Single-cell RNA-seq analysis reveals penaeid shrimp hemocyte subpopulations and cell differentiation process

Keiichiro Koiwai [a, *], Takashi Koyama [b, c], Soichiro Tsuda [d], Atsushi Toyoda [e], Kiyoshi Kikuchi [b], Hiroaki Suzuki [f], Ryuji Kawano [a]

[a] Department of Biotechnology and Life Science, Tokyo University of Agriculture and Technology, Koganei-shi, Tokyo 184-8588, Japan

[b] Fisheries Laboratory, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Hamamatsu, Shizuoka 431-0214, Japan

[c] Graduate School of Fisheries and Environmental Sciences, Nagasaki University, Bunkyo-machi, Nagasaki-shi, Nagasaki 852-8521, Japan

[d] bitBiome Inc., Waseda University Incubation Center, Nishiwaseda, Shinjuku-ku, Tokyo 169-0051, Japan

[e] Advanced Genomics Center, National Institute of Genetics, Mishima, Shizuoka 411-8540, Japan

[f] Department of Precision Mechanics, Faculty of Science and Engineering, Chuo University, Bunkyo-ku, Tokyo 112-8551, Japan

* Keiichiro Koiwai

Email: koiwai@go.tuat.ac.jp

Author Contributions: K. Koiwai designed the experiments; K. Koiwai, T.K., S.T., and A.T. performed the experiments; K. Koiwai analyzed the data; K. Kikuchi, H.S., and R.K. supervised the research; K. K. wrote the paper.

Keywords: scRNA-seq; non-model; hemocytes; cell differentiation; crustacean; shrimp

This PDF file includes:

Main Text
Figures 1 to 7

Abstract

Crustacean aquaculture is expected to be a major source of fishery commodities in the near future. Hemocytes are key players of the immune system in shrimps; however, their classification, maturation, and differentiation are still under debate. To date, only discrete and inconsistent information on the classification of shrimp hemocytes has been reported, showing that the morphological characteristics are not sufficient to resolve their actual roles. Our present study using single-cell RNA sequencing, revealed nine types of hemocytes of *Marsupenaeus japonicus* based on their transcriptional profiles. We identified markers of each subpopulation and the differentiation pathways involved in their maturation. We also discovered cell growth factors that might play crucial roles in hemocyte differentiation. Different immune roles among these subpopulations were suggested from the analysis of differentially expressed immune-related genes. These results provide a unified classification of shrimp hemocytes, which improves the understanding of its immune system.

Introduction

Aquaculture is an important source of animal protein and is considered one of the most important long-term growth areas of food production, providing 60% of fish for human consumption (1) (http://www.fao.org/fishery/statistics/en). However, crustaceans that lack an adaptive immune system (2-4) are vulnerable to pathogens. This means that ordinal vaccination is not applicable to crustaceans, unlike in fish aquaculture. Shrimp is the main target species for crustacean aquaculture. Therefore, an immune priming system for shrimp, which is entirely different from conventional vaccines, needs to be developed to control the infection of pathogens. However, little is known about the immune system of crustaceans due to the lack of biotechnological tools, such as uniform antibodies and other biomarkers (5).

Hemocytes, which are immune cells of crustaceans, are traditionally divided into three morphological types based on the dyeing of intracellular granules, which was established by Bauchau and colleagues (6-8). However, there have been additional reports on the classification of the hemocytes of shrimp; they were classified into four, eight, and five types based on electron microscopy (9), another dyeing method (10), and iodixanol density gradient centrifugation (11), respectively. As the morphology and dye staining properties of shrimp hemocytes are not absolute indicators, no unified understanding of their role has been established yet. Molecular markers, such as specific mRNAs, antibodies, or lectins, are usually available for characterizing the subpopulations of cells in model organisms, but this is not often the case for non-model organisms. Although monoclonal antibodies have been developed for some hemocytes of shrimp (12-18), their number is lower than that of humans, and their correspondence to the cell type, as well as their differentiation stage are under debate.

Recently, single-cell mRNA sequencing (scRNA-seq) techniques have dramatically changed this scene, allowing researchers to annotate non-classified cells solely based on the mRNA expression patterns of each cell. In particular, droplet-based mRNA sequencing, such as Drop-seq, developed by Macosko, Basu

59 (19), has gained popularity for classifying cells and identifying new cell types. The enormous amount of
60 biological data obtained from scRNA-seq leads us to classify cells into specific groups, analyze their
61 heterogeneity, predict the functions of single-cell populations based on the gene expression profiles, and
62 determine the cell proliferation or development pathways based on the pseudo-time ordering of a single cell
63 (20, 21). More recently, hemocytes of invertebrate, fly, and mosquito models have been subjected to these
64 types of microfluidic-based scRNA-seq to reveal their functions (22, 23).

65 Here, we performed scRAN-seq analysis on *Marsupenaeus japonicus* hemocytes to classify the
66 hemocyte types and to characterize their functions using the custom-built Drop-seq platform. To perform
67 scRNA-seq, a high-quality gene reference is essential; however, such reference genomes are scarce for
68 crustaceans because of the extremely high proportion of simple sequence repeats (5). We circumvented this
69 problem by preparing reference genomes using hybrid *de novo* assembly of short- and long-read RNA
70 sequencing results. The sequences obtained from the scRNA-seq were mapped onto the reference genes
71 successfully. Our scRNA-seq uncovered the transcriptional profiles of a few thousand *M. japonicus*
72 hemocytes. We identified the markers of each population and the differentiation pathways associated with
73 their maturation. We also discovered the cell growth factors that might play crucial roles in hemocyte
74 differentiation. Different immune roles among these subpopulations were also suggested from the analysis
75 of differentially expressed immune-related genes. Our results present a unified classification of shrimp
76 hemocytes and a deeper understanding of the immune system of shrimp.

77

78 **Results**

79 **scRNA-seq clustering of *Marsupenaeus japonicus* hemocytes**

80 Our study utilized scRNA-seq to determine the cellular subtypes with a distinct transcriptional
81 expression (Fig. 1 A). To map the scRNA-seq sequences from *M. japonicus* hemocytes, we first prepared *de*
82 *novo* assembly of the reference genes using hybrid assembly of short- and long-read RNA sequencing results.
83 Then, by using self-built Drop-seq microfluidic chips, single hemocytes were captured and their mRNA was
84 barcoded using the droplet-based strategy. This process was performed in triplicates for three shrimp
85 individuals. Following library preparation and sequencing, the transcriptomes obtained from scRNA-seq
86 were mapped against the reference genes to discover the cell types.

87 Using the Drop-seq procedure, we profiled a total of 2,704 cells and obtained a median value of 718
88 unique molecular identifiers (UMIs) and 334 genes per cell across three replicates (Fig. S1 A and B).
89 Approximately 300 genes were detected, and the total number of mRNAs expressed among individual cells
90 varied between 100 and 1,400 (Fig. S1 C and D). There was some transcriptional variability, which may have
91 resulted from the artifacts of the Drop-seq system, because it is consistent with the original Drop-seq paper
92 (19) and with recent findings in other organisms, such as fish (24), flies (23), and mosquitoes (22, 25).

93 Applying the SCTransform batch correction method integrated into the Seurat package allowed us
94 to remove the individual differences. SCTransform successfully integrated all three shrimp datasets, among

which we identified a total of nine clusters (Fig. 1 B) and obtained 3,334 commonly expressed genes. Each cluster contained the following number of cells: Hem 1, 65 cells (2.4%); Hem 2, 461 cells (17.0%); Hem 3, 313 cells (11.6%); Hem 4, 524 cells (19.4%); Hem 5, 92 cells (3.4%); Hem 6, 553 cells (20.5%); Hem 7, 135 cells (5.0%); Hem 8, 170 cells (6.3%); and Hem 9, 391 cells (14.5%), respectively.

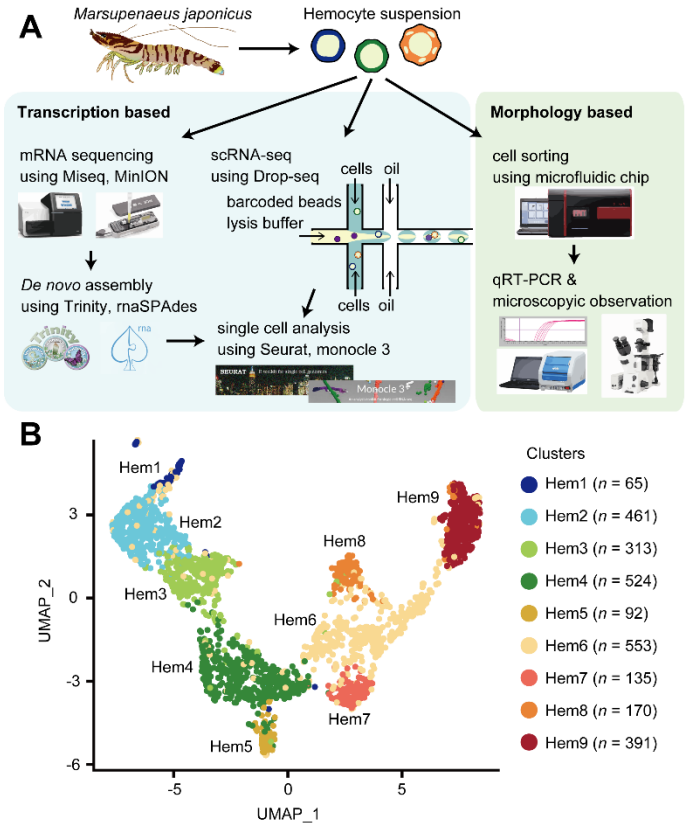

Figure 1. scRNA-seq of penaeid shrimp *M. japonicus* hemocytes. (A) Schematics of the microfluidics-based scRNA-seq, preparation of the *de novo* assembled gene list, *in silico* analysis workflow and morphology-based cell classification. (B) UMAP (uniform manifold approximation and projection) plot of SCTransform batch corrected and integrated of hemocytes from three shrimps (n = 2,704).

Cluster specific markers and their functional prediction

A total of 40 cluster-specific markers were predicted using the Seurat FindMarkers tool (Fig. 2, Dataset S1, and Dataset S2). For each cluster, seven (Hem 1), six (Hem 2), one (Hem 3), four (Hem 4), eight (Hem 5), five (Hem 7), one (Hem 8), and eight (Hem 9) markers were selected. Their functions were then annotated using BLASTX searching for penaeid shrimp identical proteins downloaded from a public database.

Hem 1 specific markers, *histone acetyltransferase* lysine acetyltransferase 6A (*HAT KAT6A)*, *gamma-aminobutyric acid* (*GABA*) *transporter*, and *polypyrimidine tract-binding protein 1* (*PTBP1*) are genes related to cell proliferation, cell migration, and colony formation in human tumor studies. HAT

114 KAT6A is known to be a chromatin regulator that controls fundamental cellular processes and is implicated
115 in regulating tumor progression (26). Autocrine/paracrine signaling via GABA receptors negatively controls
116 ES cell and peripheral neural crest stem cell proliferation (27). This GABA signaling pathway critically
117 regulates proliferation independently of differentiation, apoptosis, and overt damage to DNA (27). PTBP1 is
118 a multi-functional RNA-binding protein that is overexpressed in glioma, a type of tumor that occurs in the
119 brain, and a decreased expression of PTBP inhibits cell migration and increases the adhesion of cells to
120 fibronectin and vitronectin (28, 29). PTBP has been shown to be involved in germ cell differentiation in
121 *Drosophila melanogaster* and is essential for the development of *Xenopus laevis* (30). These pieces of
122 evidence strongly suggest that these Hem 1 markers are related to cell proliferation in shrimp hemocytes.
123 Among the six markers found in cluster Hem 2, three markers were annotated with hemocyte
124 transglutaminase (HemTGase), and one marker showed high similarity with von Willebrand factor D and
125 epidermal growth factor (EGF) domain-containing protein (VWDE). Interestingly, both Hem 1 and Hem 2
126 showed a high expression of TGase, an immature hemocyte marker of crayfish and shrimp. When the
127 extracellular TGase is digested, hemocytes start to differentiate into mature hemocytes (31-33). The high
128 expression of TGase suggests that both Hem 1 and Hem 2 are in the early stage of hemocytes. In cluster Hem
129 3, the only identified marker showed high similarity with tubulointerstitial nephritis antigen (TINAGL). In
130 cluster Hem 4, all markers showed similarity with the hypothetical protein.

131 In clusters Hem 5 to Hem 9, many of the cluster-specific markers showed similarity with immune-
132 related genes of penaeid shrimp, and very few genes were unknown. We postulate that this is partly because
133 the public database is rich in immune-related genes because of their importance. In cluster Hem 5, markers
134 showed a high similarity with c-type lysozyme, viral responsive protein (VRP), fibrous sheath CABYR-
135 binding (FSCB)-like, chitin binding-like protein, anti-lipopolysaccharide factor (ALF)-A1, and
136 PDGF/VEGF-related factor 1. In cluster Hem 7, markers showed high similarity with selenium-dependent
137 glutathione peroxidase (Se-GPX), BigPEN, ALF-C1, and hemocyte Kunitz protease inhibitor (KPI). In
138 cluster Hem 8, the marker showed high similarity with KPI. In cluster Hem 9, the markers showed similarity
139 with insulin-like growth factor-binding protein (IGFBP)-related protein 1, hemocyte KPI, crustacean
140 hematopoietic factor (CHF)-like protein, single whey acidic protein domain (SWD)-containing protein, and
141 penaeidin-II.

142 Altogether, only 70% (28/40) of markers were annotated from the BLAST searches on the penaeid
143 shrimp identical protein data. We also performed GO annotation using eggNOG-mapper (34, 35)
144 (http://eggnog-mapper.embl.de/) to predict the functions of contigs, but only about 10% of the contigs were
145 annotated, and we could not reach GO analysis.

146 From the results of this cluster-specific marker analysis, we found that Hem 1 and Hem 2 are
147 immature hemocytes. Therefore, we analyzed the whole maturation process of hemocytes using pseudo-
148 temporal analysis. Additionally, cell growth-related genes, such as *VWDE*, *TINGAL*, *IGFBP-related protein*,
149 and *CHF*, were identified as cluster-specific markers. We predicted their functions to pursue the linkage

between these genes and the hemocyte differentiation process. Finally, the immune-related genes, which were enriched in Hem 5 to 9, were analyzed to predict the immune function of each cluster.

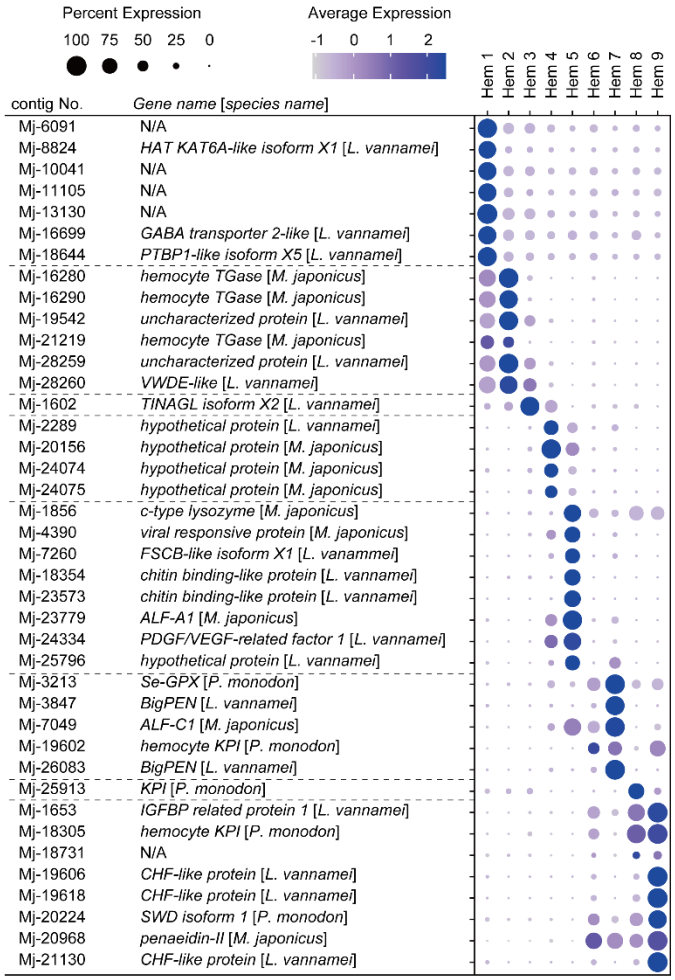

Figure 2. Dot plot representing the marker genes per cluster based on the average expression predicted using the Seurat FindMarker tool. Color gradient of the dot represents the expression level, while the size represents the percentage of cells expressing any gene per cluster.

Pseudo temporal ordering of cells delineates hemocyte lineages

To investigate the dynamics of hemocyte differentiation, we performed lineage-tree reconstruction using the monocle3 learn_graph function. The differentiation and proliferation of hemocytes in shrimp and other crustaceans are still under debate (36). Since the cell cycle- and hemocyte-type-specific markers are better studied in *Drosophila*, we checked the commonly expressed genes of *M. japonicus* that are similar to the markers of *Drosophila* to determine whether they are present in shrimp. In the search for cell cycle-specific markers (Dataset S3), a large portion of cells in Hem 1 expressed the G2/M-related genes of *Drosophila*: *heterochromatin protein 1b* (*HP1b*); Mj-13141, *CCCTC-binding factor* (*CTCF*); Mj-14904,

166 *suppressor of variegation 205* (*Su(var)205*); Mj-27796 (Fig. 3 A and Fig. S2). *Drosophila HP1* and
167 *Su(var)205* are known to be essential for the maintenance of the active transcription of the euchromatic genes
168 functionally involved in cell-cycle progression, including those required for DNA replication and mitosis
169 (37, 38). CTCF has zinc finger domains and plays an important role in the development and cell division of
170 fly and mammalian cells (39, 40). In the BLAST search of two genes related to cell-cycle progression, that
171 is, Mj-13141 and Mj-27796, on identical proteins of penaeid shrimp, they showed high similarity with the
172 *chromobox protein homolog 1-like* (Dataset S1). A chromobox family protein contributes to
173 lymphomagenesis by enhancing stem cell self-renewal and/or by increasing the replicative potential of cancer
174 stem cells in human tumor cells (41). BLAST search of Mj-14904 on identical proteins of penaeid shrimp
175 showed similarity with the zinc finger protein that contains the same domain of CTCF (Dataset S1). These
176 findings suggested that hemocytes grouped as Hem 1 are tightly regulated by these G2/M phase-related genes
177 to promote cell division.

178 Among the specific markers of four types of *Drosophila* hemocytes, four genes in prohemocytes,
179 11 genes in plasmatocytes, 11 genes in lamellocytes, and 6 genes in crystal cells showed similarity with the
180 shrimp genes (Dataset S4). However, no genes were significantly expressed in certain clusters (Fig. S3).
181 Therefore, we concluded that these *Drosophila* hemocyte markers cannot be adapted as markers for the
182 hemocytes of shrimp.

183 We considered Hem 1 to be the initial state of hemocytes and set it as the starting point in the
184 differentiation process because Hem 1 expressed cell proliferation-related genes, TGase (an immature
185 hemocyte marker), and G2/M phase-related genes (37-42). From this pseudo temporal ordering analysis, we
186 found four main lineages starting from Hem 1 to Hem 5, Hem 7, Hem 8, and Hem 9 at the endpoints (Fig. 3
187 B and Fig. S4 A-D). In crayfish, hematopoietic stem cells are present in hematopoietic tissue (HPT), and two
188 types of hemocyte lineages starting from a hematopoietic stem cell exist (43). In *Penaeus monodon*, hyaline
189 cells (i.e., agranulocytes) are considered as the young and immature hemocytes of two types of matured
190 hemocytes (9). Our pseudo temporal ordering analysis revealed that the hemocytes of *M. japonicus*
191 differentiate from a single subpopulation into four major populations. The differentiation process of *M.*
192 *japonicus* hemocytes was continuous, not discrete, which was in agreement with previous arguments on the
193 crustacean hematopoiesis mechanism (9, 43, 44).

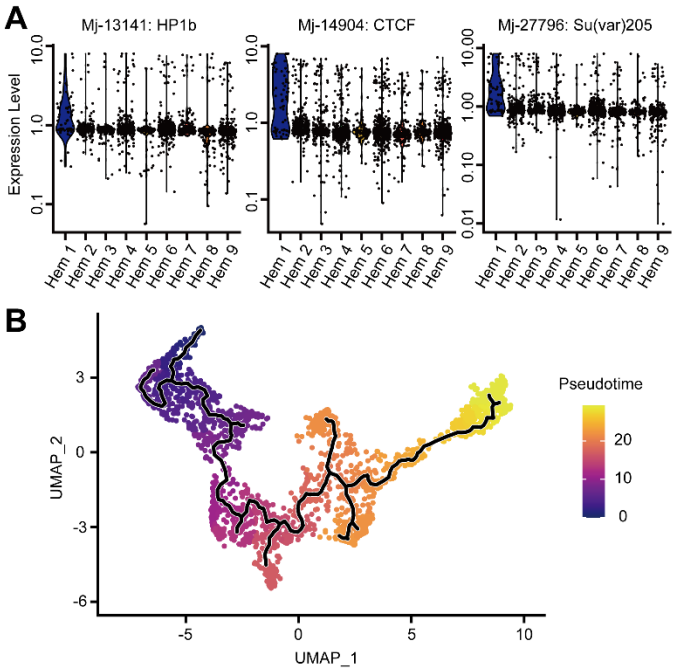

Figure 3. Pseudo temporal ordering of hemocyte lineages. (A) Violin plots displaying normalized expression levels of each cell cycle related genes across all clusters. (B) Visualization of clusters (from Figure 1B) onto the pseudo time map using monocle 3.

Expression of cell growth-related genes

The exploration of cluster-specific markers revealed that cell growth-related genes were specifically expressed in certain clusters. Therefore, we highlighted five genes that are predicted to be involved in cell growth and differentiation: *VWDE-like*, *TINAGL*, *PDGF/VEGF-related factor 1*, *IGFBP-related protein 1*, and *CHF-like* (Fig. 4, Fig. S5, and Dataset S5).

VWDE-like was highly expressed in clusters Hem 1 to Hem 3 (Fig. 4 A and B). The expression of VWDE is a common feature of blastemas, which are capable of regenerating limbs and fins in a variety of highly regenerative species (45). VWDE contains several epidermal growth factor (EGF) domains and is expected to be a downstream effector once a blastema has been established, but it is not a driver of blastema formation (45). VWDE-like of *M. japonicus* was expressed more strongly in Hem 2 than in Hem 1, indicating that VWDE-like also works as a downstream effector against Hem 1, which is predicted to be composed of undifferentiated hemocytes.

TINAGL, which was expressed in Hem 3 and 4, is a secreted extracellular protein that is essential for early angiogenesis in developing zebrafish embryos (46) and humans (47). TINAGL of *M. japonicus* has been proposed to participate in angiogenic cell differentiation. TINAGL is also known as a suppressor of cancer progression and metastasis by binding to receptors of EGF in humans (48). Interestingly, TINAGL of *M. japonicus* was expressed in Hem 3 and Hem 4, in which the expression of VWDE-like was weakened

216 (Fig. 4 A and B). This suggests that TINAGL might suppress the differentiation function of VWDE-like in
217 Hem 3 and 4 in shrimp.
218 Cells expressing PDGF/VEGF-related factors were dominant in Hem 4 and Hem 5 (Fig. 4 A and
219 B). The vascular endothelial growth factor (VEGF) signaling pathway is essential for vasculogenesis, cell
220 proliferation, and tumor migration in mammals (49, 50). Furthermore, in *Drosophila*, VEGF homologs
221 control the number of circulating hemocytes (51). The high expression of PDGF/VEGF-related factors in
222 Hem 4 and Hem 5 showed that Hem 3 would differentiate into Hem 5 through Hem 4 via the VEGF signaling
223 pathway.
224 Hem 6 to Hem 9 expressed insulin-like growth factor (IGF) binding protein (IGFBP) (Fig. 4 A and
225 B). IGFBP delivers IGFs to the target cells in mammal studies and is essential for cell growth or
226 differentiation (52). The high expression of a receptor of the insulin-like peptide at mature hemocytes in the
227 mosquito suggests that the insulin signaling pathway regulates hemocyte proliferation (53). The silencing
228 and overexpression of IGFBP caused a decrease and increase in the growth of hemocytes, respectively, in
229 the loss and gain function study of abalone *Haliotis diversicolor* (54). These studies indicate that the IGFBP-
230 related insulin signaling pathway is important for hemocyte proliferation and differentiation in invertebrates.
231 IGFBP might play an essential role in the differentiation of hemocytes from Hem 4 to Hem 6–9 in shrimp.
232 The high expression of IGFBP is determined in the brain and gonads of *Litopenaeus vannamei* (55). This
233 fact also suggests that IGFBP plays a possible role in organ growth and maturation in shrimp.
234 Up to this point, the analyzed genes were proliferation- and differentiation-promoting, but there was
235 also the specific expression of the hemocyte homeostasis regulatory gene, crustacean hematopoietic factor
236 (CHF) (43) (Fig. 4 A and B). CHF is a hematopoietic factor of crayfish, and the silencing of CHF leads to an
237 increase in the apoptosis of cells in HPT, and a reduction in the number of circulating hemocytes (33).
238 Additionally, the silencing of laminin, a receptor of CHF, reduces the number of circulating hemocytes by
239 decreasing the number of agranulocytes, as opposed to granulocytes, in *P. vannamei* (55). CHF-like was
240 expressed at cluster Hem 9 (Fig. 4 A and B), in which hemocytes are predicted to be matched here. Taken
241 together, CHF-like expressed from matured hemocytes, Hem 9, might work as a hematopoietic factor against
242 agranulocytes or regulate the homeostasis of agranulocytes.

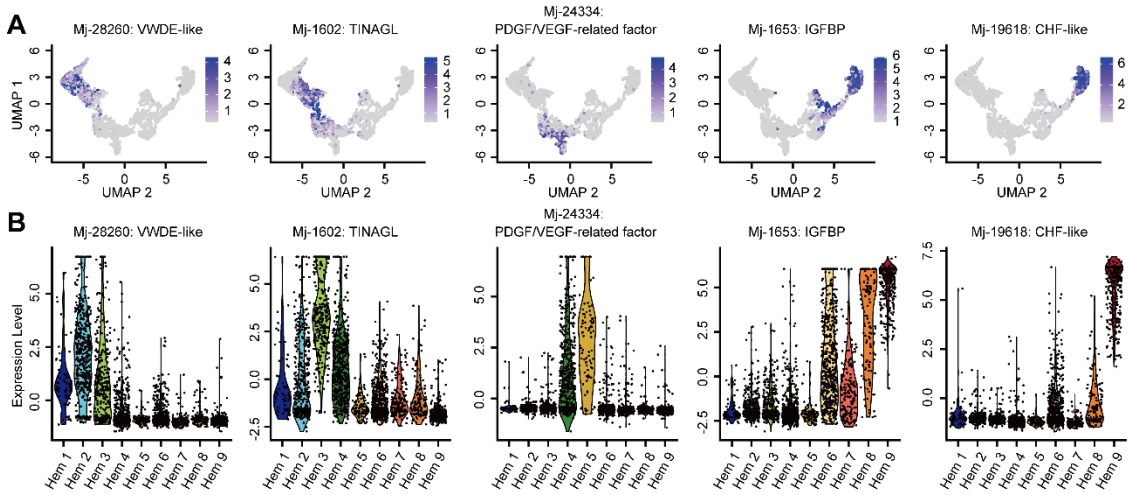

Figure 4. Cell growth-related gene expression across all clusters. (A) Expression pattern of the cell growth-related genes across cell clusters. (B) Violin plots displaying normalized expression levels of each cell growth-related genes across all clusters.

Expression of immune-related genes in single cells

Hemocytes of shrimp play key roles in their immunity; therefore, we selected immune-related genes from 3,334 commonly expressed genes, and then analyzed their expression levels to deduce the detailed immune functions of each cluster.

Antimicrobial peptides (AMPs) play the most important role in the immunity of shrimps and are well known to be stored in granulocytes (56, 57). The expression patterns of AMPs revealed that the major AMPs of penaeid shrimp were expressed in clusters Hem 5 to Hem 9 (Fig. 5 A-F, Fig. S6 and Dataset S6). Therefore, Hem 5 to Hem 9 were predicted to be granulocytes. These AMP-expressing clusters were broadly classified into two groups. One group was Hem 5, which expressed several AMPs, such as ALF-A1 and c-type lysozyme (Fig. 5 D and F), but also expressed other immune-related genes different from AMPs: *chitin-binding protein* and *virus responsible protein* (*VRP*) (Fig. 2). Chitin-binding protein is located on the cell surface and interacts with virus envelope proteins in shrimp (58). VRP is distributed in granulocytes, and infection by pathogenic viruses causes an increase in VRP transcripts (59). The other group is Hem 6 to 9, which strongly expresses major AMPs, such as penaeidin, stylicin, and SWD, suggesting that this group corresponds to the most studied immune-related cell subpopulation to date. The level of penaeidin-positive hemocytes is increased after bacterial infection (60), but is reduced after a virus infection (61). Stylicin and SWD mRNA expression was decreased after virus infection in *M. japonicus* and *P. monodon*.

Together, the functions of these two groups can be characterized as follows. In Hem 5, VRP, whose level is known to be increased by viral infection, was specifically expressed, while AMPs were downregulated. Thus, Hem 5 is predicted to be the group that contributes to the immune response against viruses. Additionally, a virus infection causes an increase in the expression of both VEGF and its receptors

269 in hemocytes to regulate a downstream signaling pathway (62-64). The high expression of PDGF/VEGF-
270 related factors at Hem 5 (Fig. 2 and 4) supports our prediction that Hem 5 plays an essential role in shrimp
271 biodefense against viruses. Previously, a subtype of granulocytes, termed semi-granular cells (SGC), were
272 isolated from *L. vannamei*, in which lysozyme was highly expressed. It is possible that Hem 5 in *M. japonicus*
273 corresponds to SGC found in *L. vannamei* because c-type lysozyme was strongly expressed in Hem 5, in our
274 study. Conversely, the AMPs specifically expressed at Hem 6–9 were upregulated by bacterial infection.
275 Therefore, these clusters contribute to bacterial defense. Different types of ALF play different roles in shrimp
276 immunity, which improves the synergism in shrimp antimicrobial defenses (65). The expression patterns of
277 ALFs were also different in each cluster (Fig. 5 C and D, Fig. S6 and Dataset S6), suggesting that different
278 clusters have different immunological functions.

279 Some of the hemocyte-type-specific markers related to their immune function have been studied in
280 crayfish: prophenoloxidase (proPO) in matured hemocytes, copper/zinc superoxide dismutase (SOD) in
281 SGC, and Kazal-type proteinase inhibitor (KPI) in granular cells (GC), and transglutaminase (TGase) in
282 immature cells (66). We BLAST searched homologies between these genes and common expressed genes to
283 check their potential as specific markers. As a result, 31 genes were found to have homology with the genes
284 listed above (Fig. S6 and Dataset S6). TGase was strongly expressed in immature clusters Hem 1 and Hem
285 2, and can, therefore, be used as a marker of immature hemocytes (Fig. 5 I and Fig. S6). ProPOs were not
286 highly expressed in the present single-cell study. Both KPI and SOD were mostly expressed at both Hem 8
287 and Hem 9, and their expression levels were similar between these clusters (Fig. 5 G and H and Fig. S6).
288 These results suggest that the functional segregation of hemocytes in shrimp is different from that in crayfish,
289 in which these molecular markers are expressed distinctly between GC and SGC. The KPI of *M. japonicus*
290 is expressed in only some hemocytes in healthy shrimps, and bacterial infection causes an increase in KPI
291 expression (67). Therefore, it can be predicted that the number of KPI-positive hemocytes in Hem 8 and 9
292 would be increased upon a pathogen infection, in shrimp.

293 In the mosquito, scRNA-seq revealed a new subpopulation called "antimicrobial granulocytes" that
294 expressed characteristic AMPs (22). Similarly, in *M. japonicus*, the expression patterns of immune-related
295 genes were also different among certain clusters, suggesting that shrimp hemocytes are more heterogeneous
296 than previously thought. It is anticipated that the class of granulocytes discussed in previous studies is actually
297 a mixture of clusters exhibiting different roles.

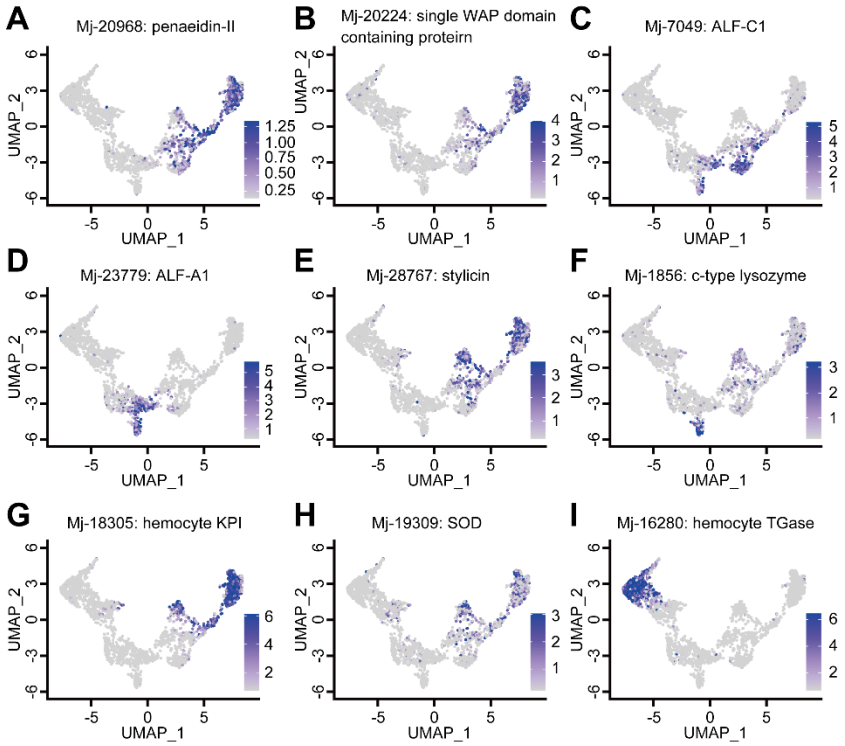

Figure 5. Expression pattern of the immune-related genes across cell clusters. (A) Penaeidin-II, (B) single WAP domain containing protein, (C) ALF-CA, (D) ALF-A1, (E) stylicin, (F) c-type lysozyme, (G) KPI, (H) SOD, and (I) TGase.

Validation of marker genes and the relationship between clusters and morphology

Our scRNA-seq results revealed nine major subpopulations and their marker genes, and the possible differentiation trajectory of *M. japonicus* hemocytes. Next, we examined the correlation between the morphology and expression of marker genes. Two major populations of hemocytes were sorted based on the forward versus side scatter plot obtained using microfluidic-based fluorescence-activated cell sorter (FACS). The sorted populations were observed using microscopy. FACS was able to separate hemocytes into two morphologically different populations (Fig. 6 A–E): smaller cells with low internal complexity in region 1 (R1) (50.7% ± 5.2%) and larger cells with high internal complexity in region 2 (R2) (47.7% ± 3.4%). From DIC and dye staining imaging (Fig. 6 A and B), we observed that cells in the R1 region contained no or few granules in the cytoplasm. The nucleus occupied a large portion of the volume in these cells (Fig. 6 C). Conversely, those in the R2 region had many granules in the cytoplasmic region, which occupied a large portion of cells (Fig. 6 D).

We conducted qRT-PCR analysis to determine the expression of some representative genes in these populations. The results showed that the ΔΔCt values of the transcripts of HemTGase were higher in R1 hemocytes than in R2 hemocytes, while the values of penaeidin, crustin, stylicin, proPO, and Cu/ZnSOD2 were higher in R2 hemocytes (Fig. 6 F). The ΔΔCt values of transcripts of c-type lysozyme and VRP were

similar in R1 and R2 hemocytes. The combination of FACS and qRT-PCR results confirmed that the gene expression of the two populations in shrimp hemocytes was roughly divided based on morphology, that is, agranulocytes and granulocytes, which is consistent with our scRNA-seq analysis. Granulocytes found in the R2 region expressed major AMPs, such as penaeidin, crustin, and stylicin, suggesting that they consist of clusters Hem 6 to Hem 9. However, c-type lysozyme and VRP, which are markers of Hem 5 (Fig. 2), were expressed in cells in both R1 and R2 regions, indicating that Hem 5 exists in both populations and is indistinguishable from the morphological characteristics. The number of cells occupying R1 and R2 was 50.7% ± 5.2% and 47.7% ± 3.4% (from n = 4), while the total number of cells classified as Hem 1 – 4 and Hem 5 – 9 in our single-cell analysis was 50.4% (1,363 cells) and 49.6% (1,341 cells), respectively.

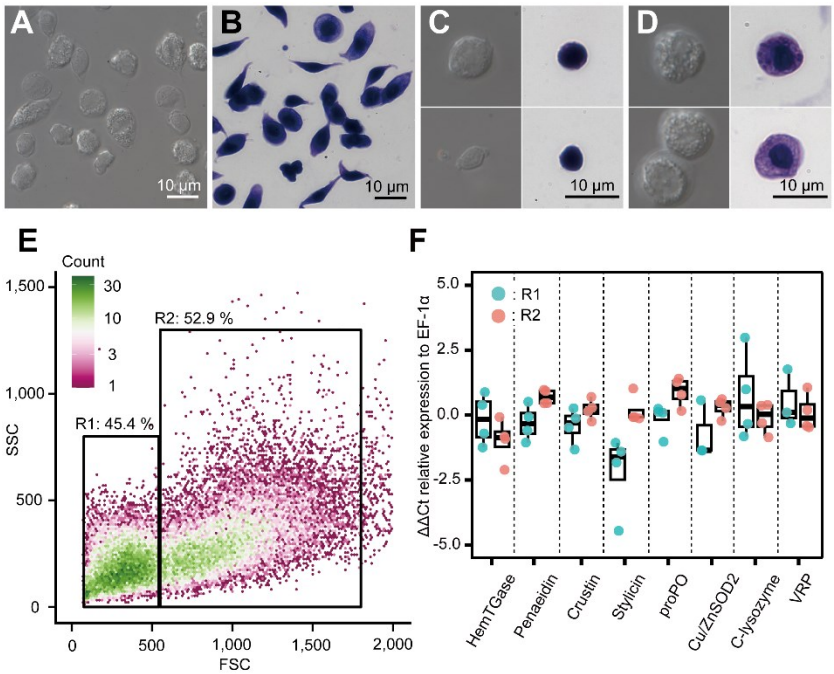

Figure 6. Morphological analysis of hemocytes and transcript profiles based on morphology. (A) DIC (Differential Interference Contrast) image of unsorted total hemocytes. (B) Dye stained total hemocytes. (C) DIC imaging and dye staining of R1 sorted hemocytes. (D) DIC imaging and dye staining of R2 sorted hemocytes. (E) FACS (fluorescence-activated cell sorting) analysis of hemocytes. Based on the FSC (forward scatter) and SSC (side scatter) two-dimensional space, two regions (R1 and R2) were obtained. (F) Differential gene expression analysis between R1 and R2 of hemocytes sorted using FACS. ΔΔCt values were analyzed using qRT-PCR. Higher ΔΔCt values indicate a higher accumulation of mRNA transcripts.

Discussion

Our single-cell transcriptome analysis revealed that there are nine subpopulations of hemocytes in shrimp. This result differs from any previous classification strategy based on various approaches, such as simple staining (7-9), monoclonal antibodies (18), flow cytometry (68), and lectin-binding profiles (69). We

341 hope that future studies will clarify the relationships between these phenotypic features and our
342 classification to comprehend the role of each cluster in the immune system of shrimp.

343 The cluster-specific markers and cell proliferation-related genes found here helped us to understand
344 how shrimp hemocytes differentiate. A strong expression of TGase, cell proliferation, and G2/M state-related
345 genes in Hem 1 suggested that hemocytes in this cluster are oligopotent and located upstream in the
346 differentiation process. In crustaceans, especially shrimp and crayfish, it is known that hemocytes are
347 produced in HPT, and that differentiated hematopoietic cells from HPT circulate in the body fluid (9, 36,
348 66). However, G2/M state hemocytes of *M. japonicus* exist in the hemolymph and account for only 0.63% ±
349 0.28% of the circulating hemocytes (70). Hem 1 accounts for only 2.4% of the analyzed cells. This similarity
350 in the fraction also indicates that only a very small fraction of oligopotent or initial state hemocytes exist
351 among the circulating hemocytes.

352 Our results also revealed that the growth-related genes were expressed at specific clusters (Fig. 4 A
353 and B). From these results, we propose that shrimp hemocytes differentiate as follows (Fig. 7 A and B): 1)
354 Hem 1 is the initial state of circulating hemocytes and has an oligopotent ability, which leaked out from HPT;
355 2) Hem 3 is differentiated from Hem 1 by VWDE-like through Hem 2; 3) Hem 4 is differentiated from Hem
356 3 by TINAGL; 4) Hem 5 is differentiated from Hem 4 by PDGF/VEGF-related factor; 5) Hem 7 to Hem 9
357 are differentiated from Hem 4 by IGFBP-related protein through Hem 6; 6) CHF-like is expressed in Hem 9
358 to maintain the immature hemocytes, Hem 1 to Hem 4. Loss and gain function studies of these differentiating
359 factors are necessary to prove the full differentiation process of penaeid shrimp hemocytes. Since we
360 identified a group of proliferating cells and differentiation factors here, the next step is to establish a way to
361 isolate and cultivate them. The markers of each cluster identified here will be good tools for isolating specific
362 cell types.

363 Crayfish's hemocytes and *Drosophila* hemocytes have been classified into three and four major
364 types according to the marker genes, respectively (23, 36). However, these crayfish or *Drosophila* marker
365 genes were found to be inadequate as cluster-specific markers for penaeid shrimp in our study. Insects and
366 crustaceans are thought to become independent about 500 million years ago (71, 72), and shrimp and crayfish
367 are thought to have become evolutionarily independent about 450 million years ago (73). Thus, it is
368 straightforward to reason that the functions of these genes have changed during the evolutionary process.
369 Therefore, we should not simply argue that the morphological and functional similarities between shrimp
370 and *Drosophila*/crayfish hemocytes are the same.

371 Most of the markers that were characteristically expressed in Hem 1 and Hem 4 could not be
372 functionally predicted by the BLAST search (Fig. 2). We speculate that the characteristic genes of Hem 1 are
373 associated with cell duplication and that Hem 4 is associated with the regulation of cell differentiation. The
374 division and differentiation mechanisms of shrimp hemocytes are still largely unknown, and no techniques
375 on culture shrimp hemocytes *in vitro* have been reported. The analysis of these unknown gene characteristics

may reveal these mechanisms. Notably, no specific marker gene was found in Hem 6, probably because these cells are yet to be specialized, unlike those in Hem 7 to Hem 9.

In conclusion, we succeeded in classifying shrimp hemocytes into nine subpopulations based on their transcriptional profiles, while they were only classified into two groups using FACS. Furthermore, our results imply that hemocytes differentiate from a single initial population. Although we have not yet successfully cultured crustacean hemocytes in passaging cultures, information on these subpopulations and marker genes will provide a foothold for hemocyte culture studies. Despite our success in the classification of hemocytes, we have not yet been able to fully understand the functions of each hemocyte group in detail. One reason for this is that the functions of some marker genes are still unknown. The present single-cell transcriptome data serves as a platform providing the necessary information for the continuous study of shrimp genes and their functions. Additionally, we have only determined the subpopulations of hemocytes in the normal state, so that our next goal will be to analyze the hemocytes in the infected state of certain diseases at different levels of cell maturation. In this way, we will be able to identify the major subpopulations that may work against the infectious agent. Likewise, it will be interesting to see how hemocytes from fertilized eggs mature. Additionally, single-cell analysis of hematopoietic tissues can also be expected to reveal more detailed differentiation mechanisms. Unlike terrestrial invertebrates, such as *Drosophila* and mosquitoes, shrimps are creatures that live in the ocean and are evolutionarily distant from each other. Understanding the immune system of shrimp species will require continuous effort, but we believe that it will provide efficient solutions to aquaculture problems.

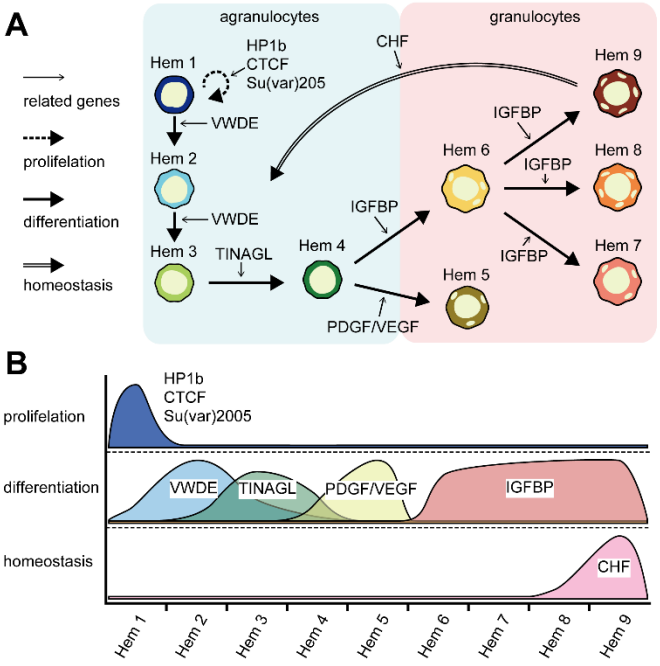

Figure 7. Model of the roles of cell growth-related genes in hemocyte differentiation. (A) Shrimp hemocyte differentiation process, 1) Hem 1 is the initial state of circulating hemocytes, which have oligopotent ability,

398 which leaked out from the hematopoietic tissue (HPT); 2) Hem 3 is differentiated from Hem 1 by VWDE-
399 like through Hem 2; 3) Hem 4 is differentiated from Hem 3 by TINAGL; 4) Hem 5 is differentiated from
400 Hem 4 by PDGF/VEGF-related factor; 5) Hem 7 to Hem 9 are differentiated from Hem 4 by IGFBP-related
401 protein through Hem 6; 6) CHF-like is expressed in Hem 9 to maintain the immature hemocytes, Hem 1 to
402 Hem 4. (B) Schematics of cell growth or homeostasis-related gene expression across clusters. The gene
403 expression distribution was cluster-specific.
404
405 **Materials and Methods**
406 **Shrimp and cell preparation**
407 Twenty-three female kuruma shrimp, *Marsupenaeus japonicus*, with an average weight of 20 g,
408 were purchased from a local distributor and maintained in artificial seawater with a 34 ppt salinity with a
409 recirculating system at 20°C. Hemolymph was collected using an anticoagulant solution suitable for penaeid
410 shrimp (7) from an abdominal site. The collected hemolymph was centrifuged at 800 x *g* for 10 min to collect
411 the hemocytes, which were then washed twice with PBS and the osmolarity was adjusted to kuruma shrimp
412 (KPBS: 480 mM NaCl, 2.7 mM KCl, 8.1 mM $Na_2HPO_4 \cdot 12H_2O$, 1.47 mM KH_2PO_4, pH 7.4).
413
414 **Preparation of expressing gene list of hemocytes**
415 *De novo* assembled transcript data were prepared as a reference for mapping Drop-seq data, because
416 the genome sequence of *M. japonicus* is still unknown. To improve the quality of the *de novo* assembled
417 transcript sequences, we prepared long-read mRNA sequences using MinION (Oxford Nanopore
418 Technologies) direct RNA sequencing to conduct hybrid *de novo* assembly. Poly(A) tailed RNA was purified
419 from 58 µg of total RNA from the hemocytes of 16 shrimp using Dynabeads Oligo$(dT)_{25}$ (Thermo Fisher
420 Scientific), and 500 ng of poly-(A) RNA was ligated to adaptors using a direct RNA sequencing kit (Oxford
421 Nanopore Technologies) according to the manufacturer's manual version DRS_9080_v2_revL_14Aug2019.
422 Finally, 44 ng of the library was obtained and sequenced using MinION, by using a MinION flow cell R9.4.1
423 (Oxford Nanopore Technologies). All sequencing experiments were performed using MinKNOW v3.6.5
424 without base calling. Raw sequence data were then base-called using Guppy v3.6.1. Once the raw signal from
425 the MinION fast5 files was converted into fastq files, the sequencing errors were corrected using TALC v1.01
426 (74)(https://gitlab.igh.cnrs.fr/lbroseus/TALC) by using the Illumina short reads sequence of kuruma shrimp
427 hemocytes (DDBJ Sequence Read Archive (DRA) accession number DRA004781). The corrected long-read
428 sequences from MinION and short-read sequences from Illumina Miseq were hybrid *de novo* assembled
429 using rnaSPAdes v3.14.1 (75)(https://cab.spbu.ru/software/rnaspades/) and Trinity 2.10.0
430 (76)(https://github.com/trinityrnaseq/trinityrnaseq/wiki). All assembled *de novo* transcripts were merged and
431 subjected to the EvidentialGene program v2022.01.20 (http://arthropods.eugenes.org/EvidentialGene/) to
432 remove similar sequences with a default parameter. The remaining sequences were renamed as Mj-XXX and
433 used as a hemocyte-expressing gene list. The assembled sequences and code used to perform base-calling

434 and *de novo* assembly are available on GitHub at https://github.com/KeiichiroKOIWAI/Drop-
435 seq_on_shrimp.
436
437 **Single-cell and single-bead encapsulation by a microfluidic device and exonuclease and reverse**
438 **transcribe reaction on a bead**
439 The drop-seq procedure was used to encapsulate single hemocytes and single mRNA capture beads
440 together into fL-scale microdroplets, as previously described (19). The following steps were performed in
441 triplicates for three shrimp individuals. Briefly, the self-built Drop-seq microfluidic device was prepared by
442 molding polydimethylsiloxane (PDMS; Sylgard 184, Dow Corning Corp.) from the microchannel structure
443 formed by the negative photoresist (SU-8 3050, Nippon Kayaku Co.). Using this device, droplets containing
444 a cell and a Barcoded Bead SeqB (ChemGenes Corporation) were produced up to 2 mL per sample using a
445 pressure pump system (Droplet generator, On-chip Biotechnologies Co., Ltd.). During the sample
446 introduction, the vial bottles containing cells and beads were shaken using a vortex mixer to prevent
447 sedimentation and aggregation (77). Droplets were collected from the channel outlet into the 50-mL corning
448 tube and incubated at 80°C for 10 min in a water bath to promote hybridization of the poly(A) tail of mRNA
449 and oligo d(T) on beads. After incubation, droplets were broken promptly and barcoded beads with captured
450 transcriptomes were reverse transcribed using Maxima H Minus Reverse Transcriptase (Thermo Fisher
451 Scientific) at RT for 30 min, then at 42°C for 90 min. Then, the beads were treated with Exonuclease I (New
452 England Biolabs) to obtain single-cell transcriptomes attached to microparticles (STAMP). The first-strand
453 cDNAs on beads were amplified using PCR. The beads obtained above were distributed throughout PCR
454 tubes (1,000 beads per tube), wherein 1× KAPA HiFi HS Ready Mix (KAPA Biosystems) and 0.8 μM 1st
455 PCR primer were included in a 25 μL reaction volume. PCR amplification was achieved using the following
456 program: initial denaturation at 95°C for 3 min; 4 cycles at 98°C for 20 s, 65°C for 45 s, and 72°C for 6 min;
457 12 cycles of 98°C for 20 s, 67°C for 20 s, and 72°C for 6 min; and a final extension at 72°C for 5 min. The
458 amplicons were pooled, double-purified with 0.9× AMPure XP beads (Beckman Coulter), and eluted in 100
459 μL of ddH_2O. Sequence-ready libraries were prepared according to Picelli, Bjorklund (78). A total of 1 ng of
460 each cDNA library was fragmented using home-made Tn5 transposome in a solution containing 10 mM
461 TAPS-NaOH (pH 8.5), 5 mM $MgCl_2$, and 10% dimethylformamide at 55°C for 10 min. The cDNA fragments
462 were purified using a DNA Clean & Concentrator Kit (Zymo Research) and eluted in 25 μL of ddH_2O. The
463 index PCR reaction was performed by adding 12 μL of the elute to a mixture consisting of 1× Fidelity Buffer,
464 0.3 mM dNTPs, 0.5 U KAPA HiFi DNA polymerase (KAPA Biosystems), 0.2 μM P5 universal primer, and
465 0.2 μM i7 index primer. Each reaction was achieved as follows: initial extension and subsequent denaturation
466 at 72°C for 3 min and 98°C for 30 s; 12 cycles of 98°C for 10 s, 63°C for 30 s, and 72°C for 30 s; and a final
467 extension at 72°C for 5 min. The amplified library was purified using 0.9× AMPure XP beads and sequenced
468 (paired-end) on an Illumina NextSeq 500 sequencer (NextSeq 500/550 High Output v2 kit (75 cycles); 20
469 cycles for read1 with custom sequence primer, 8 cycles for index read, 64 cycles for read2. Before performing

470 the Drop-seq on kuruma shrimp hemocytes, we validated the protocol by performing the same procedure
471 using a mixture of HEK293 and NIH3T3 cells and sequencing the test library using a Miseq Reagent Kit v3
472 (150 cycles).
473
474 **Analysis of single-cell data**
475 Paired-end reads were processed and mapped to the reference *de novo* assembled gene list of
476 hemocytes following the Drop-seq Core Computational Protocol version 2.0.0, and the corresponding Drop-
477 seq tools v2.3.0 (https://github.com/broadinstitute/Drop-seq) provided by McCarroll Lab
478 (http://mccarrolllab.org/dropseq/). The Picard suite (https://github.com/broadinstitute/picard) was used to
479 generate the unaligned bam files. The steps included the detection of barcode and UMI sequences, filtration
480 and trimming of low-quality bases and adaptors or poly(A) tails, and the alignment of reads using bowtie2
481 v2.4.1 (79) (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml). The cumulative distribution of reads
482 from the aligned bam files was obtained using BAMTagHistogram, and the number of cells was inferred
483 using Drop-seq tools.
484
485 **Data integration**
486 After digital expression data from 3 shrimps were read using Seurat v3.2.1 (80,
487 81)(https://satijalab.org/seurat/), SCTransform (81) was performed to remove the technical variation and to
488 select common expressing genes, while retaining biological heterogeneity. We ran a PCA using the
489 expression matrix of the top 3,000 most variable genes. The total number of principal components (PCs)
490 required to compute and store was 50. The UMAP was then performed using the following parameters:
491 n.neighbors, min.dist, and n.components were 10L, 0.1, and 2, respectively, to visualize the data in the two-
492 dimensional space, and then the clusters were predicted with a resolution of 0.5.
493
494 **Functional prediction of commonly expressed genes**
495 The predicted functions of selected assembled sequences as common expressing genes across three
496 replicates were searched using BLAST program v2.2.31 (82, 83)
497 (https://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/) on penaeid shrimp identical proteins
498 (downloaded from NCBI Identical Protein Groups on 19th of August, 2020) with the blastx parameter of e-
499 value as 0.0001 and num_alignments as 3. Then, the functions of each cluster were predicted based on the
500 marker genes. Marker genes were predicted using the Seurat FindAllMarkers tool with the following
501 parameters: min.pct as 0.5, logfc.threshold as 1, and test.
502
503 **Visualization of genes with distinctive functions on single hemocytes**
504 To visualize the genes with distinctive functions of shrimp, we extracted the distinct sequences from
505 commonly expressed genes based on their blastx results against penaeid shrimp identical proteins. Here, we

506 focused on the cell growth-related genes listed in Dataset S5 and on immune-related genes, such as
507 antimicrobial peptides (AMPs), transglutaminase, copper/zinc superoxide dismutase, and Kazal-type
508 proteinase inhibitors listed in Dataset S6. From these genes that showed a characteristic expression among
509 single-cell data, dot plots, violin plots, or feature plot visualizations were applied using Seurat functions.
510
511 **Comparison with *Drosophila* marker genes**
512 To check whether the *Drosophila* marker genes are applicable to shrimp, we performed a BLAST
513 search on *the Drosophila* cell cycle and cell type markers (https://github.com/hbc/tinyatlas). Common genes
514 among the three shrimp species were tblastx searched for *Drosophila melanogaster* genes (dmel-all-gene-
515 r6.34.fasta; downloaded from FlyBase https://flybase.org/) with the parameters of e-value as 0.0001 and
516 num_alignments as 3.
517
518 **Pseudo temporal ordering of cells using Monocle 3**
519 The integrated data of Seurat were transferred to Monocle 3 (84) (https://github.com/cole-trapnell-
520 lab/monocle3) to calculate a cell trajectory using the learn_graph function. We assigned the start point based
521 on the expression of cell proliferation-related genes and *Drosophila* marker genes of the cell cycle.
522
523 **Cell sorting of hemocytes and qRT-PCR of marker genes**
524 To validate the Drop-seq results on hemocytes, populations of hemocytes in the forward scatter
525 (FSC) and side scatter (SSC) two-dimensional space were sorted using a microfluidic cell sorter (On-chip
526 sort, On-chip Biotechnologies Co., Ltd.) from four shrimp individuals (Fig. S7). In the FSC/SSC two-
527 dimensional space, two main populations were predicted as Region1 (R1): small/simple and Region2 (R2):
528 large/complexity populations, which were defined as agranulocytes and granulocytes, respectively. After
529 sorting, some sorted hemocytes were immediately fixed in 2% formalin in KPBS and stained with a May-
530 Grunwald and Giemsa staining solution to observe the cellular components. Non-stained and stained
531 hemocytes were subjected to microscopy IX71 (Olympus Corporation) to observe their structures.
532 Total RNA was also collected from sorted cells and pre-sorted cells. The concentration of RNA was
533 measured using a nanodrop, and cDNA was transcribed using a High-Capacity cDNA Reverse Transcription
534 Kit (Thermo Fisher Scientific). Constructed cDNA was diluted five times with TE buffer and subjected to
535 qRT-PCR using KOD SYBR qPCR (TOYOBO Co. Ltd.), following the manufacturer's protocol. The
536 expression of each gene was calculated using the ΔΔCT method against elongation factor-1 alpha and total
537 hemocytes.
538
539 **Data files and analysis code**
540 The raw sequence data of newly sequenced *M. japonicus* transcriptomic reads were archived in the
541 DDBJ Sequence Read Archive (DRA) of the DNA Data Bank of Japan as follows: MinION mRNA direct

542 sequencing: DRA010948; Drop-seq shrimp rep1: DRA010950; shrimp rep2: DRA010951; shrimp rep3:
543 DRA010952; mixture sample of HEK293 and 3T3 cells: DRA010949. The Seurat digital expression data
544 were archived in the Genomic Expression Archive of the DNA Data Bank of Japan: E-GEAD-403. Fast5
545 data of MinION direct RNA sequencing will be made available upon request from the authors. The code used
546 to perform *de novo* assembly, clustering, and marker analysis is available on GitHub at
547 https://github.com/KeiichiroKOIWAI/Drop-seq_on_shrimp. The key resources are listed in Dataset S7.
548

549 **Acknowledgments**

550 We would like to thank Fumiko Sunaga for her technical support in performing cell sorting and
551 analysis, and Editage (www.editage.com) for English language editing. This work was supported by JSPS
552 KAKENHI (JP19J00539 and JP20K15603) to K. Koiwai; a Grant-in-Aid for Scientific Research on
553 Innovative Areas (17H06425) to K. Kikuchi.
554

555 **Competing Interest Statement:** The authors declare no conflict of interest.
556

557 **References**

558 1. FAO. Fishery and aquaculture statistics 2018/FAO annuaire. Rome: FAO; 2020.

559 2. Jiravanichpaisal P, Lee BL, Söderhäll K. Cell-mediated immunity in arthropods: hematopoiesis,
560 coagulation, melanization and opsonization. Immunobiology. 2006;211(4):213-36.

561 3. Tassanakajon A, Somboonwiwat K, Supungul P, Tang S. Discovery of immune molecules and their
562 crucial functions in shrimp immunity. Fish Shellfish Immunol. 2013;34(4):954-67.

563 4. Flegel TW. A future vision for disease control in shrimp aquaculture. Journal of the World Aquaculture
564 Society. 2019;50(2):249-66.

565 5. Zhang X, Yuan J, Sun Y, Li S, Gao Y, Yu Y, et al. Penaeid shrimp genome provides insights into
566 benthic adaptation and frequent molting. Nat Commun. 2019;10(1):356.

567 6. Bauchau AG. Cruataceans. Invertabrate blood cells. 1981;2:385-420.

568 7. Söderhäll K, Smith VJ. Separation of the haemocyte populations of *Carcinus maenas* and other marine
569 decapods, and prophenoloxidase distribution. Dev Comp Immunol. 1983;7(2):229-39.

570 8. Johansson MW, Keyser P, Sritunyalucksana K, Söderhäll K. Crustacean haemocytes and
571 haematopoiesis. Aquaculture. 2000;191(1-3):45-52.

572 9. van de Braak CB, Botterblom MH, Liu W, Taverne N, van der Knaap WP, Rombout JH. The role of the
573 haematopoietic tissue in haemocyte production and maturation in the black tiger shrimp (*Penaeus*
574 *monodon*). Fish Shellfish Immunol. 2002;12(3):253-72.

575 10. Kondo M, Tomonaga S, Takahashi Y. Granulocytes with cytoplasmic deposits of kuruma prawn.
576 Aquaculture Science. 2012;60(1):151-2.

577 11. Dantas-Lima JJ, Tuan VV, Corteel M, Grauwet K, An NTT, Sorgeloos P, et al. Separation of *Penaeus*
578 *vannamei* haemocyte subpopulations by iodixanol density gradient centrifugation. Aquaculture.
579 2013;408-409:128-35.

580 12. Rodriguez J, Boulo V, Mialhe E, Bachère E. Characterisation of shrimp haemocytes and plasma
581 components by monoclonal antibodies. J Cell Sci. 1995;108:1043-50.

582 13. van de Braak CB, Taverne N, Botterblom MH, van der Knaap WP, Rombout JH. Characterisation of
583 different morphological features of black tiger shrimp (*Penaeus monodon*) haemocytes using
584 monoclonal antibodies. Fish Shellfish Immunol. 2000;10(6):515-30.

585 14. Sung HH, Wu PY, Song YL. Characterisation of monoclonal antibodies to haemocyte subpopulations
586 of tiger shrimp (*Penaeus monodon*): immunochemical differentiation of three major haemocyte types.
587 Fish Shellfish Immunol. 1999;9(3):167-79.

588 15. Sung HH, Sun R. Use of monoclonal antibodies to classify hemocyte subpopulations of tiger shrimp
589 (*Penaeus monodon*). Journal of Crustacean Biology. 2002;22(2):337-44.

590 16. Winotaphan P, Sithigorngul P, Muenpol O, Longyant S, Rukpratanporn S, Chaivisuthangkura P, et al.
591 Monoclonal antibodies specific to haemocytes of black tiger prawn *Penaeus monodon*. Fish Shellfish
592 Immunol. 2005;18(3):189-98.

593 17. Lin Y, Zhan W, Li Q, Zhang Z, Wei X, Sheng X. Ontogenesis of haemocytes in shrimp
594 (*Fenneropenaeus chinensis*) studied with probes of monoclonal antibody. Dev Comp Immunol.
595 2007;31(11):1073-81.

596 18. Xing J, Chang Y, Tang X, Sheng X, Zhan W. Separation of haemocyte subpopulations in shrimp
597 *Fenneropenaeus chinensis* by immunomagnetic bead using monoclonal antibody against granulocytes.
598 Fish Shellfish Immunol. 2017;60:114-8.

599 19. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, et al. Highly parallel genome-wide
600 expression profiling of individual cells using nanoliter droplets. Cell. 2015;161(5):1202-14.

601 20. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, et al. The dynamics and regulators
602 of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol.
603 2014;32(4):381-6.

604 21. Soneson C, Robinson MD. Bias, robustness and scalability in single-cell differential expression
605 analysis. Nat Methods. 2018;15(4):255-61.

606 22. Raddi G, Barletta ABF, Efremova M, Ramirez JL, Cantera R, Teichmann SA, et al. Mosquito cellular
607 immunity at single-cell resolution. Science. 2020;369(6507):1128-32.

608 23. Tattikota SG, Cho B, Liu Y, Hu Y, Barrera V, Steinbaugh MJ, et al. A single-cell survey of *Drosophila*
609 blood. eLife. 2020;9:e54818.

610 24. Carmona SJ, Teichmann SA, Ferreira L, Macaulay IC, Stubbington MJ, Cvejic A, et al. Single-cell
611 transcriptome analysis of fish immune cells provides insight into the evolution of vertebrate immune
612 cell types. Genome Res. 2017;27(3):451-61.

613 25. Severo MS, Landry JJM, Lindquist RL, Goosmann C, Brinkmann V, Collier P, et al. Unbiased
614 classification of mosquito blood cells by single-cell genomics and high-content imaging. Proc Natl
615 Acad Sci U S A. 2018;115(32):E7568-E77.

616 26. Lv D, Jia F, Hou Y, Sang Y, Alvarez AA, Zhang W, et al. Histone acetyltransferase KAT6A
617 upregulates PI3K/AKT signaling through TRIM24 binding. Cancer Res. 2017;77(22):6190-201.

618 27. Andang M, Hjerling-Leffler J, Moliner A, Lundgren TK, Castelo-Branco G, Nanou E, et al. Histone
619 H2AX-dependent GABA(A) receptor regulation of stem cell proliferation. Nature.
620 2008;451(7177):460-4.

621 28. Cheung HC, Hai T, Zhu W, Baggerly KA, Tsavachidis S, Krahe R, et al. Splicing factors PTBP1 and
622 PTBP2 promote proliferation and migration of glioma cell lines. Brain. 2009;132(Pt 8):2277-88.

623 29. Shibayama M, Ohno S, Osaka T, Sakamoto R, Tokunaga A, Nakatake Y, et al. Polypyrimidine tract-
624 binding protein is essential for early mouse development and embryonic stem cell proliferation. FEBS
625 J. 2009;276(22):6658-68.

626 30. Jiao H, Dong P, Yan L, Yang Z, Lv X, Li Q, et al. TGF-beta1 Induces polypyrimidine tract-binding
627 protein to alter fibroblasts proliferation and fibronectin deposition in keloid. Sci Rep. 2016;6:38033.

628 31. Lin X, Söderhäll K, Söderhäll I. Transglutaminase activity in the hematopoietic tissue of a crustacean,
629 *Pacifastacus leniusculus*, importance in hemocyte homeostasis. BMC Immunol. 2008;9:58.

630 32. Huang CC, Sritunyalucksana K, Söderhäll K, Song YL. Molecular cloning and characterization of tiger
631 shrimp (*Penaeus monodon*) transglutaminase. Dev Comp Immunol. 2004;28(4):279-94.

632 33. Lin X, Söderhäll K, Söderhäll I. Invertebrate hematopoiesis: an astakine-dependent novel
633 hematopoietic factor. J Immunol. 2011;186(4):2073-9.

634 34. Huerta-Cepas J, Szklarczyk D, Heller D, Hernandez-Plaza A, Forslund SK, Cook H, et al. eggNOG 5.0:
635 a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms
636 and 2502 viruses. Nucleic Acids Res. 2019;47(D1):D309-D14.

637 35. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, et al. Fast genome-
638 wide functional annotation through orthology assignment by eggNOG-Mapper. Mol Biol Evol.
639 2017;34(8):2115-22.

640 36. Söderhäll I. Crustacean hematopoiesis. Dev Comp Immunol. 2016;58:129-41.

641 37. De Lucia F, Ni JQ, Vaillant C, Sun FL. HP1 modulates the transcription of cell-cycle regulators in
642 *Drosophila melanogaster*. Nucleic Acids Res. 2005;33(9):2852-8.

643 38. Paro R, Hogness DS. The polycomb protein shares a homologous domain with a heterochromatin-
644 associated protein of *Drosophila*. Proc Natl Acad Sci U S A. 1991;88(1):263-7.

645 39. Mohan M, Bartkuhn M, Herold M, Philippen A, Heinl N, Bardenhagen I, et al. The *Drosophila*
646 insulator proteins CTCF and CP190 link enhancer blocking to body patterning. EMBO J.
647 2007;26(19):4203-14.

648 40. Rasko JE, Klenova EM, Leon J, Filippova GN, Loukinov DI, Vatolin S, et al. Cell growth inhibition by
649 the multifunctional multivalent zinc-finger factor CTCF. Cancer Res. 2001;61(16):6002-7.

650 41. Scott CL, Gil J, Hernando E, Teruya-Feldstein J, Narita M, Martinez D, et al. Role of the chromobox
651 protein CBX7 in lymphomagenesis. Proc Natl Acad Sci U S A. 2007;104(13):5389-94.

652 42. Junkunlo K, Söderhäll K, Noonin C, Söderhäll I. PDGF/VEGF-related receptor affects
653 transglutaminase activity to control cell migration during crustacean hematopoiesis. Stem Cells Dev.
654 2017;26(20):1449-59.

655 43. Lin X, Söderhäll I. Crustacean hematopoiesis and the astakine cytokines. Blood. 2011;117(24):6417-
656 24.

657 44. Noonin C, Lin X, Jiravanichpaisal P, Söderhäll K, Söderhäll I. Invertebrate hematopoiesis: an anterior
658 proliferation center as a link between the hematopoietic tissue and the brain. Stem Cells Dev.
659 2012;21(17):3173-86.

660 45. Leigh ND, Sessa S, Dragalzew AC, Payzin-Dogru D, Sousa JF, Aggouras AN, et al. von Willebrand
661 factor D and EGF domains is an evolutionarily conserved and required feature of blastemas capable of
662 multitissue appendage regeneration. Evol Dev. 2020;22(4):297-311.

663 46. Brown LJ, Alawoki M, Crawford ME, Reida T, Sears A, Torma T, et al. Lipocalin-7 is a matricellular
664 regulator of angiogenesis. PLoS One. 2010;5(11):e13905.

665 47. Mary S, Kulkarni MJ, Mehendale SS, Joshi SR, Giri AP. Tubulointerstitial nephritis antigen-like 1
666 protein is downregulated in the placenta of pre-eclamptic women. Clin Proteomics. 2017;14:8.

667 48. Shen M, Jiang YZ, Wei Y, Ell B, Sheng X, Esposito M, et al. Tinagl1 suppresses triple-negative breast
668 cancer progression and metastasis by simultaneously inhibiting integrin/FAK and EGFR signaling.
669 Cancer Cell. 2019;35(1):64-80.

670 49. Neufeld G, Cohen T, Gengrinovitch S, Poltorak Z. Vascular endothelial growth factor (VEGF) and its
671 receptors. The FASEB Journal. 1999;13(1):9-22.

672 50. Peichev M, Naiyer AJ, Pereira D, Zhu Z, Lane WJ, Williams M, et al. Expression of VEGFR-2 and
673 AC133 by circulating human CD34(+) cells identifies a population of functional endothelial precursors.
674 Blood. 2000;95(3):952-8.

675 51. Munier AI, Doucet D, Perrodou E, Zachary D, Meister M, Hoffmann JA, et al. PVF2, a PDGF/VEGF-
676 like growth factor, induces hemocyte proliferation in *Drosophila* larvae. EMBO Rep. 2002;3(12):1195-
677 200.

678 52. Varma Shrivastav S, Bhardwaj A, Pathak KA, Shrivastav A. Insulin-like growth factor binding protein-
679 3 (IGFBP-3): unraveling the role in mediating IGF-independent effects within the cell. Front Cell Dev
680 Biol. 2020;8:286.

681 53. Castillo J, Brown MR, Strand MR. Blood feeding and insulin-like peptide 3 stimulate proliferation of
682 hemocytes in the mosquito *Aedes aegypti*. PLoS Pathog. 2011;7(10):e1002274.

683 54. Wang G, Li N, Zhang L, Zhang L, Zhang Z, Wang Y. IGFBP7 promotes hemocyte proliferation in
684 small abalone *Haliotis diversicolor*, proved by dsRNA and cap mRNA exposure. Gene.
685 2015;571(1):65-70.

686 55. Charoensapsri W, Sangsuriya P, Lertwimol T, Gangnonngiw W, Phiwsaiya K, Senapin S. Laminin
687 receptor protein is implicated in hemocyte homeostasis for the whiteleg shrimp *Penaeus (Litopenaeus)*
688 *vannamei*. Dev Comp Immunol. 2015;51(1):39-47.

689 56. Rosa RD, Barracco MA. Antimicrobial peptides in crustaceans. Isj-Invert Surviv J. 2010;7(2):262-84.

690 57. Bachère E, Gueguen Y, Gonzalez M, de Lorgeril J, Garnier J, Romestand B. Insights into the anti-
691 microbial defense of marine invertebrates: the penaeid shrimps and the oyster *Crassostrea gigas*.
692 Immunol Rev. 2004;198:149-68.

693 58. Chen KY, Hsu TC, Huang PY, Kang ST, Lo CF, Huang WP, et al. *Penaeus monodon* chitin-binding
694 protein (PmCBP) is involved in white spot syndrome virus (WSSV) infection. Fish Shellfish Immunol.
695 2009;27(3):460-5.

696 59. Elbahnaswy S, Koiwai K, Zaki VH, Shaheen AA, Kondo H, Hirono I. A novel viral responsive protein
697 (MjVRP) from *Marsupenaeus japonicus* haemocytes is involved in white spot syndrome virus
698 infection. Fish Shellfish Immunol. 2017;70:638-47.

699 60. Munoz M, Vandenbulcke F, Saulnier D, Bachère E. Expression and distribution of penaeidin
700 antimicrobial peptides are regulated by haemocyte reactions in microbial challenged shrimp. Eur J
701 Biochem. 2002;269(11):2678-89.

702 61. Zhang K, Koiwai K, Kondo H, Hirono I. White spot syndrome virus (WSSV) suppresses penaeidin
703 expression in *Marsupenaeus japonicus* hemocytes. Fish & Shellfish Immunol. 2018;78:233-7.

704 62. Li S, Wang Z, Li F, Yu K, Xiang J. A novel vascular endothelial growth factor receptor participates in
705 white spot syndrome virus infection in *Litopenaeus vannamei*. Front Immunol. 2017;8:1457.

706 63. Wang Z, Li S, Li F, Xie S, Xiang J. Identification and function analysis of a novel vascular endothelial
707 growth factor, LvVEGF3, in the Pacific whiteleg shrimp *Litopenaeus vannamei*. Dev Comp Immunol.
708 2016;63:111-20.

709 64. Wang Z, Li S, Yu Y, Yu K, Zhang X, Xiang J, et al. Identification and characterization of two novel
710 vascular endothelial growth factor genes in *Litopenaeus vannamei*. Fish Shellfish Immunol.
711 2019;84:259-68.

712 65. Rosa RD, Vergnes A, de Lorgeril J, Goncalves P, Perazzolo LM, Sauné L, et al. Functional divergence
713 in shrimp anti-lipopolysaccharide factors (ALFs): from recognition of cell wall components to
714 antimicrobial activity. PLoS One. 2013;8(7):e67937.

715 66. Söderhäll I. Recent advances in crayfish hematopoietic stem cell culture: a model for studies of
716 hemocyte differentiation and immunity. Cytotechnology. 2013;65(5):691-5.

717 67. Mai HN, Nguyen HT, Koiwai K, Kondo H, Hirono I. Characterization of a Kunitz-type protease
718 inhibitor (MjKuPI) reveals the involvement of MjKuPI positive hemocytes in the immune responses of
719 kuruma shrimp *Marsupenaeus japonicus*. Dev Comp Immunol. 2016;63:121-7.

720 68. Koiwai K, Alenton RRR, Shiomi R, Nozaki R, Kondo H, Hirono I. Two hemocyte sub-populations of
721 kuruma shrimp *Marsupenaeus japonicus*. Mol Immunol. 2017;85:1-8.

722 69. Koiwai K, Kondo H, Hirono I. Isolation and molecular characterization of hemocyte sub-populations in
723 kuruma shrimp *Marsupenaeus japonicus*. Fisheries Science. 2019;85(3):521-32.

724 70. Sequeira T, Tavares D, Arala-Chaves M. Evidence for circulating hemocyte proliferation in the shrimp
725 *Penaeus japonicus*. Dev Comp Immunol. 1996;20(2):97-104.

726 71. Rota-Stabelli O, Daley AC, Pisani D. Molecular timetrees reveal a Cambrian colonization of land and a
727 new scenario for ecdysozoan evolution. Curr Biol. 2013;23(5):392-8.

728 72. Thomas GWC, Dohmen E, Hughes DST, Murali SC, Poelchau M, Glastad K, et al. Gene content
729 evolution in the arthropods. Genome Biol. 2020;21(1):15.

730 73. Wolfe JM, Breinholt JW, Crandall KA, Lemmon AR, Lemmon EM, Timm LE, et al. A phylogenomic
731 framework, evolutionary timeline and genomic resources for comparative studies of decapod
732 crustaceans. Proc Biol Sci. 2019;286(1901):20190079.

733 74. Broseus L, Thomas A, Oldfield AJ, Severac D, Dubois E, Ritchie W. TALC: Transcript-level Aware
734 Long-read Correction. Bioinformatics. 2020.

735 75. Bushmanova E, Antipov D, Lapidus A, Prjibelski AD. rnaSPAdes: a *de novo* transcriptome assembler
736 and its application to RNA-Seq data. Gigascience. 2019;8(9):1-13.

737 76. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome
738 assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011;29(7):644-52.

739 77. Biočanin M, Bues J, Dainese R, Amstad E, Deplancke B. Simplified Drop-seq workflow with
740 minimized bead loss using a bead capture and processing microfluidic chip. Lab Chip.
741 2019;19(9):1610-20.

742 78. Picelli S, Bjorklund AK, Reinius B, Sagasser S, Winberg G, Sandberg R. Tn5 transposase and
743 tagmentation procedures for massively scaled sequencing projects. Genome Research.
744 2014;24(12):2033-40.

745 79. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357-9.

746 80. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across
747 different conditions, technologies, and species. Nat Biotechnol. 2018;36(5):411-20.

748 81. Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using
749 regularized negative binomial regression. Genome Biol. 2019;20(1):296.

750 82. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture
751 and applications. BMC Bioinformatics. 2009;10:421.

752 83. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. Journal of
753 Molecular Biology. 1990;215(3):403-10.

754 84. Cao J, Spielmann M, Qiu X, Huang X, Ibrahim DM, Hill AJ, et al. The single-cell transcriptional
755 landscape of mammalian organogenesis. Nature. 2019;566(7745):496-502.

756