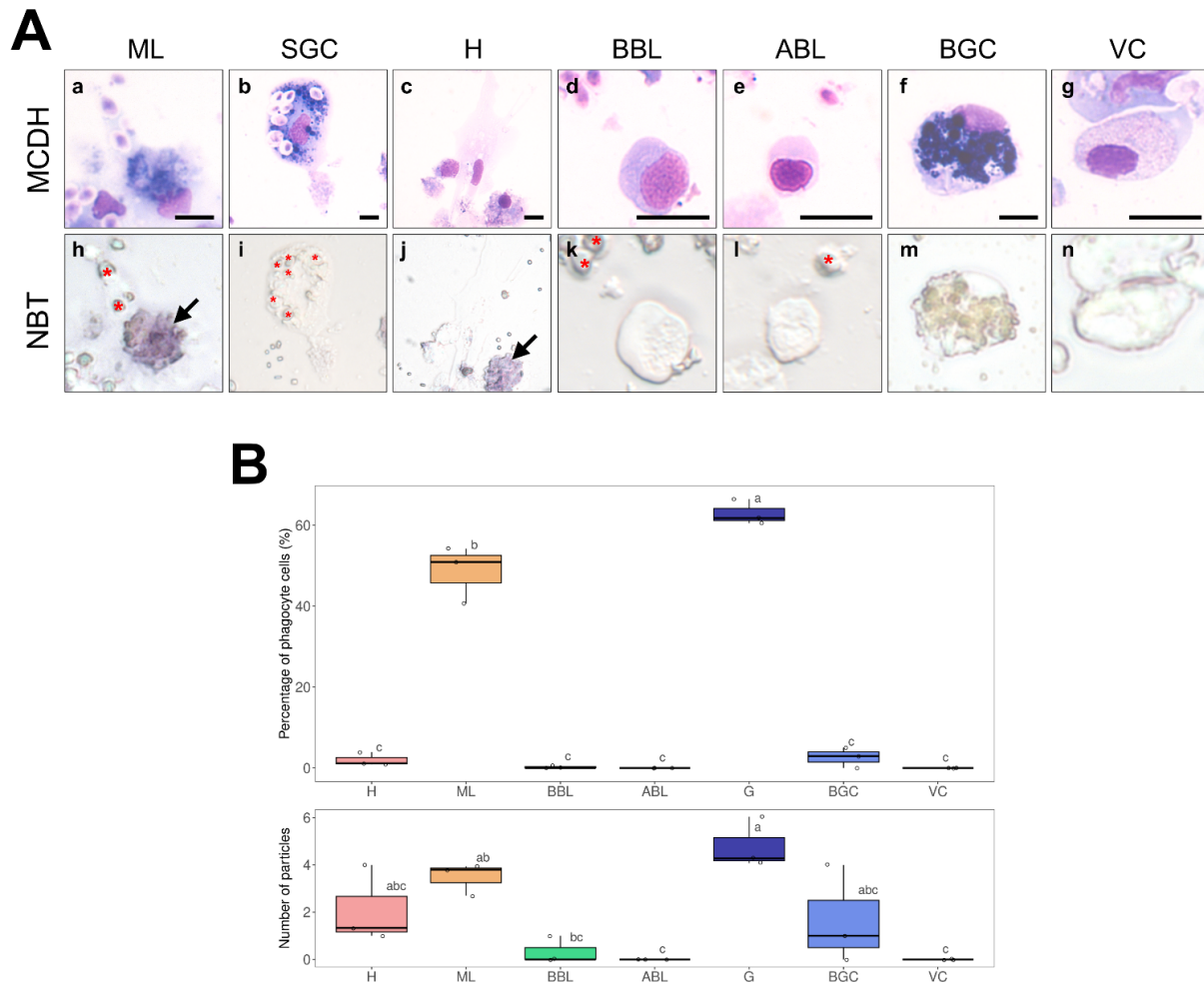


**Figure S5. Cluster specificity and expression level of the 14 selected cluster markers A)** Violin graph showing the average expression level (Log2FC) of the 14 selected marker transcripts specific to the different scRNA-seq clusters. **B)** Identification of cells expressing the selected markers on the UMAP plot. Positive cells are colored purple according to the Log2FC value. **LACC24** : Laccase 24, **CLEC** : C-type lectin domain-containing protein, **EGFL** : EGF-like domain-containing protein 8, **LEVAR** : Putative regulator of levamisole receptor-1, **TGC** : TGc domain-containing protein, **XBOX** : X-box binding protein-like protein, **MLDP** : ML domain-containing protein, **HMGB1** : High mobility group protein B1, **CUBN** : Cubilin, **GAL** : Galectin, **CAV** : Caveolin, **NAT1** : Natterin-

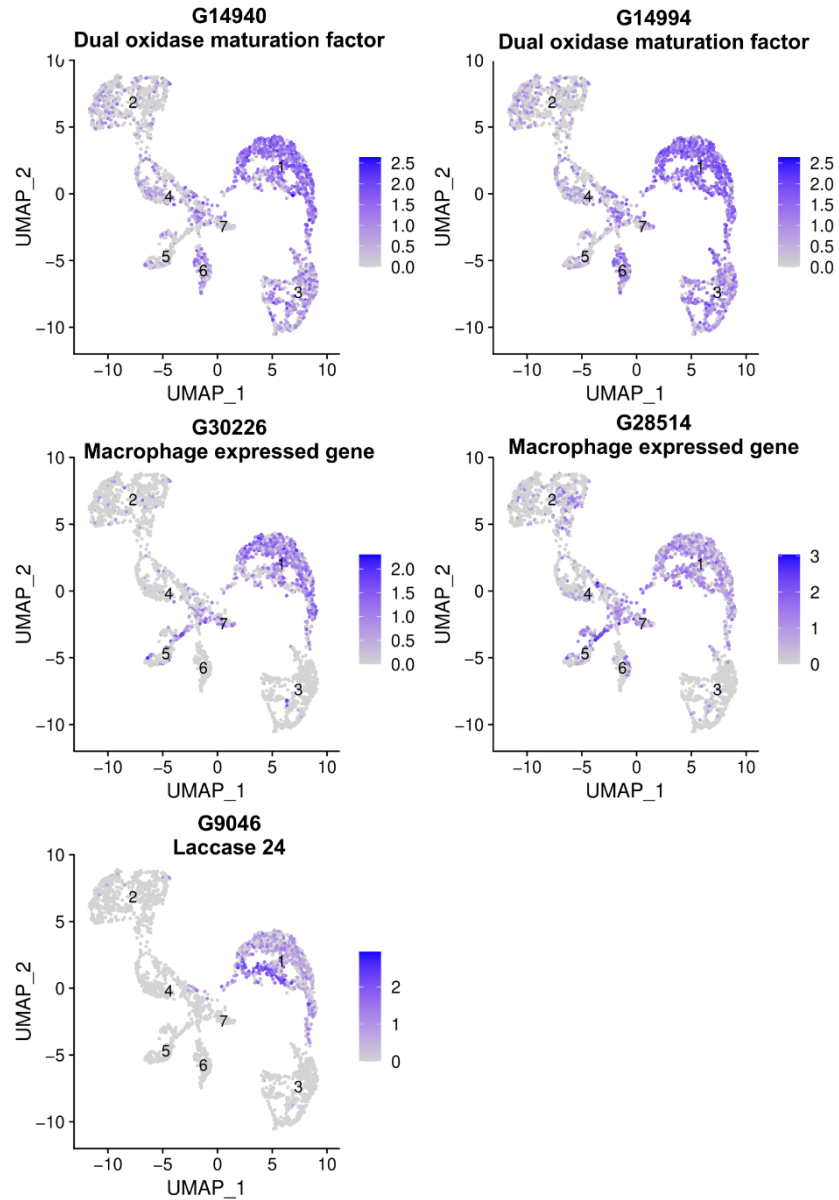
1, **MOX** : DBH-like monooxygenase protein 1, **GPROT** : G protein receptor F1-2 domain-containing protein.



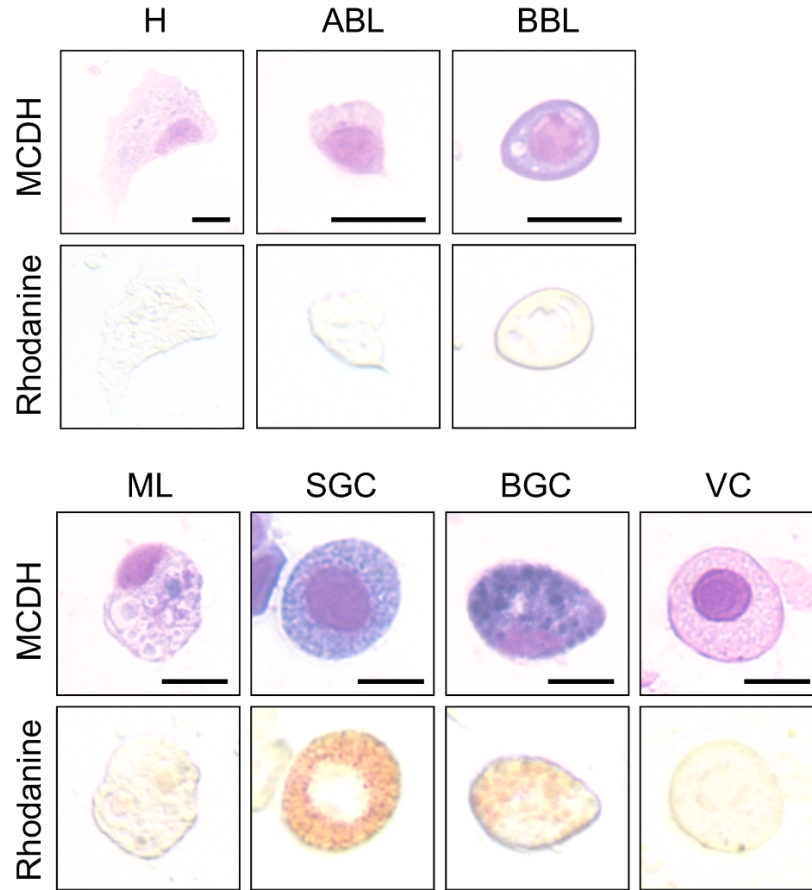
**Figure S6. Measurement of the ability of *C.gigas* hemocytes to phagocytose *Vibrio tasmaniensis* LMG20012<sup>T</sup>.** Oyster hemocytes were challenged with non-pathogenic *Vibrio tasmaniensis* LMG20012<sup>T</sup> and phagocytosis was measured by observing intracellular bacteria after MCDH staining. **(A)** MCDH staining of hemocytes after phagocytosis assay. Scale bar : 10µm. **(B)** Bar plot showing the proportion of each cell type and the proportion of phagocytic cells. **H** : Hyalinocytes, **ABL** : Acidophilic Blast-Like cells, **BBL** : Basophilic Blast-Like cells, **ML** : Macrophage-Like cells, **SGC** : Small Granule Cells, **VC** : Vesicular Cells and **BGC** : Big Granule Cells.



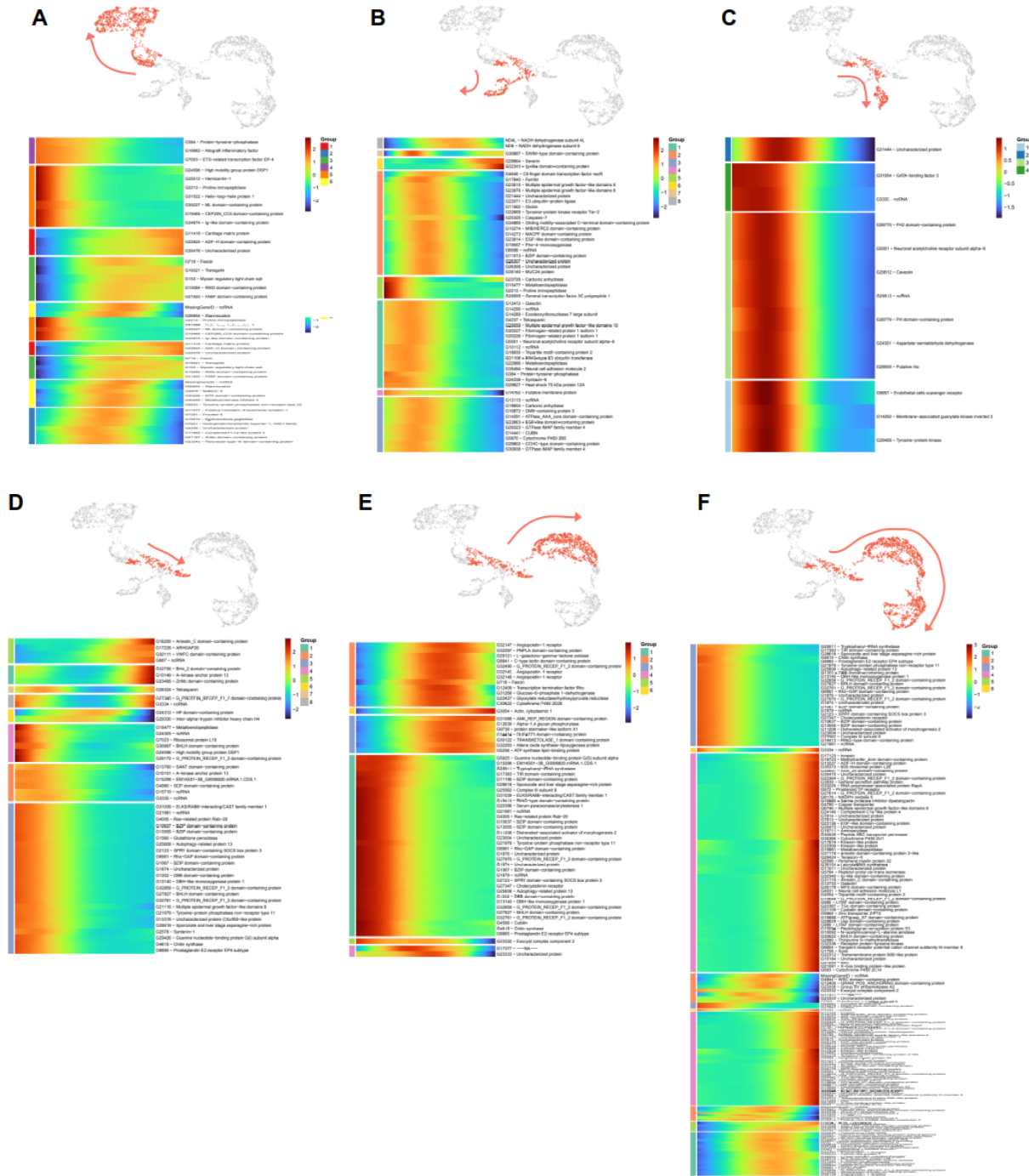
**Figure S7. NBT (NitroBlueTetrazolium) staining of oyster hemolymph exposed to zymosan particles. (A)** Hemocyte morphology after MCDH staining : Macrophage Like (a), Small Granule Cells (b), Hyalinocyte (c), Basophilic (d) and Acidophilic (e) Blast cells, Big Granule Cells (f) and Vesicular Cells (g). NBT staining of the different hemocyte types (h-n). Red stars show zymosan and bacteria particles. Black arrows identify Macrophage-Like cells. Scale bar : 10  $\mu$ m. **(B)** Quantification of the phagocytic activity of each cell type for zymosan particles from 3 independent experiments. Results of quantification of the phagocytic activity of each cell type and number of zymosan particles per cell type. The graph shows the result of 3 independent experiments.



**Figure S8. Uniform Manifold Approximation and Projection (UMAP) plots of cells expressing Macrophage-Like markers.** Cluster numbers are indicated on each cluster. Each point in the UMAP plot represents a single hemocyte, and the clustering of these points reveals the distinct transcriptional profiles of macrophage-like specific markers within the hemocyte population.

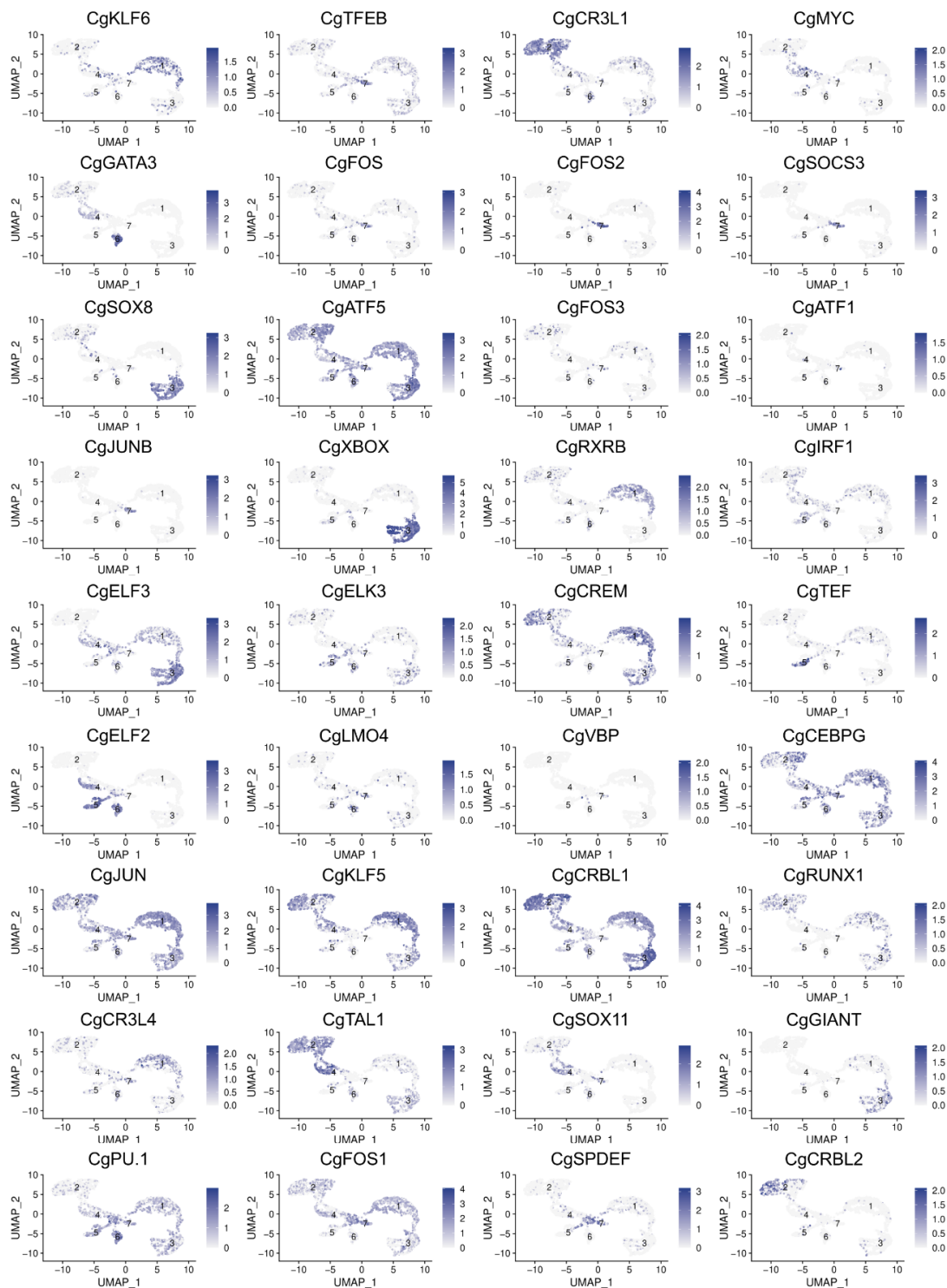


**Figure S9. Labeling of intracellular copper stores in *C.gigas* hemocytes.** MCDH (upper panels) and rhodanine (lower panels) staining of oyster hemocytes to reveal copper accumulation. Cells were first processed for copper staining and then stained according to MCDH protocol. **H** : Hyalinocytes, **ABL** : Acidophilic Blast-Like cells, **BBL** : Basophilic Blast-Like cells, **ML** : Macrophage-Like cells, **SGC** : Small Granule Cells, **VC** : Vesicular Cells and **BGC** : Big Granule Cells. Bar : 10μm.



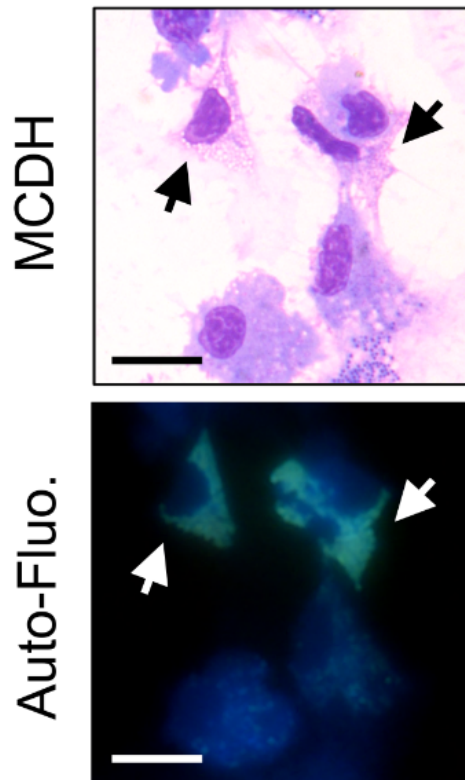
**Figure S10. Results of scRNA-seq trajectory analysis using Monocle3.** Analysis was performed using cluster 4 as the zero pseudotime. **(A)** Lineage from immature cells to hyalinocytes, **(B)** to cluster 5 cells, **(C)** to cluster 6 cells, **(D)** to vesicular cells, **(E)** to macrophage-like cells and **(F)** to small granule cells. For each lineage, cell trajectories are shown

in red and heat maps of pseudo time-dependent genes are shown. Blue indicates low expression, and red indicates high expression. Pseudotime flows from right to left.



**Figure S11. Uniform Manifold Approximation and Projection (UMAP) plots of cells expressing transcription factors. 28 UMAP representation for the transcription factors**

identified in the scRNA-seq dataset. Each UMAP plot shows cells expressing the transcription factor in purple. Log2FC expression level is also reported.



**Figure S12. Observation of autofluorescence of vesicular cells in hemolymph.** Freshly punctured total hemolymph was cytospun, directly observed under a microscope using a DAPI filter set (DAPI Blue ex : 350/50 nm, DC : 400 nm and em : 460/50 nm) and then processed for MCDH staining. Arrows indicate autofluorescent cells. Scale bar : 10  $\mu$ m