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

Keiichiro Koiwai[1,2]*, Takashi Koyama[3,4], Soichiro Tsuda[5], Atsushi Toyoda[6], Kiyoshi Kikuchi[3], Hiroaki Suzuki[7], Ryuji Kawano[1]

[1]Department of Biotechnology and Life Science, Tokyo University of Agriculture and Technology, Koganei, Japan; [2]Laboratory of Genome Science, Tokyo University of Marine Science and Technology, Minato, Japan; [3]Fisheries Laboratory, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Hamamatsu, Japan; [4]Graduate School of Fisheries and Environmental Sciences, Nagasaki University, Nagasaki, Japan; [5]bitBiome Inc, Shinjuku, Japan; [6]Advanced Genomics Center, National Institute of Genetics, Mishima, Japan; [7]Department of Precision Mechanics, Faculty of Science and Engineering, Chuo University, Bunkyo, Japan

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 six types of hemocytes of *Marsupenaeus japonicus* based on their transcriptional profiles. We identified markers of each subpopulation and predicted the differentiation pathways involved in their maturation. We also predicted 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.

*For correspondence:
koiwai@kaiyodai.ac.jp

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 (*FAO, 2020*) (http://www.fao.org/fishery/statistics/en). However, crustaceans that lack an adaptive immune system (*Jiravanichpaisal et al., 2006*; *Tassanakajon et al., 2013*; *Flegel, 2019*) 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 (*Zhang et al., 2019*).

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 (*Bauchau, 1981*; *Söderhäll and Smith, 1983*; *Johansson et al., 2000*). 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 (*van de Braak et al., 2002*), another
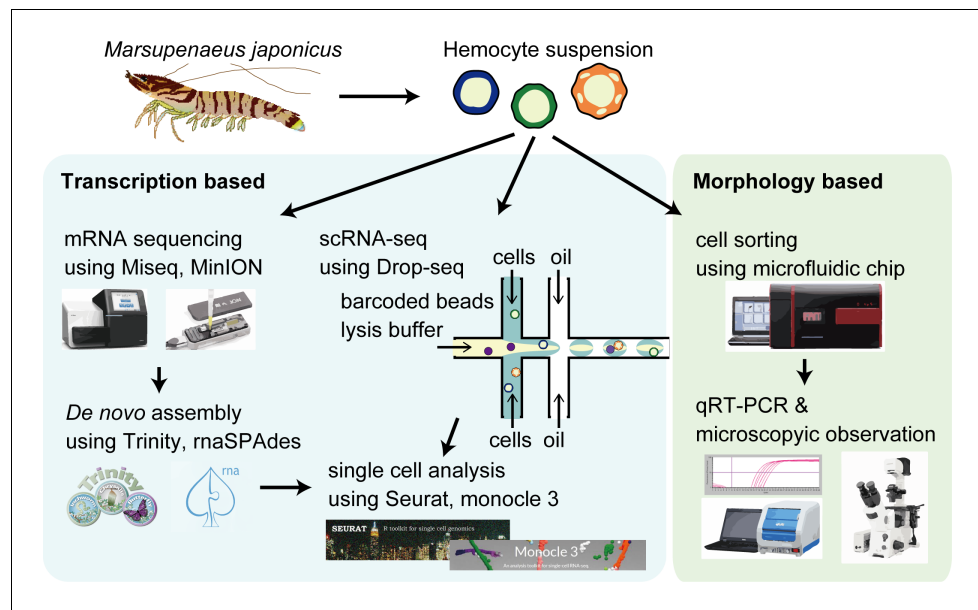
Figure 1. The schematic of single-cell mRNA sequencing (scRNA-seq) analysis of penaeid shrimp *M. japonicus* hemocytes. Single hemocytes were analyzed through the microfluidics-based Drop-seq, mRNA sequencing for the preparation of de novo assembled gene sets, in silico analysis workflow, and morphology-based cell classification.

dyeing method (*Kondo et al., 2012*), and iodixanol density gradient centrifugation (*Dantas-Lima et al., 2013*), 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 (*Rodriguez et al., 1995*; *van de Braak et al., 2000*; *Sung et al., 1999*; *Sung and Sun, 2002*; *Winotaphan et al., 2005*; *Lin et al., 2007*; *Xing et al., 2017*), their number is lower than that of humans, and their correspondence to the cell type, as well as their differentiation stage, is 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 (*Macosko et al., 2015*), has gained popularity for classifying cells and identifying new cell types. The enormous amount of biological data obtained from scRNA-seq leads us to classify cells into specific groups, analyze their heterogeneity, predict the functions of single-cell populations based on the gene expression profiles, and determine the cell proliferation or development pathways based on the pseudo-time ordering of a single cell (*Trapnell et al., 2014*; *Soneson and Robinson, 2018*). More recently, hemocytes of invertebrate, fly, and mosquito models have been subjected to these types of microfluidic-based scRNA-seq to reveal their functions (*Raddi et al., 2020*; *Tattikota et al., 2020*; *Cho et al., 2020*; *Cattenoz et al., 2020*; *Fu et al., 2020*).

Here, we performed scRNA-seq analysis on *Marsupenaeus japonicus* hemocytes to classify the hemocyte types and characterize their functions using the custom-built Drop-seq platform. To perform scRNA-seq, a high-quality gene reference is essential; however, such reference genomes are scarce for crustaceans because of the extremely high proportion of simple sequence repeats (*Zhang et al., 2019*). We circumvented this problem by preparing reference genes using hybrid de novo assembly of short- and long-read RNA sequencing results. The sequences obtained from the scRNA-seq were mapped onto the reference genes successfully. Our scRNA-seq uncovered the transcriptional profiles of a few thousand *M. japonicus* hemocytes. We identified the markers of each population and the differentiation pathways associated with their maturation. We also discovered the cell growth factors that might play crucial roles in hemocyte differentiation. Different immune
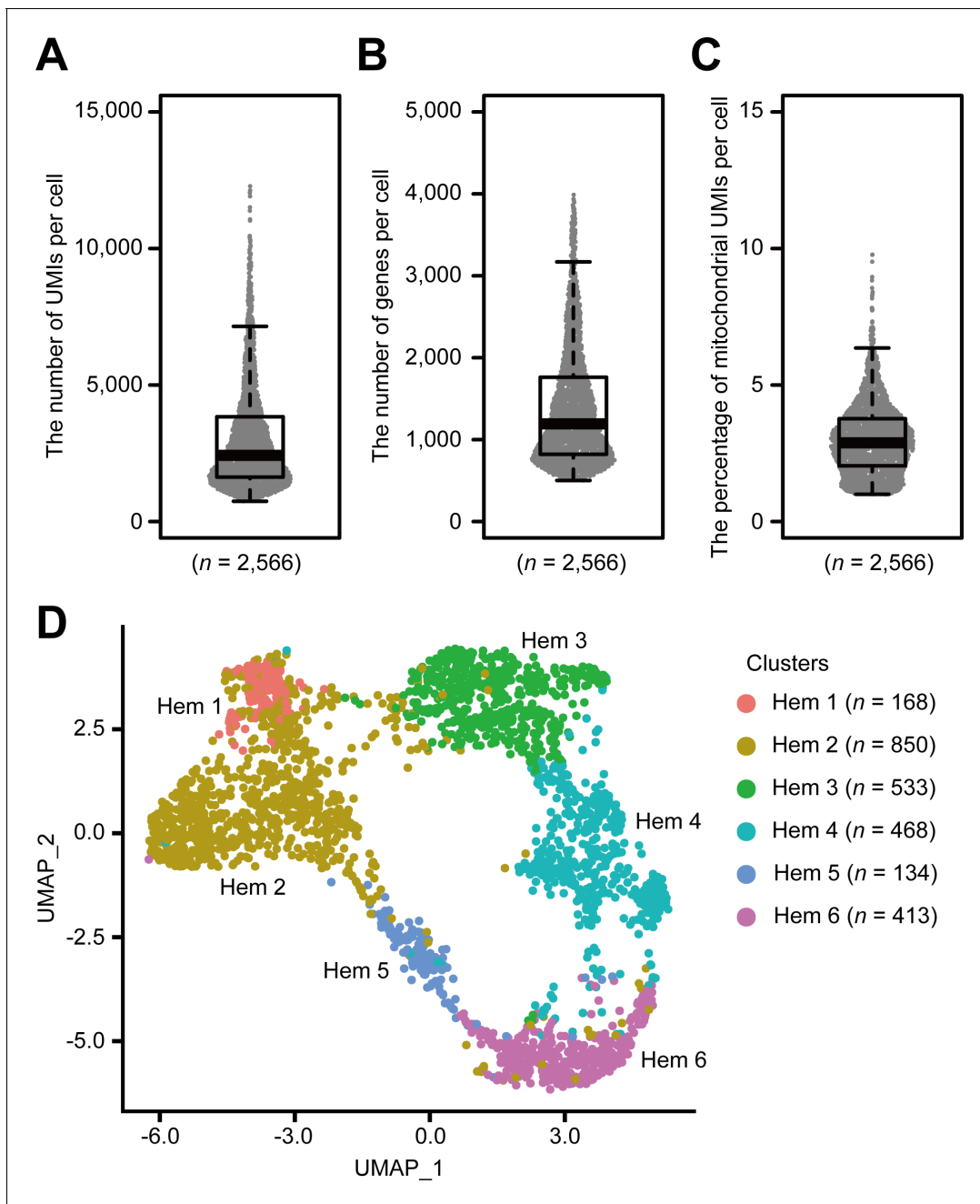
Figure 2. Single-cell mRNA sequencing (scRNA-seq) analysis of penaeid shrimp *M. japonicus* hemocytes. Distribution and the median of the number of transcripts (unique molecular identifiers [UMIs]) (A), genes (B), and percentage of mitochondrial UMIs (C) detected per cell. Uniform manifold approximation and projection (UMAP) plot of SCTransform batch corrected and integrated of hemocytes from three shrimps (*n* = 2566) (D). scRNA-seq analysis of penaeid shrimp *M. japonicus* hemocytes.

The online version of this article includes the following source data and figure supplement(s) for figure 2:

Source data 1. Excel sheets pertaining to UMIs, genes , mitochondrial UMIs detected per hemocyte used for *Figure 2A-C*.

Figure supplement 1. Sinaplots show the distribution of the number of transcripts (scored by unique molecular identifiers [UMIs]) (A), genes (B), and percentage of mitochondrial UMIs (C) detected per hemocyte on individual shrimp.

Figure supplement 1—source data 1. Excel sheets pertaining to UMIs, genes , mitochondrial UMIs detected per hemocyte on individual shrimp used for *Figure 2—figure supplement 1A-C*.

roles among these subpopulations were also suggested from the analysis of differentially expressed

immune-related genes. Our results present a unified classification of shrimp hemocytes and a deeper understanding of the immune system of shrimp.

Results

scRNA-seq clustering of *M. japonicus* hemocytes

Our study utilized scRNA-seq to determine the cellular subtypes with a distinct transcriptional expression (*Figure 1A*). To map the scRNA-seq sequences from *M. japonicus* hemocytes, we first prepared a high-quality de novo assembled reference genes using hybrid assembly of short- and long-read RNA sequencing results. The quality of assembled genes was quality-checked by BUSCO against arthropoda database. BUSCO tool showed 91.5% (35.6% complete and single-copy genes and 55.9% complete and duplicated genes) as complete genes, showing good completeness of the assembly. Some of the assembled genes constructed in this study showed homology to the same genes despite processed EvidentialGene program to remove similar sequences. Cluster database at high identity with tolerance (CD-HIT) programs was also applied on our draft assembled genes, but it was not able to cluster better than EvidentialGene program. Since we thought that subjective clustering of specific genes by ourselves would result in bias, we conducted our analysis based on the results of clustering by EvidentialGene program. Then, by using self-built Drop-seq microfluidic chips, single hemocytes were captured and their mRNA was barcoded using the droplet-based strategy. This process was performed in triplicates for three shrimp individuals. Following library preparation and sequencing, the transcriptomes obtained from scRNA-seq were mapped against the reference genes to discover the cell types.

Using the Drop-seq procedure, we profiled a total of 2566 cells and obtained a median value of 2427 unique molecular identifiers (UMIs), 1193 genes, and 2.9% of mitochondrial genes per cell across three replicates (*Figure 2*, *Figure 2—figure supplement 1A–C*). There was some transcriptional variability, which may have resulted from the artifacts of the Drop-seq system, because it is consistent with the original Drop-seq paper (*Macosko et al., 2015*) and with recent findings in other organisms, such as fish (*Carmona et al., 2017*), flies (*Tattikota et al., 2020*; *Cho et al., 2020*), and mosquitoes (*Raddi et al., 2020*). We constructed three Drop-seq libraries, which accumulates 1500 cells each from three individual shrimps. The sequence depth of read number per single-cell counts about 800,000 reads per cell based on the calculation that 400 million reads from NextSeq 500/550 High Output v2 kit/three individual shrimp/1500 cells. Conventionally, this read number is sufficient for any biological clustering with reference to the additional information of Drop-seq original paper; however, we obtained a median value of 2427 UMIs per cell, which is lower than other organisms such as *Drosophila melanogaster* counting 8000 UMIs (*Tattikota et al., 2020*) or HEK (human) and 3T3 (mouse) cells counting 9000 UMIs (*Macosko et al., 2015*), respectively. This is due to the lack of genomic information of *M. japonicus*; hence, we prepared de novo assembled transcripts for the read mapping. As the genomic analysis of *M. japonicus* further progresses, we believe that these UMI values will be improved.

Applying the SCTransform batch correction method integrated into the Seurat package allowed us to remove the individual differences. SCTransform successfully integrated all three shrimp datasets (*Figure 2—figure supplement 1D*), among which we identified a total of six clusters (*Figure 2D*). Each cluster contained the following number of cells: Hem 1, 168 cells (6.5%); Hem 2, 850 cells (33.1%); Hem 3, 533 cells (20.8%); Hem 4, 468 cells (18.2%); Hem 5, 134 cells (5.2%); and Hem 6, 413 cells (16.1%).

Since not all the de novo assembled transcripts were expressed in this single-cell data, the highly expressed transcript that represents 50% of the total normalized expression data was extracted as 'Ex50' for further analysis. The functions of 1362 transcripts composed Ex50 were predicted by blastx program on penaeid shrimp proteins and by eggNOG-mapper annotation for the eukaryotic orthologous groups (KOGs) and gene ontologies (GOs) (*Supplementary file 1*).

KOGs and GOs analysis on clusters

Among 1362 transcripts composed Ex50, 865 (70.9%) transcripts on shrimp proteins, 886 (65.1%) transcripts on KOG annotation, and 614 (45.1%) transcripts on GO annotation were annotated.

In the result of KOGs, seven KOG annotations (i.e., chromatin structure and dynamics, amino acid transport and metabolism, lipid transport and metabolism, transcription, replication, recombination and repair, signal transduction mechanisms, and intracellular trafficking, secretion, and vesicular transport) were highly expressed among most of the cells in cluster Hem 1 (*Figure 3A*). In the result of GOs, two GO annotations (i.e., molecular transducer activity and growth) were highly expressed in cluster Hem 1. Furthermore, one GO annotation, immune system process, was highly expressed in cluster Hem 6 (*Figure 3B*). This result showed that a group of genes involved in cell division and metabolism are highly expressed in Hem 1. It was also predicted that immune-related genes were strongly expressed in Hem 6. Genes involved in basic metabolism, such as lipid metabolism and transcription, were expressed in almost all clusters, but among those the expression tended to be stronger in Hem 1.

Cluster-specific markers and their functional prediction

A total of 415 cluster-specific markers were predicted using the Seurat FindMarkers tool (*Figure 4A*; *Supplementary file 2*). For each cluster, 167 (Hem 1), 29 (Hem 2), 38 (Hem 3), 37 (Hem 4), 42 (Hem 5), and 102 (Hem 6) markers were identified. Important markers in each cluster are described in the dot plot figure (*Figure 4B*).

Many of Hem 1-specific markers were related to cell proliferation, cell migration, and colony formation. For example, *sin3 histone deacetylase corepressor complex component SDS3* (*SUDS3*) (Mj-9135) and *histone deacetylase* (*HDAC*) (Mj-18057) are known to be essential for the proliferation through controlling cell cycle progression, DNA replication and repair, and cell death in mammal studies (*Dannenberg et al., 2005*; *McDonel et al., 2012*). *Polypyrimidine tract-binding protein 1* (*PTBP1*) (Mj-18644) is also a gene related to cell proliferation, cell migration, and colony formation in human tumor studies (*Cheung et al., 2009*; *Shibayama et al., 2009*; *Monzón-Casanova et al., 2020*). It is a multifunctional RNA-binding protein that is overexpressed in glioma, a type of tumor that occurs in the brain, and a decreased expression of PTBP inhibits cell migration and increases the adhesion of cells to fibronectin and vitronectin (*Cheung et al., 2009*; *Shibayama et al., 2009*). PTBP has been shown to be involved in germ cell differentiation in *D. melanogaster* and is essential for the development of *Xenopus laevis* (*Jiao et al., 2016*). *Cell division cycle 7-related protein kinase* (*CDC7*) (Mj-22708 and Mj-22709) is required for continuous DNA replication in mammalian cells, and reduced levels of Cdc7 kinase are viable but exhibit significantly reduced body size and impaired cell proliferation (*Kim et al., 2003*). The high expression of homologues of cell proliferation-related genes suggests that Hem 1 is a cluster related to cell proliferation in shrimp hemocytes.

Among the Hem 2 markers, three markers were annotated with *hemocyte transglutaminase* (*HemTGase*) (Mj-16280, Mj-16290, and Mj-21219), two markers were annotated with *phenoloxidase-activating factor 2-like* (*PPAF2-like*) (Mj-15851 and Mj-23817), and two markers were annotated with *single VWC domain protein 5* (*Vago 5*) (Mj-24191 and Mj-24193). *TGase* is an immature hemocyte marker of crayfish and shrimp. When the extracellular TGase is digested, hemocytes start to differentiate into mature hemocytes (*Lin et al., 2008*; *Huang et al., 2004*; *Lin et al., 2011*). The high expression of *HemTGase* suggests that Hem 2 is in the early stage of hemocytes. PPAF2 has a role to activate phenoloxidase (PO) cascade involving melanization (*Boonchuen et al., 2021*). Vago 5 is a kind of shrimp cytokine functionally similar to interferon and found to be involved in shrimp antiviral immunity (*Li et al., 2015*; *Gao et al., 2019*). PPAF2 and Vago 5 do not directly exhibit defense mechanisms, but work by activating other genes and other immune system. Therefore, Hem 2 may work as an immune activator against other hemocytes.

In cluster Hem 3, two markers were annotated with *anti-lipopolysaccharide factor-A1* (*ALF-A1*) (Mj-19322 and Mj-23779), and one marker was annotated with *viral responsive protein* (*VRP*) (Mj-4390). In cluster Hem 4, one marker was annotated with *anti-lipopolysaccharide factor-like* (*ALF-like*) (Mj-7049), one marker was annotated with *myeloid differentiation factor 2–related lipid-recognition protein* (*ML protein*) (Mj-10040), and three markers were annotated with *penaeidin-II* (Mj-18245, Mj-19281, and Mj-20968). ALF and penaeidin-II are well-studied antimicrobial peptides (AMPs) (*Destoumieux et al., 1997*; *Bachère et al., 2004*; *An et al., 2016*; *Liu et al., 2006*; *Mekata et al., 2010*; *Rosa and Barracco, 2010*; *Rosa et al., 2013*). These AMPs are stored in the granules of hemocytes. VRP protein is distributed in granular-containing hemocytes and induced its expression by virus infection (*Elbahnaswy et al., 2017*). ML protein recognizes a lipid component of virus and induces the expression of Vago 5 (*Gao et al., 2019*), which is highly expressed in Hem 2. The
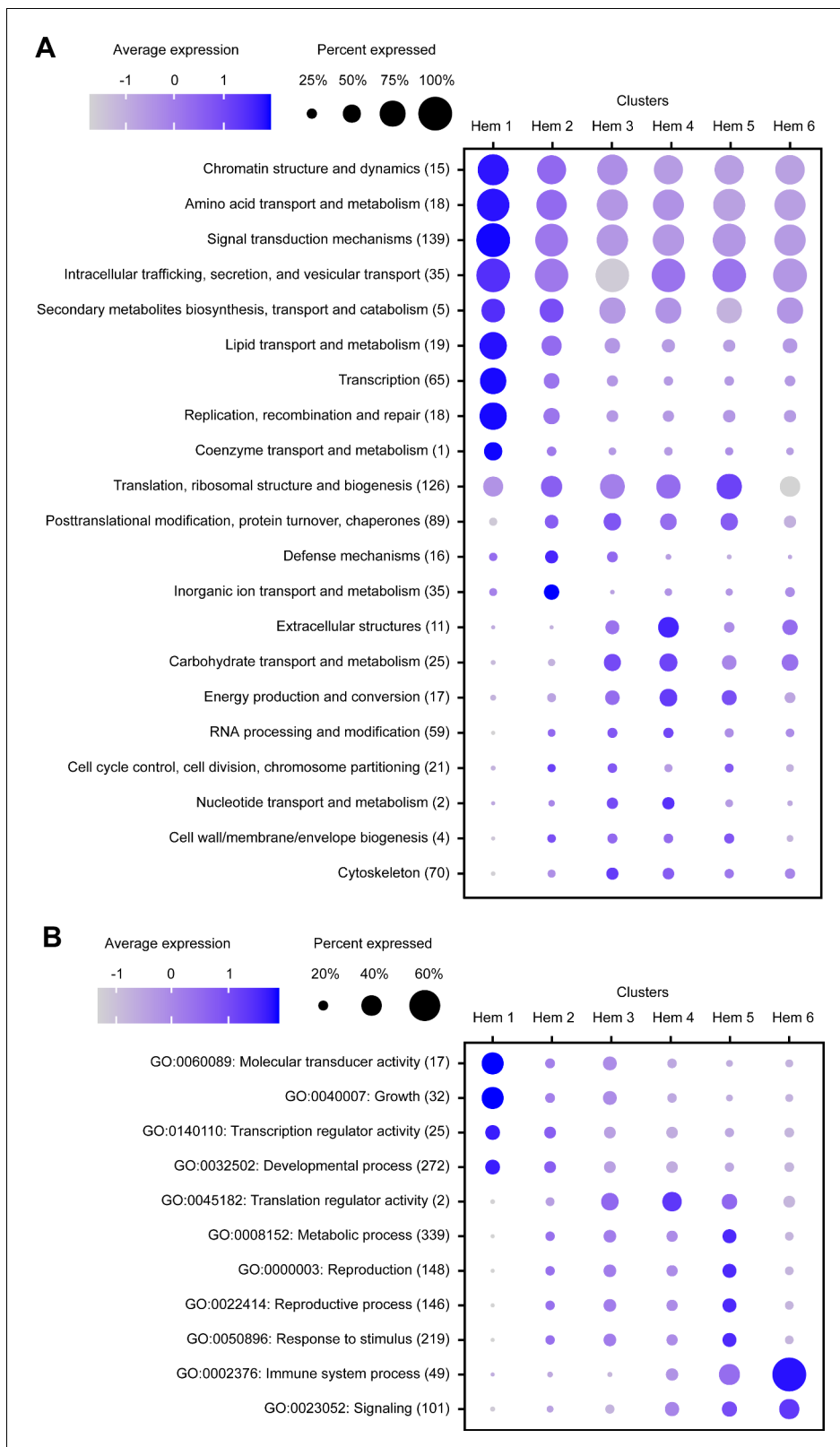
Figure 3. Dot plot profiling of the eukaryotic orthologous group (KOG) and gene ontology (GO) analyses in each cluster. Dot plot representing the average expression of KOGs (**A**) and GOs (**B**) per cluster. Color gradient of dots represents the expression level, while the size represents the percentage of cells expressing any genes per cluster. The numbers in parentheses represent the number of genes estimated as distinct function of KOGs or GOs.
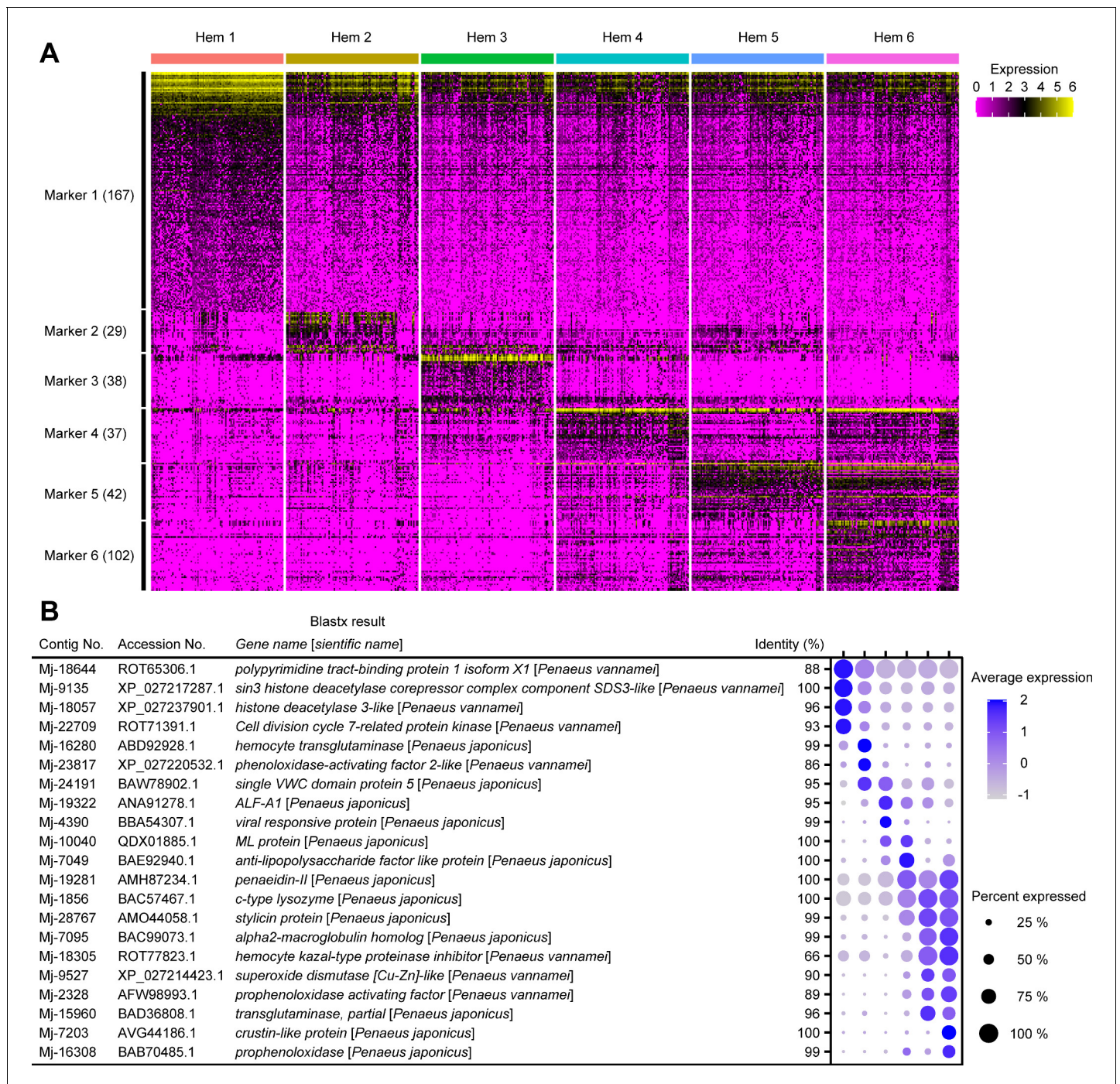
Figure 4. The cluster-specific marker genes predicted using the Seurat FindMarkers tool. Heat map profile of the marker in each cluster (A). Color gradient represents the expression level of each single cell. The numbers in parentheses represent the number of genes estimated as markers of each cluster. Important marker genes in each cluster (B). Color gradient of the dot represents the expression level, while the size represents the percentage of cells expressing any genes per cluster. Detailed blast results of each marker on penaeid shrimp are listed in *Supplementary file 2*.

expression of AMPs and VRP supporting Hem 3 and 4 are granular-containing hemocytes, and the expression of virus infection-related genes, VRP and ML protein, suggests that Hem 4 might play important roles when viruses infect shrimp.

In clusters Hem 5 and Hem 6, many of the cluster-specific markers showed similarity with immune-related genes of penaeid shrimp. In clusters Hem 5 and Hem 6, markers showed a high

similarity with *alpha2-macroglobulin*, *c-type lysozyme*, *stylicin*, *crustin*, *penaeidin-II*, *hemocyte kazal-type proteinase inhibitor* (*KPI*), *superoxide dismutase* (*SOD*), *prophenoloxidase* (*proPO*), *PPAF*, and *TGase*. Hem 5 and 6 were found to have high expressions of immune-related genes by GO analysis, and our maker analysis results support this observation.

G2/M and S phase clustering and pseudo-temporal ordering of hemocytes delineate hemocyte lineages

Cell cycle of each cell was calculated by Seurat CellCycleScoring function based on the homologue genes of *Drosophila* G2/M and S phase markers (https://github.com/hbc/tinyatlas, *Kirchner and Barrera, 2019*). In the search for cell cycle-specific markers, we found that a large portion of cells in Hem 1 expressed the G2/M-related genes of *Drosophila* (*Supplementary file 3*; *Figure 5—figure supplement 2*). For examples, HP1 and Su(var)205 are known to be essential for the maintenance of the active transcription of the euchromatic genes functionally involved in cell-cycle progression, including those required for DNA replication and mitosis in *Drosophila* (*De Lucia et al., 2005*; *Paro and Hogness, 1991*). CTCF has zinc finger domains and plays an important role in the development and cell division of fly and mammalian cells (*Mohan et al., 2007*; *Rasko et al., 2001*). These findings suggested that hemocytes grouped as Hem 1 are tightly regulated by these G2/M phase-related genes to promote cell division. Among the total of 2566 cells, G1 phase consisted of 66% (1693 cells), G2/M phase consisted of 9% (236 cells), and S phase consisted of 25% cells (637 cells), respectively. With the exception of Hem 1, the percentage of G1 phase was the highest, and S phase was the second highest among all clusters; Hem 2 to Hem 6 (*Figure 5A, B*). On the other hand, Hem 1 showed the highest percentage of G2/M phase, followed by G1 phase and the lowest percentage of S phase. KOG and GO analyses of each cluster showed that Hem 1 cluster has high expression of chromatin dynamics, duplication, and growth functions (*Figure 3*). Thereby, we determined that Hem 1 is a cluster in the early stage of hemocyte development, which might have the potential of self-renewing or just came out from the hematopoietic tissue (HPT). A previous study using flow cytometry reported a significantly lower percentage of S phase, about 0.5% (*Sequeira et al., 1996*). It is possible that the genes used in the analysis of S phase had lower homology with *Drosophila* markers than the gene set used in the analysis of G2/M phase, and thus did not score as accurately as *Drosophila*.

We further performed lineage-tree reconstruction using the Monocle 3 learn_graph function to investigate the dynamics of hemocyte differentiation because the differentiation and proliferation pathways of hemocytes in shrimp and other crustaceans are still under debate (*Söderhäll, 2016*). We considered Hem 1 to be the initial state of hemocytes and set it as the starting point in the differentiation process because Hem 1 expressed cell proliferation-related genes, *TGase* (an immature hemocyte marker), and G2/M phase-related genes (*De Lucia et al., 2005*; *Paro and Hogness, 1991*; *Mohan et al., 2007*; *Rasko et al., 2001*; *Scott et al., 2007*; *Junkunlo et al., 2017*; *Junkunlo et al., 2019*). From this pseudo-temporal ordering analysis, we found two main lineages starting from Hem 1 to Hem 4, and Hem 6 at the endpoints (*Figure 5C*). In crayfish, hematopoietic stem cells are present in HPT, and two types of hemocyte lineages starting from a hematopoietic stem cell exist (*Lin and Söderhäll, 2011*). In *Penaeus monodon*, hyaline cells (i.e., agranulocytes) are considered as the young and immature hemocytes of two types of matured hemocytes (*van de Braak et al., 2002*). Our pseudo-temporal ordering analysis revealed that the hemocytes of *M. japonicus* differentiate from a single subpopulation into two major populations. The differentiation process was continuous, not discrete, which was in agreement with previous arguments on the crustacean hematopoiesis mechanism (*van de Braak et al., 2002*; *Lin and Söderhäll, 2011*; *Noonin et al., 2012*).

Possibility of application of *Drosophila* hemocyte-type marker to shrimp

Since the hemocyte type-specific markers are better studied in *Drosophila*, we checked the Ex50 that are similar to the markers of *Drosophila* to determine whether they are applicable to shrimp. Among the specific markers of four types of *Drosophila* hemocytes, 8 genes in prohemocytes, 14 genes in plasmatocytes, 35 genes in lamellocytes, and 5 genes in crystal cells showed similarity with the shrimp genes (*Figure 6*; *Supplementary file 4*).
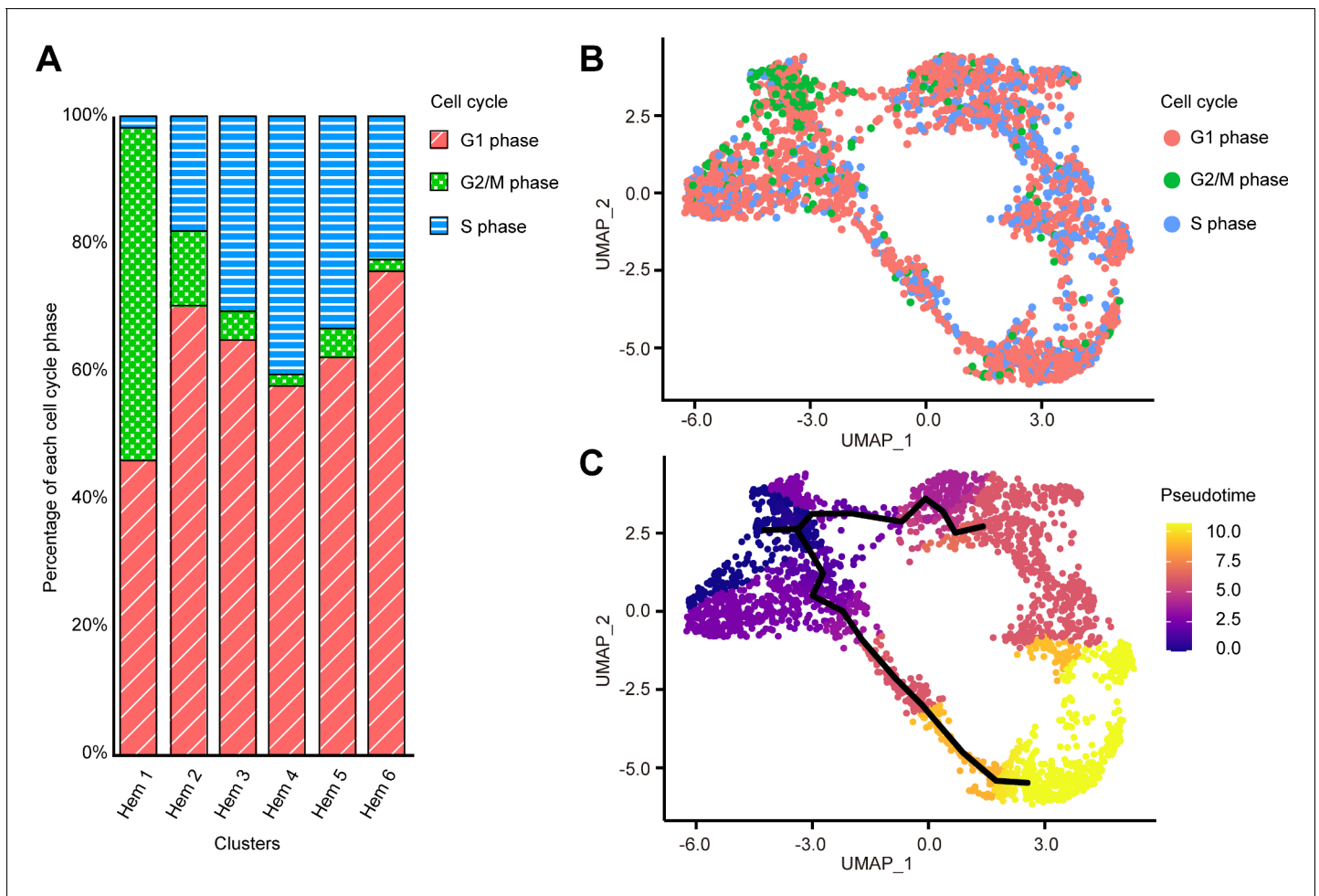
Figure 5. Cell cycle distribution of each cluster and pseudo-temporal ordering of hemocyte lineages. Percentage of each cell cycle on clusters (**A**). Uniform manifold approximation and projection plot of cell cycles of hemocytes from three shrimps (*n* = 2566) (**B**). Visualization of clusters onto the pseudotime map using monocle 3 (**C**). The black lines indicate the main path of the pseudotime ordering of the cells. Color gradient of each dot represents the pseudotime.

The online version of this article includes the following source data and figure supplement(s) for figure 5:

Source data 1. Source data of the percentage of each cell state per cluster used for *Figure 5A*.
Figure supplement 1. Uniform manifold approximation and projection plot of cell cycles from individual shrimps.
Figure supplement 2. Dot plots profiling of cell cycle-related genes in each cluster.

In case of prohemocyte markers, Mj-7095 and Mj-21032 showed cluster-specific expression and were annotated on *thioester-containing protein 4* (*Tep4*) with identity 30%, which participates in the cellular immune response to certain Gram-negative bacteria in *Drosophila* (*Figure 6*). Prohemocytes are progenitor cells in *Drosophila*; however, the expression of these marker genes was not strong in Hem 1, which was expected to be progenitor cells in shrimp, but rather was high in Hem 5 and 6. In the blastx results on penaeid shrimp proteins, Mj-7095 and Mj-21032 were annotated on *alpha2-macroglobulin* gene (*Supplementary file 1*). Alpha2-macroglobulin (a2m) of shrimp contributes clotting pathway to eliminate bacterial infection, and its proteins are contained within the secretory granules of hemocytes (*Chaikeeratisak et al., 2012*). Therefore, it is not reasonable to use these markers as a common marker with prohemocytes in shrimp.

Drosophila plasmatocytes are known as small round cell and professional phagocytes reminiscent of the cells from the mammalian monocyte/macrophage lineage (*Meister and Lagueux, 2003*). Mj-6 and Mj-14846 showed cluster-specific expression on the plasmatocyte markers (*Figure 6*). Mj-6 was annotated with *hemolection* (*Hml*) of *Drosophila* and *hemocytin-like* of *Litopenaeus vannamei*. Mj-
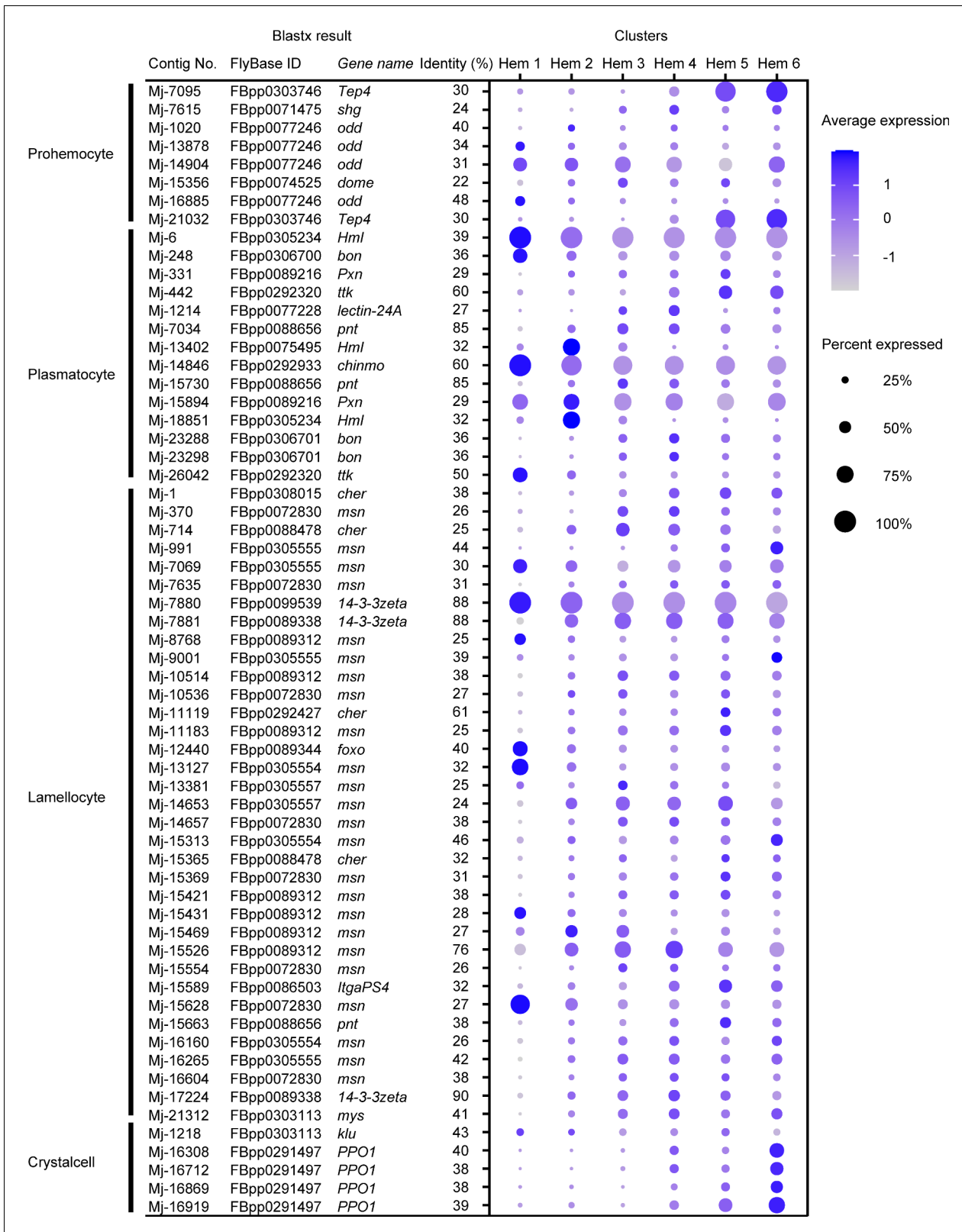
Figure 6. Dot plots profiling of *Drosophila* hemocyte-type markers in each cluster. Color gradient of the dot represents the expression level, while the size represents the percentage of cells expressing any gene per cluster. Detailed blast results are listed in *Supplementary file 4*.

14846 was annotated with *chronologically inappropriate morphogenesis* (*chinmo*) of *Drosophila* and *genetic suppressor element 1-like* of *L. vannamei*. Hml of *Drosophila* is involved in the clotting reaction. It has been estimated that some of the hemocytes expressing *Hml* are self-renewing in *Drosophila* single-cell study (*Tattikota et al., 2020*). *chinmo* plays a role in the proliferation of developing hemocytes in *Drosophila* (*Flaherty et al., 2010*). Since our results also predict that Hem 1 is a group of progenitor or self-renewing cells, these genes could be a potentially useful marker for identifying progenitor or self-renewing cells of shrimp.

Mj-7880 showed characteristic expression within the identical cluster among lamellocyte markers (*Figure 6*). Mj-7880 was annotated with *14-3-3zeta* of *Drosophila* and *14-3-3-like protein* of *L. vannamei*. There are various isoforms of *14-3-3* genes, and *14-3-3zeta* is a marker for lamellocytes, but it is also known that 14-3-3 proteins function in normal cell cycle progression in *Drosophila* (*Su et al., 2001*). In shrimp, Mj-7880 is probably working as cell cycle-specific roles, which is why Mj-7880 is strongly expressed in Hem 1.
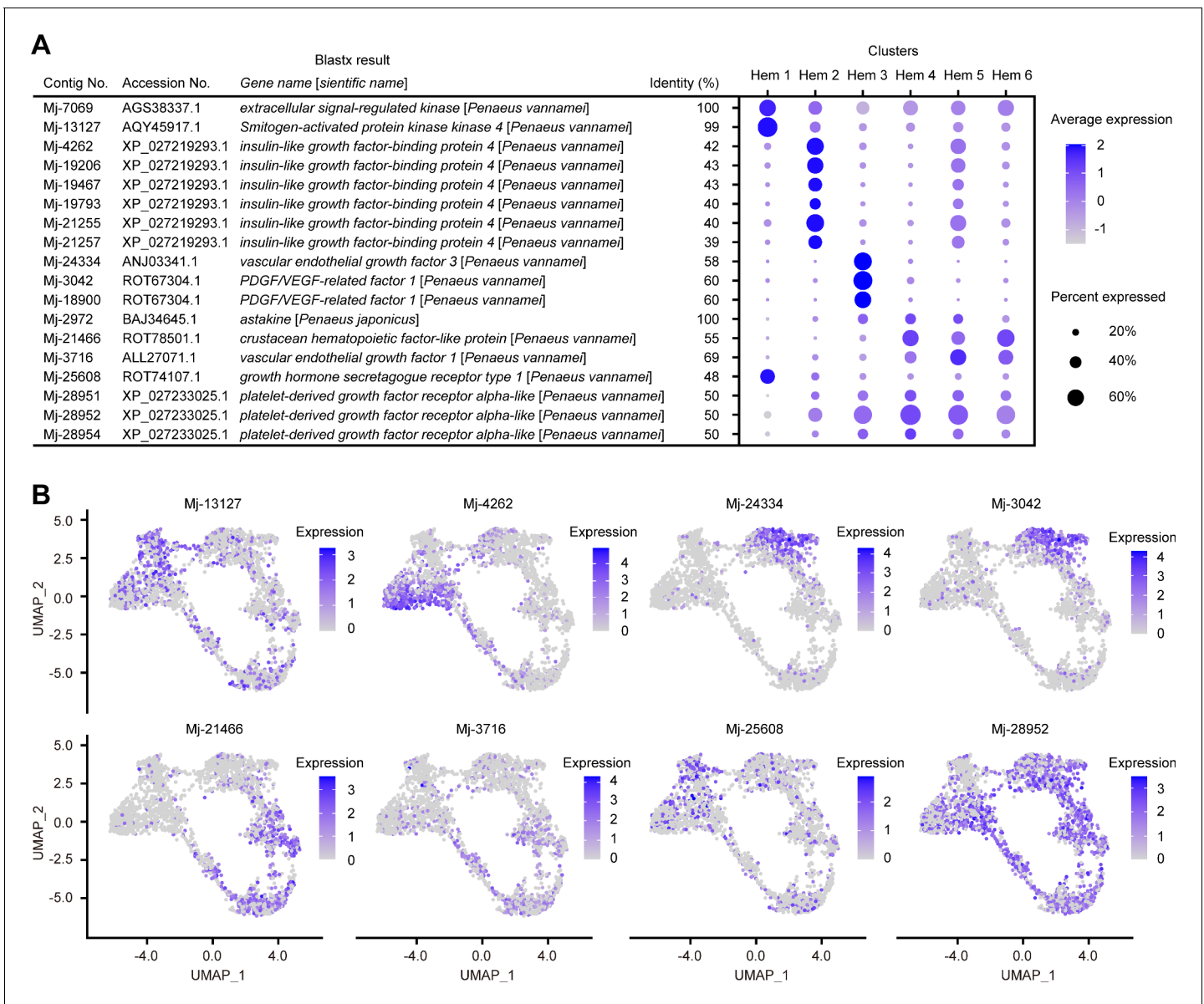
Among *Drosophila* crystal cell markers, four contigs, Mj-16308, Mj-16712, Mj-16869, and Mj-16919, were annotated on *prophenoloxidase 1* (*PPO1*) of *Drosophila* (*Figure 6*). These four contigs were also annotated on *proPO* of *M. japonicus* and *phenoloxidase 3-like* of *L. vannamei*. From the single-cell study of *Drosophila* hemocytes, the expression level of *PPO1* increased with the maturation of crystal cells (*Tattikota et al., 2020*). Crystal cells contain about −5% of total number of hemocytes of *Drosophila* and contain the enzymes necessary for humoral melanization that accompanies a number of immune reactions (*Meister and Lagueux, 2003*). It is very interesting to note that the ratio of Hem 6 is larger in shrimp compared to in *Drosophila* crystal cells, but the expression of *proPO* is highest in Hem 6, the estimated endpoint of differentiation by pseudotime analysis (*Figure 5C*). It is not possible to assign crystal cells of *Drosophila* and Hem 6 of shrimp as homologous cells because of the unknown similarities in morphology and other functions, but at least *PPO/proPO* can be considered to be a marker gene commonly expressed in mature hemocytes of both *Drosophila* and shrimp.

Expression of cell growth-related genes

The exploration of cluster-specific markers revealed that cell growth-related genes were specifically expressed in certain clusters. We identified 10 cell growth-related genes from Ex50 that are predicted to be involved in cell growth and differentiation: *extracellular signal-regulated kinase* (*ERK*), *mitogen-activated protein kinase kinase 4* (*MKK4*), *insulin-like growth factor-binding protein 4* (*IGFBP-4*), *vascular endothelial growth factor 1* (*VEGF-1*), *VEGF-3*, *astakine*, *crustacean hematopoietic factor-like protein* (*CHF-like*), *PDGF/VEGF-related factor 1*, *growth hormone secretagogue receptor type 1* (*GHSR*), and *platelet-derived growth factor receptor alpha-like* (*PDGFRA-like*) (*Figure 7A*; *Supplementary file 1*).

Both *ERK* and *MKK4* are a series of genes involved in mitogen-activated protein kinase (MAPK) pathways. In *Drosophila,* MAPK pathway is activated by signaling from several receptors, causing hemocyte proliferation and differentiation (*Zettervall et al., 2004*). Although it has been investigated that MAPK pathway contributes to the defense responses of shrimp (*Yang et al., 2012*; *He et al., 2018*; *Luo et al., 2020*), it is not yet known whether this pathway directly affects the proliferation or differentiation of hemocytes. However, the high expression of *ERK* and *MKK4* in Hem 1 strongly suggests that MAPK pathway is involved in the proliferation of hemocytes in shrimp. Therefore, not only immune responses but also cell-type variation is a good point of focus for future research on studies of MAPK pathway in shrimp.

Hem 2 highly expressed *IGFBP-4* (*Figure 7*). IGFBP delivers IGFs to the target cells in mammal studies and is essential for cell growth or differentiation (*Varma Shrivastav et al., 2020*). The high expression of a receptor of the insulin-like peptide at mature hemocytes in the mosquito suggests that the insulin signaling pathway regulates hemocyte proliferation (*Castillo et al., 2011*). The silencing and overexpression of IGFBP caused a decrease and increase in the growth of hemocytes of abalone *Haliotis diversicolor* (*Wang et al., 2015*). These studies indicate that the IGFBP-related insulin signaling pathway is important for hemocyte proliferation and differentiation in invertebrates. IGFBP-4 might play an essential role in the differentiation of hemocytes from Hem 1 to Hem 2 in shrimp. The high expression of *IGFBP* is determined in the brain and gonads of *L. vannamei* (*Ventura-López et al., 2017*). This fact also suggests that IGFBP plays a possible role in organ growth and maturation in shrimp. It is worth noting that the expression of *IGFBP-4* is strong at the site

A

Contig No.	Accession No.	Gene name [*sientific name*]	Identity (%)
Mj-7069	AGS38337.1	*extracellular signal-regulated kinase* [*Penaeus vannamei*]	100
Mj-13127	AQY45917.1	*Smitogen-activated protein kinase 4* [*Penaeus vannamei*]	99
Mj-4262	XP_027219293.1	*insulin-like growth factor-binding protein 4* [*Penaeus vannamei*]	42
Mj-19206	XP_027219293.1	*insulin-like growth factor-binding protein 4* [*Penaeus vannamei*]	43
Mj-19467	XP_027219293.1	*insulin-like growth factor-binding protein 4* [*Penaeus vannamei*]	43
Mj-19793	XP_027219293.1	*insulin-like growth factor-binding protein 4* [*Penaeus vannamei*]	40
Mj-21255	XP_027219293.1	*insulin-like growth factor-binding protein 4* [*Penaeus vannamei*]	40
Mj-21257	XP_027219293.1	*insulin-like growth factor-binding protein 4* [*Penaeus vannamei*]	39
Mj-24334	ANJ03341.1	*vascular endothelial growth factor 3* [*Penaeus vannamei*]	58
Mj-3042	ROT67304.1	*PDGF/VEGF-related factor 1* [*Penaeus vannamei*]	60
Mj-18900	ROT67304.1	*PDGF/VEGF-related factor 1* [*Penaeus vannamei*]	60
Mj-2972	BAJ34645.1	*astakine* [*Penaeus japonicus*]	100
Mj-21466	ROT78501.1	*crustacean hematopoietic factor-like protein* [*Penaeus vannamei*]	55
Mj-3716	ALL27071.1	*vascular endothelial growth factor 1* [*Penaeus vannamei*]	69
Mj-25608	ROT74107.1	*growth hormone secretagogue receptor type 1* [*Penaeus vannamei*]	48
Mj-28951	XP_027233025.1	*platelet-derived growth factor receptor alpha-like* [*Penaeus vannamei*]	50
Mj-28952	XP_027233025.1	*platelet-derived growth factor receptor alpha-like* [*Penaeus vannamei*]	50
Mj-28954	XP_027233025.1	*platelet-derived growth factor receptor alpha-like* [*Penaeus vannamei*]	50

B

Figure 7. Cell growth-related gene expressions on clusters and single hemocytes. Dot plots profiling of cell growth-related genes in each cluster (**A**). Color gradient of the dot represents the expression level, while the size represents the percentage of cells expressing any gene per cluster. Expression profiling of cell growth-related genes on uniform manifold approximation and projection plot (**B**). Color gradient of each dot represents the expression level. The details of the identified genes are listed in ***Supplementary file 1***.

where Hem 3 and Hem 5 seem to be in the early stage of differentiation, which may indicate that IGFBP-4 is acting to initiate differentiation into Hem 3 and Hem 5 from Hem 2 (***Figure 7B***).

Cells expressing *VEGF-3* and *PDGF/VEGF-related factor* 1 were dominant in Hem 3. The vascular endothelial growth factor (VEGF) signaling pathway is essential for vasculogenesis, cell proliferation, and tumor migration in mammals (***Neufeld et al., 1999***; ***Peichev et al., 2000***). Furthermore, in *Drosophila*, VEGF homologs control the number of circulating hemocytes (***Munier et al., 2002***). The high expression of *VEGF-3* and *PDGF/VEGF-related factor* 1 in Hem 3 indicated that Hem 2 would differentiate into Hem 3 the VEGF signaling pathway. Interestingly, a different homologue of VEGF, *VEGF 1*, was highly expressed in Hem 4, Hem 5, and Hem 6. These hemocytes are estimated as endpoints of differentiation by pseudotime analysis (***Figure 5C***). This means that the different types of VEGFs, *VEGF-1* and *VEGF-3*, may induce differentiation into different hemocytes.

Astakines, which were expressed in Hem 3, Hem 4, and Hem 5, are produced by granular-containing hemocytes and are released into the plasma (*Söderhäll et al., 2005*). Astakines function as decreasing the extracellular TGase activity and inducing the differentiation of hemocyte precursor into maturation (*Söderhäll, 2016*). Our analysis showed that *astakine* was expressed from a part of matured hemocytes, which is predicted as granular-containing hemocytes. Therefore, as in the previous study (*Lin et al., 2011*), it is thought that *astakine* act on precursor hemocytes, Hem 1 and Hem 2, to promote maturation in shrimp.

From Ex50 genes, we identified two candidate receptors related to cell growth, *GHSR* and *PDGFRA-like*. GHSR is a receptor for ghrelin and stimulates growth hormone secretion in vertebrates (*Smith et al., 2001*). Growth hormone-releasing peptide-6 (GHRP-6) is a synthetic peptide that mimics the effect of the endogenous ligand ghrelin and has a strong growth hormone secretagogues activity. In shrimp, GHRP-6 functions as a growth promoter peptide. More interestingly, the total hemocyte number in hemolymph was increased by supplying of GHRP-6 (*Martínez et al., 2017*). Therefore, GHSR is thought to play an important role in increasing hemocyte number in shrimp. PDGFRA, as the name suggests, is the receptor for platelet-derived growth factor (PDGF). It regulates cellular growth and differentiation by binding to its ligand (*Lei and Kazlauskas, 2009*). In crayfish, PDGF regulates hemocyte differentiation from immature to matured (*Junkunlo et al., 2017*). The strong expression of *PDGFRA* in shrimp was detected in Hem 2 to Hem 6, not in Hem 1, suggesting that PDGF signals are important for hemocyte maturation but not for proliferation of hemocytes in shrimp.

So far, we described the expression of proliferation- and differentiation-promoting genes. However, there was also the specific expression of the hemocyte homeostasis regulatory gene, *crustacean hematopoietic factor* (*CHF*) (*Lin and Söderhäll, 2011*). CHF is a hematopoietic factor of crayfish, and the silencing of CHF leads to an increase in the apoptosis of cells in HPT and a reduction in the number of circulating hemocytes (*Jiao et al., 2016*). Additionally, the silencing of laminin, a receptor of CHF, reduces the number of circulating hemocytes by decreasing the number of agranulocytes, as opposed to granulocytes, in *P. vannamei* (*Charoensapsri et al., 2015*). *CHF-like* was expressed in cluster Hem 4, Hem 5, and Hem 6 (*Figure 7*). Taken together, *CHF-like* expressed from matured hemocytes might work as a hematopoietic factor against immature hemocytes, such as Hem 1 and Hem 2.

Expression of immune-related genes

Hemocytes of shrimp play key roles in their immunity; therefore, we identified immune-related genes from Ex50, and then analyzed their expression levels to deduce the detailed immune functions of each cluster (*Supplementary file 1*).

AMPs and antimicrobial enzymes play the most important role in the immunity of shrimps and are well known to be stored in granulocytes (*Bachère et al., 2004*; *Rosa and Barracco, 2010*). The expression patterns of AMPs and antimicrobial enzymes revealed that the major AMPs of penaeid shrimp were expressed in clusters Hem 4 to Hem 6 (*Figure 8A, B*, *Figure 8—figure supplement 1*). Therefore, Hem 4 to Hem 6 were predicted to be granulocytes. Whereas almost all the expressions of AMP genes were increased as hemocytes grew (*Figure 5C*, *Figure 8B*), the expression patterns of *ALFs* were different from other AMPs (*Figure 8—figure supplement 1*). It is predicted that the diversity of shrimp *ALFs* contributes to create synergism-improving shrimp antimicrobial defenses (*Rosa et al., 2013*). Analysis of each function of *ALFs* may lead to a more detailed understanding of immune function in each cluster of hemocytes in shrimp.

Clotting is an important innate immune response in shrimp. Transglutaminase (TGase) plays critical roles in clotting and hematopoiesis, and there are two types of TGases in shrimp (*Lin et al., 2008*; *Söderhäll, 2016*; *Lin and Söderhäll, 2011*; *Chen et al., 2005*). In crayfish, different types of TGases are localized in different types of hemocyte, in immature hemocytes and granular hemocytes, respectively (*Junkunlo et al., 2020*). In our scRNA-seq, different types of TGases were expressed in different cluster of hemocytes (*Figure 8—figure supplement 2*), indicating that shrimp TGases are also expressed in different places. More interestingly, it is reported that TGase is not expressed in all granulocytes in crayfish (*Junkunlo et al., 2020*). Our result was in accordance with the previous study because not 100% of the hemocytes expressed TGase in clusters Hem 5 and Hem 6. Alpha2-a2m of shrimp contributes to clotting pathway and are stored in the granules of hemocytes (*Chaikeeratisak et al., 2012*). Since our results also showed strong expression of a2m in
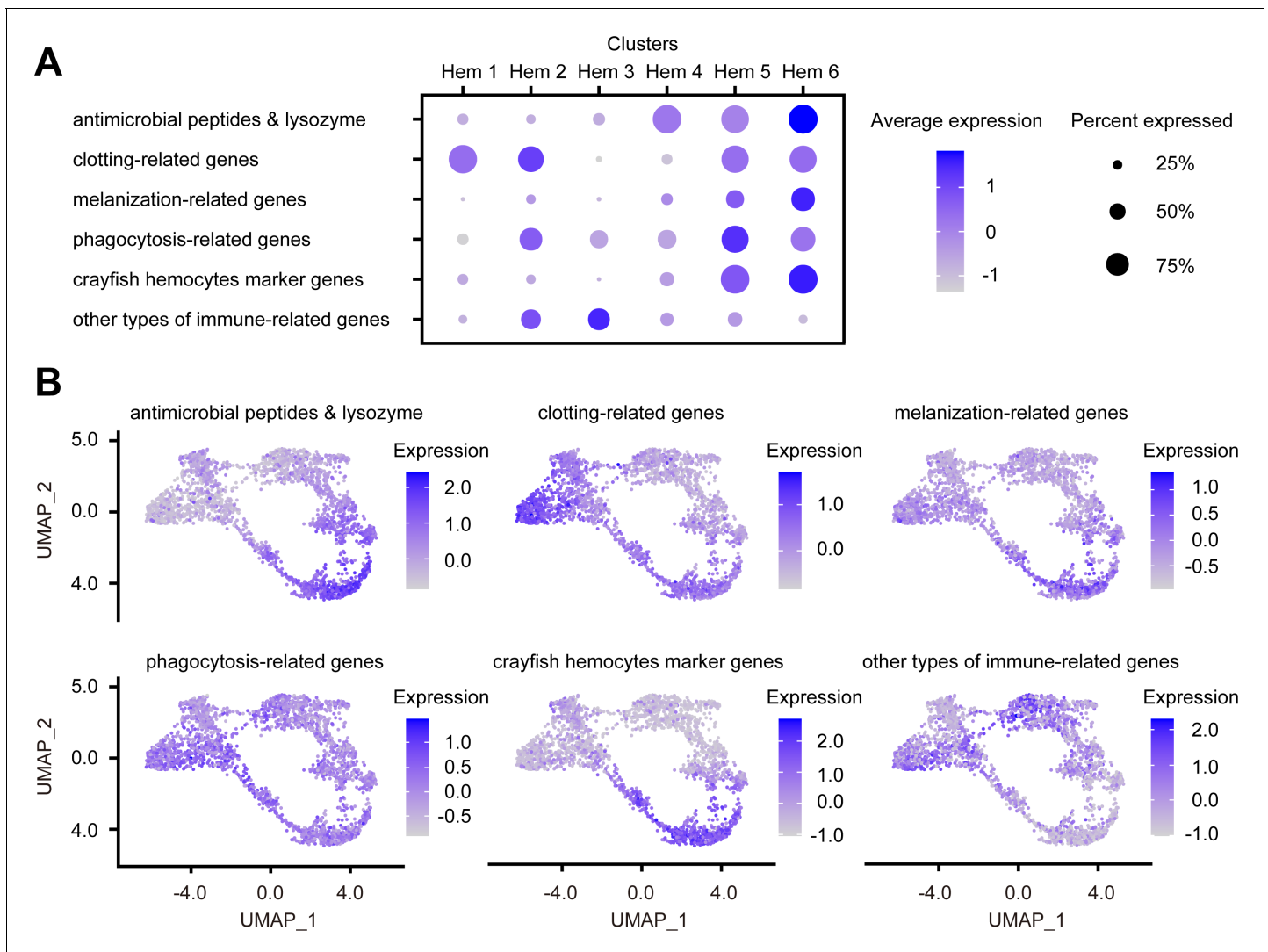
Figure 8. Dot plots and uniform manifold approximation and projection (UMAP) plot profiling of immune-related genes. Dot plot representing the average expression of each immune-related gene per cluster (A). Color gradient of the dot represents the expression level, while the size represents the percentage of cells expressing any genes per cluster. The numbers in parentheses represents the number of genes estimated as immune-related. Expression profiling of immune-related genes on UMAP plot (B). Color gradient of each plot represents the expression level. The details of the identified genes are listed in *Supplementary file 1*.

The online version of this article includes the following figure supplement(s) for figure 8:

Figure supplement 1. Dot plot representing the antibacterial peptides and lysozyme-related genes per cluster based on average expression.
Figure supplement 2. Dot plot representing the clotting-related genes per cluster based on average expression.
Figure supplement 3. Dot plot representing the melanization-related genes per cluster based on average expression.
Figure supplement 4. Dot plot representing the phagocytosis-related genes per cluster based on average expression.
Figure supplement 5. Dot plot representing the crayfish hemocyte marker genes per cluster based on average expression.
Figure supplement 6. Dot plot representing the other types of immune-related genes per cluster based on average expression.

Hem 5 and Hem 6, these clusters are predicted to be a granulocyte. Coagulation factor and hemocytin are related to clotting pathway in invertebrates (*Lavine and Strand, 2002*; *Kotani et al., 1995*). Strong expression of the whole clotting-related genes was predicted in Hem 1, Hem 2, Hem 5, and Hem 6, except Hem 3 and Hem 4 (*Figure 8*). Although it is not possible to estimate which genes are working in clotting pathway and how, it is clear that there were populations, Hem 3 and Hem 4, that are not involved in clotting.

Melanization is performed by PO and controlled by the prophenoloxidase (proPO) activation cascade in shrimp (*Amparyup et al., 2013*). Except prophenoloxidase-activating factor 2-like (PPAF2-

like) and QM protein, melanization-related genes were expressed in Hem 4 to Hem 6. This result is consistent with the fact that *proPO* was a marker gene for mature hemocytes (*Figure 8—figure supplement 3*). In the previous study, the strong PO activity was observed in granules of hemocytes in crayfish (*Giulianini et al., 2007*), suggesting that PO activity and granules are closely related. In addition, it is considered that mature granulocytes are involved in melanization. Although the detailed functions of PPAF2-like and QM protein were still unknown except that these genes are involved in proPO cascade in shrimp, it is reasonable to conclude that melanization or proPO cascade is the product of a cascade reaction through the entire hemocytes, not just granulocytes.

Phagocytosis is a crucial defense mechanism against bacteria and viruses in shrimp. Many types of genes are involved in phagocytosis, such as receptors for recognition, downstream signal pathways, and intracellular regulators (*Liu et al., 2020*). In kuruma shrimp *M. japonicus*, *integrin alpha PS* is reported as a marker for hemocytes that mediates phagocytosis (*Koiwai et al., 2018*). However, it is reported that different subpopulations of hemocytes seem to exhibit specific preferences of different bacteria or viruses in phagocytosis (*Liu et al., 2020*). Since there are many cascades for phagocytosis and various molecules work together, it is difficult to determine which cluster is centralized, but various phagocytosis-related genes are expressed relatively strongly in Hem 5 (*Figure 8—figure supplement 4*). *Rab7* was found to be highly expressed in Hem 1, which is known to be involved in the endosomal trafficking pathway of virus-infected penaeid shrimp (*Zhao et al., 2015*). However, there have been no detailed functional studies of *Rab7* on whether it is involved in phagocytosis in shrimp, so it is early to determine its role in Hem 1.

Some of the hemocyte type-specific markers related to their immune function have been studied in crayfish: proPO in matured hemocytes, copper/zinc superoxide dismutase (SOD) in semi-granular cells (SGC), and Kazal-type proteinase inhibitor (KPI) in granular cells (GC), and TGase in immature cells (*Söderhäll, 2013*). The usefulness of proPO and TGase as markers has already been discussed above. Both *KPI* and *SOD* were mostly expressed at both Hem 5 and Hem 6, and their expression levels were similar between these clusters (*Figure 8—figure supplement 5*). These results suggest that the functional segregation of hemocytes in shrimp is different from that in crayfish.

We also identified other immune-related genes (*Figure 8—figure supplement 6*). *Notch* and *pellino* regulate AMPs expression through the signaling pathway (*Zhang et al., 2020*; *Li et al., 2014*; *Ning et al., 2018*). As discussed in the section explaining cluster-specific markers, *ML protein, VRP,* and *Vago 5* are related to virus infection. Molecules that directly affect defense, such as AMPs and clotting-related genes, are more abundantly expressed in Hem 4 to Hem 6, whereas molecules that regulate immune system seem to be more abundantly expressed in Hem 1 to Hem 3. In other words, it was predicted that agranulocytes play a role in signaling for immune system, and that granulocytes are responsible for direct biological defense functions. In addition, the expression of virus infection-related genes, *ML protein*, *VRP,* and *Vago 5*, suggests Hem 2, Hem 3, and Hem 4 might play important roles when virus infect shrimp.

In the mosquito, scRNA-seq revealed a new subpopulation called 'antimicrobial granulocytes' that expressed characteristic AMPs (*Raddi et al., 2020*). Similarly, in *M. japonicus*, the expression patterns of immune-related genes were also different among certain clusters, suggesting that shrimp hemocytes are more heterogeneous than previously thought. It is anticipated that the class of granulocytes discussed in previous studies is actually a mixture of clusters exhibiting different roles.

Validation of marker genes and the relationship between clusters and morphology

Our scRNA-seq results revealed six major subpopulations and their marker genes, and the possible differentiation trajectory of kuruma shrimp *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 (*Figure 9A–E*, *Figure 9—figure supplement 1*): 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 differential interference contrast (DIC) and dye staining imaging (*Figure 9A–D*), we observed that cells in the R1 region contained no or few granules in the cytoplasm. The nucleus occupied a large portion of the
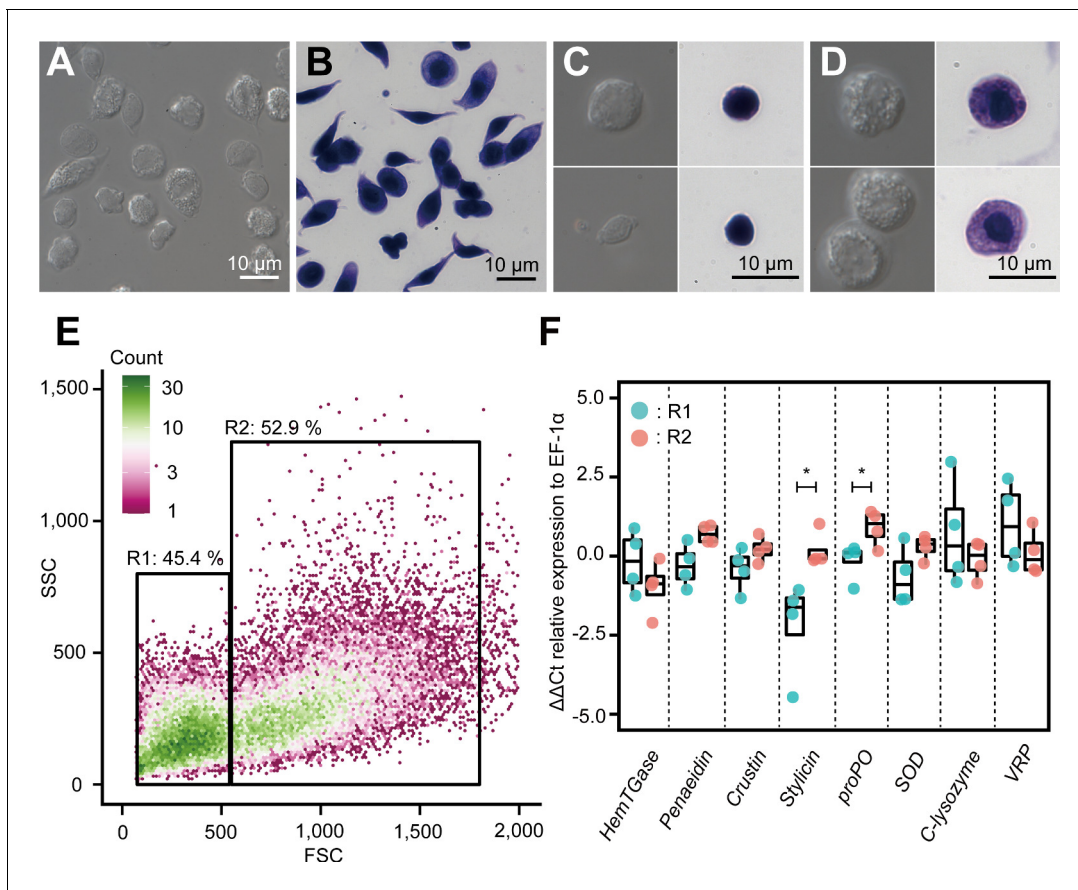
Figure 9. Morphological analysis of hemocytes and transcript profiles based on morphology. (**A**) Differential interference contrast (DIC) image of unsorted total hemocytes. (**B**) Dye-stained total hemocytes. (**C**) DIC imaging and dye staining of region 1 (R1)-sorted hemocytes. (**D**) DIC imaging and dye staining of region 2 (R2)-sorted hemocytes. (**E**) Fluorescence-activated cell sorting (FACS) analysis of hemocytes. Based on the forward scatter (FSC) and side scatter (SSC) two-dimensional space, two regions (R1 and R2) were obtained. (**F**) Differential gene expression analysis between R1 and R2 of hemocytes sorted using FACS. $\Delta\Delta$Ct values were analyzed using qRT-PCR. Higher $\Delta\Delta$Ct values indicate a higher accumulation of mRNA transcripts. The p values shown in the figures are represented by *$p<0.05$.

The online version of this article includes the following source data and figure supplement(s) for figure 9:

Source data 1. Source data for CT values of each gene used for *Figure 9F*.
Figure supplement 1. Ffluorescence-activated cell sorting analysis of hemocytes from four individual shrimps (**A–D**).

volume in these cells (*Figure 9C*). Conversely, those in the R2 region had many granules in the cytoplasmic region, which occupied a large portion of cells (*Figure 9D*).

We conducted qRT-PCR analysis to determine the expression of some representative genes in these populations. The results showed that the average $\Delta\Delta$Ct values of the transcripts of *HemTGase* were higher in R1 hemocytes than in R2 hemocytes, while the values of *penaeidin*, *crustin*, *stylicin*, *proPO*, and *SOD* were higher in R2 hemocytes (*Figure 9F*). The average $\Delta\Delta$Ct 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. Some of these genes were not statistically different in the t-test, but due to the small sample size, we used the mean values for discussion. Granulocytes found in the R2 region expressed major AMPs, such as *penaeidin*, *crustin*, and *stylicin*, suggesting that they consist of clusters Hem 4 to Hem 6. However, *c-type lysozyme*, which was expressed in Hem 4 to Hem 6 in scRNA-seq, was higher on the cells in R1 region. The reason why only *c-lysozyme* was expressed in different population from the other AMPs was unclear. *VRP*, which is a marker of Hem 3 (*Figure 8—figure supplement 6*), was expressed in cells in both R1 and R2 regions, indicating that Hem

3 exists in both populations and is indistinguishable from the morphological characteristics. Our conclusion that the cells in R1 and R2 regions separated by FACS consist of Hem 1–3 and Hem 4–6, respectively, is quantitatively supported by the fact that Hem 1–3 and Hem 4–6 cover 60.4 and 39.6% of whole cells, respectively.

Discussion

Our single-cell transcriptome analysis revealed that there are six subpopulations of hemocytes in shrimp. This result provides us a detailed and advanced understanding of the cell classification over previous strategies based on various approaches, such as simple staining (*Söderhäll and Smith, 1983*; *Johansson et al., 2000*; *van de Braak et al., 2002*), monoclonal antibodies (*Xing et al., 2017*), flow cytometry (*Koiwai et al., 2017*), and lectin-binding profiles (*Koiwai et al., 2019*). We hope that future studies will clarify the relationships between these phenotypic features and our classification to comprehend the role of each cluster in the immune system of shrimp.

The cluster-specific markers and cell proliferation-related genes found here helped us to understand how shrimp hemocytes differentiate. A strong expression of *TGase*, cell proliferation-related genes, and G2/M state-related genes in Hem 1 suggested that hemocytes in this cluster are oligopotent and located upstream in the differentiation process. In crustaceans, especially shrimp and crayfish, it is known that hemocytes are produced in HPT, and that differentiated hematopoietic cells from HPT circulate in the body fluid (*van de Braak et al., 2002*; *Söderhäll, 2016*; *Söderhäll, 2013*). Hem 1 accounts for only 6.5% of the analyzed cells, which indicates that only a very small fraction of oligopotent or initial state hemocytes exists among the circulating hemocytes. Many of the markers that were characteristically expressed in Hem 1 could not be functionally predicted by the BLAST search (*Supplementary file 1*). We speculate that the characteristic genes of Hem 1 are associated with cell duplication or differentiation. The division and differentiation mechanisms of shrimp hemocytes are still largely unknown, and no techniques on culture shrimp hemocytes in vitro have been reported. The analysis of these unknown gene characteristics may reveal these mechanisms.

Our results also revealed that the growth-related genes were expressed on specific clusters (*Figure 7*). It was predicted that several growth factors are involved in differentiation and maturation of hemocytes, and that these factors play roles in the maturation of hemocytes through several cellular processes. In *P. monodon*, a model has been proposed in which hemocytes differentiate linear to two types of granulocytes from a single type of agranulocytes (*van de Braak et al., 2002*). Our study strongly supports this previous study. We would also like to emphasize that shrimp hemocytes are not completely separated into distinct cell types, and their differentiation process is rather continuous. Loss and gain function studies of these growth-related genes are necessary to prove the full differentiation process of penaeid shrimp hemocytes.

Crayfish's hemocytes and *Drosophila* hemocytes have been classified into three and four major types according to the marker genes, respectively (*Tattikota et al., 2020*; *Cho et al., 2020*; *Söderhäll, 2016*). The relationship between *Drosophila* and shrimp was almost impossible to establish on the hemocyte-type markers. Also, for crayfish hemocyte markers, there was no difference between clusters according to *SOD* and *KPI* that are the markers of SGC and GC, respectively. Insects and crustaceans are thought to become independent about 500 million years ago (*Rota-Stabelli et al., 2013*; *Thomas et al., 2020*), and shrimp and crayfish are thought to have become evolutionarily independent about 450 million years ago (*Wolfe et al., 1901*). Thus, it is straightforward to reason that the functions of these genes have changed during the evolutionary process. Therefore, we should not simply argue that the morphological and functional similarities between shrimp and *Drosophila*/crayfish hemocytes are the same. Interestingly, however, *proPO* (or *PPO*) was found to be highly expressed in mature hemocytes in shrimp, *Drosophila*, and crayfish alike. Therefore, it is possible that *proPO* is a marker gene for mature hemocytes, at least in common with these three species. It is unclear to what extent this trend will hold, but it is interesting to examine why this trend is conserved across species.

There have been a number of studies to investigate the function of hemocytes, but most of them were done on whole hemocytes because of the difficulty in distinguishing and isolating cell types. Here, we predicted the major functions, that is AMPs releasing, clotting, melanization, and phagocytosis, of each hemocyte population. The strong expression of the AMPs, melanization, and phagocytosis-related genes in mature hemocytes revealed that granulocytes are the cells that play a central

role in immune activity in shrimp. Furthermore, clotting-related genes were strongly expressed in Hem 1 and Hem 2, indicating that agranulocytes are the main player in clotting. In addition, the expression of genes that is known to be elevated by viral infection was high in Hem 3 and Hem 4. This may indicate that these hemocyte populations are the ones that work during virus infection. Or, the percentage of this hemocyte population may increase during viral infection, while the amount of gene expression in the hemocytes remains the same. This aspect will be clarified by studying the expression dynamics in shrimps infected with viruses or bacteria. In mature hemocytes, the final products of the immune response, such as *AMPs*, *a2m*, and *proPO*, were strongly expressed, whereas the genes responsible for the middle of the cascade, such as *PPAF*, *Notch*, *pellino*, *ML protein*, and *Vago 5*, were expressed in immature or differentiated hemocytes. This observation suggests that the immune system is not governed solely by granulocytes, but there is a communication with agranulocytes in response to foreign substances.

In conclusion, we succeeded in classifying shrimp hemocytes into six 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 HPTs 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. There is room for improvement in genomic information, library preparation efficiency, functional prediction, etc., but this kind of single-cell study in shrimp will provide efficient solutions to aquaculture problems.

Materials and methods

Key resources table

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Biological sample (*Marsupenaeus japonicus*)	Hemocytes	NA	NA	Hemocytes from hemolymph from 20 g of kuruma shrimp
Sequence-based reagent	1st PCR primer	*Macosko et al., 2015*	DOI: 10.1016/j.cell.2015.05.002	AAGCAGTGGTATCAACGCAGAGT
Sequence-based reagent	P5 universal primer	Illumina, Inc	NA	AATGATACGGCGACCACCGAGA TCTACACGCCTGTCCGCGGAAG CAGTGGTATCAACGCAGAGT*A*C
Sequence-based reagent	i7 index primer	Illumina, Inc	NA	N703: CAAGCAGAAGACGGCATACG AGATTTCTGCCTGTCTCGTGGGCTCGG N704: CAAGCAGAAGACGGCATACGAG ATGCTCAGGAGTCTCGTGGGCTCGG N705: CAAGCAGAAGACGGCATACGAGATAG GAGTCCGTCTCGTGGGCTCGG
Sequence-based reagent	Custom sequence primer	*Macosko et al., 2015*	DOI: 10.1016/j.cell.2015.05.002	GCCTGTCCGCGGAAGCAG TGGTATCAACGCAGAGTAC

Continued on next page

Continued

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Sequence-based reagent	EF-1α	*Koiwai et al., 2019*	DOI:10.1007/s12562-019-01311-5 sequence accession number: AB458256	For: ATTGCCACACCGCTCACA Rev: TCGATCTTGGTCAGCAGTTCA
Sequence-based reagent	HemTGase	*Yeh et al., 2006*	DOI: 10.1016/j.bbapap.2006.04.005 sequence accession number: DQ436474	For: GAGTCAGAAGTCGCCGAGTGT Rev: TGGCTCAGCAGGTCGTTTAA
Sequence-based reagent	Penaeidin	*An et al., 2016*	DOI: 10.1016/j.dci.2016.02.001 sequence accession number: KU057370	For: TTAGCCTTACTCTGTCAAGTGTACGCC Rev: AACCTGAAGTTCCGTAGGAGCCA
Sequence-based reagent	Crustin	*Hipolito et al., 2014*	DOI: 10.1016/j.dci.2014.06.001 sequence accession number: AB121740	For: AACTACTGCTGCGAAAGGTCTCA Rev: GGCAGTCCAGTGGCTTGGTA
Sequence-based reagent	Stylicin	*Liu et al., 2015*	DOI: 10.1016/j.fsi.2015.09.044 sequence accession number: KR063277	For: GGCTCTTCCTTTTCACCTG Rev: GTCGGGCATTCTTCATCC
Sequence-based reagent	proPO	*Koiwai et al., 2019*	DOI: 10.1007/s12562-019-01311-5 sequence accession number: AB073223	For: CCGAGTTTTGTGGAGGTGTT Rev: GAGAACTCCAGTCCGTGCTC
Sequence-based reagent	SOD	*Hung et al., 2014*	DOI: 10.1016/j.fsi.2014.07.030 sequence accession number: AB908996	For: GCCGACACTTCCGACATCA Rev: TTTTGCTTCCGGGTTGGA
Sequence-based reagent	C-lysozyme	*Hikima et al., 2003*	DOI:10.1016/s0378-1119(03)00761-3	For: ATTACGGCCGCTCTGAGGTGC Rev: CCAGCAATCGGCCATGTAGC
Sequence-based reagent	VRP	*Elbahnaswy et al., 2017*	DOI: 10.1016/j.fsi.2017.09.045 sequence accession number: LC179543	For: CTACGGTCGCTACCTTCGTTTG Rev: TCAACAACGCTTCTGAACTTATTCC
Commercial assay or kit	TRI REAGENT	Molecular Research Center, Inc	TR118	NA
Commercial assay or kit	Direct-zol RNA MiniPrep	Zymo Research	R2050	NA
Commercial assay or kit	Dynabeads Oligo(dT)25	Thermo Fisher Scientific	DB61002	NA
Commercial assay or kit	Direct RNA Sequencing kit	Oxford Nanopore Technologies	SQK-RNA002 kit	NA
Commercial assay or kit	MinION Flow Cell	Oxford Nanopore Technologies	Flow Cell R9.4.1	NA
Commercial assay or kit	Negative Photoresist	Nippon Kayaku Co., Ltd.	SU-8 3050	NA
Commercial assay or kit	Polydimethylsiloxane sylgard 184	Dow Corning Corp.	SYLGARD 184 Silicone Elastomer Kit	NA
Commercial assay or kit	Barcoded Bead SeqB	ChemGenes Corporation	MACOSKO-2011–10	NA
Commercial assay or kit	Maxima H Minus Reverse Transcriptase	Thermo Fisher Scientific	EP0751	NA

Continued on next page

Continued

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Commercial assay or kit	Exonuclease I	New England Biolabs	M0293S	NA
Commercial assay or kit	KAPA HiFi HotStart ReadyMix	Roche Ltd.	KK2601	NA
Commercial assay or kit	KAPA HiFi DNA polymerase	Roche Ltd.	KK2103	NA
Commercial assay or kit	Agencourt AMPure XP beads	Beckman Coulter	A63882	NA
Commercial assay or kit	DNA Clean and Concentrator Kit	Zymo Research	D4013	NA
Commercial assay or kit	Qubit dsDNA HS Assay Kit	Thermo Fisher Scientific	Q32851	NA
Commercial assay or kit	High-Capacity cDNA Reverse Transcription Kit	Thermo Fisher Scientific	4368814	NA
Commercial assay or kit	KOD SYBR qPCR	TOYOBO Co. Ltd.	QKD-201	NA
Cell staining solution	May-Grünwald's eosin methylene blue solution modified	Merck KGaA	101424	NA
Cell staining solution	Giemsa's Azure Eosin Methylene Blue solution	Merck KGaA	109204	NA
Software, algorithm	Guppy v3.6.1	Oxford Nanopore Technologies	NA	https://community.nanoporetech.com/
Software, algorithm	MinKNOW v3.6.5	Oxford Nanopore Technologies	NA	https://community.nanoporetech.com/
Software, algorithm	TALC v1.01	*Broseus et al., 2020a*, *Broseus et al., 2020b*	DOI: 10.1093/bioinformatics/btaa634	https://gitlab.igh.cnrs.fr/lbroseus/TALC
Software, algorithm	rnaSPAdes v3.14.1	*Bushmanova et al., 2019a*, *Bushmanova et al., 2019b*	DOI: 10.1093/gigascience/giz100	https://cab.spbu.ru/software/rnaspades/
Software, algorithm	Trinity 2.10.0	*Grabherr et al., 2011*, *Grabherr et al., 2018*	DOI: 10.1038/nbt.1883	https://github.com/trinityrnaseq/trinityrnaseq/wiki
Software, algorithm	EvidentialGene v2022.01.20	*Gilbert, 2019*	NA	http://arthropods.eugenes.org/EvidentialGene/
Software, algorithm	BUSCO v5.0.0	*Seppey et al., 1962*	DOI: 10.1007/978-1-4939-9173-0_14	https://busco.ezlab.org/
Software, algorithm	Blast+ v2.2.31	*Altschul et al., 1990*; *Camacho et al., 2009*	DOI: 10.1016/s0022-2836(05)80360-2 DOI: 10.1186/1471-2105-10-421	https://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/
Software, algorithm	Drop-seq tools v2.3.0	McCarroll Lab *Wysoker et al., 2020*	NA	https://github.com/broadinstitute/Drop-seq
Software, algorithm	Picard Toolkit	Broad Institute *Picard Toolkit, 2019*	NA	http://broadinstitute.github.io/picard/
Software, algorithm	STAR v2.7.8a	*Dobin et al., 2013*; *Dobin et al., 2021*	DOI: 10.1093/bioinformatics/bts635	https://github.com/alexdobin/STAR
Software, algorithm	Surat v4.0.1	*Butler et al., 2018*; *Stuart et al., 2019*	DOI: 10.1038/nbt.4096 DOI: 10.1016/j.cell.2019.05.031	https://satijalab.org/seurat/
Software, algorithm	Monocle 3 v0.2.3.0	*Trapnell et al., 2014*, *Trapnell et al., 2021*	DOI: 10.1038/nbt.2859	https://github.com/cole-trapnell-lab/monocle3

Shrimp and cell preparation

Twenty-three female kuruma shrimp, *M. japonicus*, with an average weight of 20 g, were purchased from a local distributor and maintained in artificial seawater with a 34 ppt salinity with a recirculating

system at 20°C. Hemolymph was collected using an anticoagulant solution suitable for penaeid shrimp (*Söderhäll and Smith, 1983*) from an abdominal site. The collected hemolymph was centrifuged at 800 × *g* for 10 min to collect the hemocytes, which were then washed twice with PBS, and the osmolarity was adjusted to kuruma shrimp (KPBS: 480 mM NaCl, 2.7 mM KCl, 8.1 mM Na$_2$H-PO$_4$·12H$_2$O, 1.47 mM KH$_2$PO$_4$, pH 7.4).

Preparation of expressing gene list of hemocytes

De novo assembled transcript data were prepared as a reference for mapping Drop-seq data because the genome sequence of *M. japonicus* is still unknown. To improve the quality of the de novo assembled transcript sequences, we prepared long-read mRNA sequences using MinION (Oxford Nanopore Technologies) direct RNA sequencing to conduct hybrid de novo assembly. Poly (A) tailed RNA was purified from 58 μg of total RNA from the hemocytes of 16 shrimp using Dynabeads Oligo(dT)$_{25}$ (Thermo Fisher Scientific), and 500 ng of poly-(A) RNA was ligated to adaptors using a direct RNA sequencing kit (Oxford Nanopore Technologies) according to the manufacturer's manual version DRS_9080_v2_revL_14Aug2019. Finally, 44 ng of the library was obtained and sequenced using MinION by using a MinION flow cell R9.4.1 (Oxford Nanopore Technologies). All sequencing experiments were performed using MinKNOW v3.6.5 without base calling. Raw sequence data were then base-called using Guppy v3.6.1. Once the raw signal from the MinION fast5 files was converted into fastq files, the sequencing errors were corrected using TALC v1.01 (*Broseus et al., 2020a*, https://gitlab.igh.cnrs.fr/lbroseus/TALC, *Broseus et al., 2020b*), by using the Illumina short reads sequence of *M. japonicus* shrimp hemocytes (DDBJ Sequence Read Archive [DRA] accession number DRA004781). The corrected long-read sequences from MinION and short-read sequences from Illumina Miseq were hybrid de novo assembled using rnaSPAdes v3.14.1 (*Bushmanova et al., 2019a*, https://cab.spbu.ru/software/rnaspades/, *Bushmanova et al., 2019b*) and Trinity 2.10.0 (*Grabherr et al., 2011*, https://github.com/trinityrnaseq/trinityrnaseq/wiki, *Grabherr et al., 2018*). All assembled de novo transcripts were merged and subjected to the EvidentialGene program v2022.01.20 (http://arthropods.eugenes.org/EvidentialGene/, *Gilbert, 2019*) to remove similar sequences with a default parameter. The remaining sequences were renamed as Mj-XXX and were subjected to BUSCO analysis to check the quality of assembling (*Seppey et al., 1962*). The assembled transcripts were used as a hemocyte-expressing gene list. The assembled sequences and code used to perform base-calling and de novo assembly are available on GitHub at https://github.com/KeiichiroKOIWAI/Drop-seq_on_shrimp (*Koiwai, 2021a*, copy archived at swh:1: rev:cddd130b4cb841168bf3e29d0fbfef0de5ef2ad7, *Koiwai, 2021b*).

Single-cell and single-bead encapsulation by a microfluidic device and exonuclease and reverse transcribe reaction on a bead

The Drop-seq procedure was used to encapsulate single hemocytes and single mRNA capture beads together into fL-scale microdroplets, as previously described (*Macosko et al., 2015*). The following steps were performed in triplicates for three shrimp individuals. Briefly, the self-built Drop-seq microfluidic device was prepared by molding polydimethylsiloxane (PDMS; Sylgard 184, Dow Corning Corp.) from the microchannel structure formed by the negative photoresist (SU-8 3050, Nippon Kayaku Co.). Using this device, droplets containing a cell and a Barcoded Bead SeqB (ChemGenes Corporation) were produced up to 2 mL per sample using a pressure pump system (Droplet generator, On-chip Biotechnologies Co., Ltd.). During the sample introduction, the vial bottles containing cells and beads were shaken using a vortex mixer to prevent sedimentation and aggregation (*Biočanin et al., 2019*). Droplets were collected from the channel outlet into the 50 mL corning tube and incubated at 80°C for 10 min in a water bath to promote hybridization of the poly(A) tail of mRNA and oligo d(T) on beads. After incubation, droplets were broken promptly and barcoded beads with captured transcriptomes were reverse transcribed using Maxima H Minus Reverse Transcriptase (Thermo Fisher Scientific) at room temperature (RT) for 30 min, then at 42°C for 90 min. Then, the beads were treated with Exonuclease I (New England Biolabs) to obtain single-cell transcriptomes attached to microparticles (STAMP). The first-strand cDNAs on beads were amplified using PCR. The beads obtained above were distributed throughout PCR tubes (1000 beads per tube), wherein 1× KAPA HiFi HS Ready Mix (KAPA Biosystems) and 0.8 μM 1st PCR primer were included in a 25 μL reaction volume. PCR amplification was achieved using the following program:

initial denaturation at 95℃ for 3 min; 4 cycles at 98℃ for 20 s, 65℃ for 45 s, and 72℃ for 6 min; 12 cycles of 98℃ for 20 s, 67℃ for 20 s, and 72℃ for 6 min; and a final extension at 72℃ for 5 min. The amplicons were pooled, double-purified with 0.9× AMPure XP beads (Beckman Coulter), and eluted in 100 μL of ddH$_2$O. Sequence-ready libraries were prepared according to *Picelli et al., 2014*. A total of 1 ng of each cDNA library was fragmented using home-made Tn5 transposome in a solution containing 10 mM TAPS-NaOH (pH 8.5), 5 mM MgCl$_2$, and 10% dimethylformamide at 55℃ for 10 min. The cDNA fragments were purified using a DNA Clean and Concentrator Kit (Zymo Research) and eluted in 25 μL of ddH$_2$O. The index PCR reaction was performed by adding 12 μL of the elute to a mixture consisting of 1× Fidelity Buffer, 0.3 mM dNTPs, 0.5 U KAPA HiFi DNA polymerase (KAPA Biosystems), 0.2 μM P5 universal primer, and 0.2 μM i7 index primer. Each reaction was achieved as follows: initial extension and subsequent denaturation at 72℃ for 3 min and 98℃ for 30 s; 12 cycles of 98℃ for 10 s, 63℃ for 30 s, and 72℃ for 30 s; and a final extension at 72℃ for 5 min. The amplified library was purified using 0.9× AMPure XP beads and sequenced (paired-end) on an Illumina NextSeq 500 sequencer (NextSeq 500/550 High Output v2 kit [75 cycles]); 20 cycles for read1 with custom sequence primer, 8 cycles for index read, and 64 cycles for read2. Before performing the Drop-seq on *M. japonicus* shrimp hemocytes, we validated the protocol by performing the same procedure using a mixture of HEK293 and NIH3T3 cells and sequencing the test library using a Miseq Reagent Kit v3 (150 cycles).

Analysis of single-cell data

Paired-end reads were processed and mapped to the reference de novo assembled gene list of hemocytes following the Drop-seq Core Computational Protocol version 2.0.0 and the corresponding Drop-seq tools v2.3.0 (https://github.com/broadinstitute/Drop-seq; *Wysoker et al., 2020*), provided by McCarroll Lab (http://mccarrolllab.org/dropseq/). The Picard suite (https://github.com/broadinstitute/picard, *Picard Toolkit, 2019*). was used to generate the unaligned bam files. The steps included the detection of barcode and UMI sequences, filtration and trimming of low-quality bases and adaptors or poly(A) tails, and the alignment of reads using STAR v2.7.8a (*Dobin et al., 2013*). The cumulative distribution of reads from the aligned bam files was obtained using BAMTagHistogram, and the number of cells was inferred using Drop-seq tools.

Data integration

After digital expression data from three shrimps were read using Seurat v4.0.1 (*Butler et al., 2018*; *Hafemeister and Satija, 2019*) (https://satijalab.org/seurat/). To remove data from low-quality cells or empty droplets, we filtered digital expression data with unique feature counts over 4000 or less than 500 and mitochondrial counts over 10% or less than 1%, then the remained 2566 dataset was used for the further analysis. SCTransform (*Hafemeister and Satija, 2019*) was performed to remove the technical variation, while retaining biological heterogeneity. We ran a PCA using the expression matrix of the top 3000 most variable genes. The total number of principal components (PCs) required to compute and store was 50. The UMAP was then performed using the following parameters: n.neighbors, min.dist, and n.components were 50L, 0.2, and 2, respectively, to visualize the data in the two-dimensional space, and then the clusters were predicted with a resolution of 0.4.

Functional prediction of the highly expressed genes and clusters

The de novo assembly constructed 30,986 transcripts. However, not all transcripts were not so expressed in our single-cell data. Thereby, we limited to the highly expressed genes as 'Ex50' that represent 50% of the total normalized expression data. The functions of Ex50 were predicted using BLAST program v2.2.31 (*Camacho et al., 2009*; *Altschul et al., 1990*) on three common penaeid shrimp, *L. vannamei*, *P. monodon*, and *M. japonicus*, proteins (totally 62,242 proteins were downloaded from NCBI Protein Groups on 10 April 2021) with the blastx parameter of e-value as 1e-6, and KOGs and GOs of each Ex50 were also predicted using TransDecoder (https://github.com/TransDecoder/TransDecoder, *Zimmermann et al., 2018*) and eggNOG-mapper (http://eggnog-mapper.embl.de/, *Cantalapiedra et al., 2021*). From the annotated result, 21 KOGs and eleven GOs were selected, then average expressions were calculated on the clusters by AddModuleScore function of Seurat. The functions of each cluster were also predicted based on the marker genes.

Marker genes were predicted using the Seurat FindAllMarkers tool with the following parameters: min.pct as 0.5, logfc.threshold as 1, and test.

Prediction of cell cycles based on *Drosophila* G2/M and S phase marker genes

To calculate cell cycle, Ex50 genes were annotated with *Drosophila* G2/M and S phase marker genes (https://github.com/hbc/tinyatlas, *Kirchner and Barrera, 2019*) using blastx parameter of e-value as 1e-6. Based on this result, we calculated cell cycle score of each single cell by Seurat CellCycleScoring function.

Pseudo-temporal ordering of cells using Monocle 3

The integrated data of Seurat were transferred to Monocle 3 (*Trapnell et al., 2014*, https://github.com/cole-trapnell-lab/monocle3, *Trapnell et al., 2021*), to calculate a cell trajectory using the learn_graph function. We assigned the start point based on the expression of cell proliferation-related genes and *Drosophila* marker genes of the cell cycle.

Comparison with *Drosophila* hemocyte marker genes

To check whether the *Drosophila* marker genes are applicable to shrimp, we performed a BLAST search on the *Drosophila* cell-type markers (https://github.com/hbc/tinyatlas, *Kirchner and Barrera, 2019*). Ex50 were blastx searched on *D. melanogaster* genes (dmel-all-gene-r6.34.fasta; downloaded from FlyBase https://flybase.org/) with the parameters of e-value as 1e-6.

Visualization of cell growth-related and immune-related genes on single hemocytes

To visualize the immune-related genes of shrimp, we identified the distinct sequences from Ex50 based on their blastx results against penaeid shrimp proteins. Here, we focused on the cell growth-related genes and on immune-related genes, such as AMPs, lysozyme, clotting-related, melanization-related, phagocytosis-related, superoxide dismutase, Kazal-type proteinase inhibitors, and other immune-related genes listed in *Supplementary file 1*. The average expressions of each immune-related genes were calculated on the clusters by AddModuleScore function of Seurat.

Cell sorting of hemocytes and qRT-PCR of marker genes

To validate the Drop-seq results on hemocytes, populations of hemocytes in the forward scatter (FSC) and side scatter (SSC) two-dimensional space were sorted using a microfluidic cell sorter (On-chip sort, On-chip Biotechnologies Co., Ltd.) from four shrimp individuals (*Figure 9—figure supplement 1*). In the FSC/SSC two-dimensional space, two main populations were predicted as R1: small/simple and R2: large/complexity populations, which were defined as agranulocytes and granulocytes, respectively. After sorting, some sorted hemocytes were immediately fixed in 2% formalin in KPBS and stained with a May-Grunwald and Giemsa staining solution to observe the cellular components. Briefly, fixed hemocytes were smeared on a glass slides and dried, stained for 5 min with 20% May-Grunwald stain solution (Merck KGaA) in 6.67 mM phosphate buffer (pH 6.6), washed with phosphate buffer, stained for 15 min with 4% Giemsa stain solution (Merck KGaA) in 6.67 mM phosphate buffer (pH 6.6), and washed with tap water, dried, mounted with malinol (Muto Pure Chemicals). Non-stained and stained hemocytes were subjected to microscopy IX71 (Olympus Corporation) to observe their structures.

Total RNA was also collected from sorted cells and pre-sorted cells. The concentration of RNA was measured using a nanodrop, and cDNA was transcribed using a High-Capacity cDNA Reverse Transcription Kit (Thermo Fisher Scientific). Constructed cDNA was diluted five times with TE buffer and subjected to qRT-PCR with a number of cycles as 40 using KOD SYBR qPCR (TOYOBO Co. Ltd.), following the manufacturer's protocol. When the CT value was not detectable, we set a CT value of 40 to use for ΔCT calculation. The expression of each gene was calculated using the ΔΔCT method against elongation factor-1 alpha and total hemocytes.

Statistics

Statistics of qRT-PCR were performed on R software. Significance between two hemocyte R1 and R2 was calculated by unpaired t-test. The p values shown in the figures are represented by *$p<0.05$.

Data files and analysis code

The raw sequence data of newly sequenced *M. japonicus* transcriptomic reads were archived in the DDBJ Sequence Read Archive (DRA) of the DNA Data Bank of Japan as follows: MinION mRNA direct sequencing: DRA010948; Drop-seq shrimp rep1: DRA010950; shrimp rep2: DRA010951; shrimp rep3: DRA010952; mixture sample of HEK293 and 3T3 cells: DRA010949. The Seurat digital expression data were archived in the Genomic Expression Archive of the DNA Data Bank of Japan: E-GEAD-403. Fast5 data of MinION direct RNA sequencing will be made available upon request from the authors. The code used to perform de novo assembly, clustering, and marker analysis is available on GitHub at https://github.com/KeiichiroKOIWAI/Drop-seq_on_shrimp, *Koiwai, 2021a*.

Additional information

Competing interests

Soichiro Tsuda: is an employee of bitBiome inc, and does not own any stock of the company. The other authors declare that no competing interests exist.

Author contributions

Keiichiro Koiwai, Conceptualization, Resources, Data curation, Formal analysis, Funding acquisition, Investigation, Visualization, Writing - original draft, Project administration, Writing - review and editing; Takashi Koyama, Resources, Investigation, Writing - review and editing; Soichiro Tsuda, Investigation, Writing - review and editing; Atsushi Toyoda, Resources, Investigation; Kiyoshi Kikuchi, Resources, Supervision, Funding acquisition; Hiroaki Suzuki, Ryuji Kawano, Resources, Supervision, Writing - review and editing

Author ORCIDs

Keiichiro Koiwai (iD) https://orcid.org/0000-0002-8890-8229
Takashi Koyama (iD) https://orcid.org/0000-0001-7140-7244
Atsushi Toyoda (iD) http://orcid.org/0000-0002-0728-7548

Hiroaki Suzuki [ID] https://orcid.org/0000-0002-8899-0955
Ryuji Kawano [ID] https://orcid.org/0000-0001-6523-0649

Decision letter and Author response
Decision letter https://doi.org/10.7554/eLife.66954.sa1
Author response https://doi.org/10.7554/eLife.66954.sa2

Additional files

Supplementary files
• Supplementary file 1. Table representing blastx researching of assembled genes against penaeid shrimp's proteins.

• Supplementary file 2. Table representing predicted marker genes per cluster.

• Supplementary file 3. Table representing blastx researching of assembled genes against *Drosophila* cell cycle markers.

• Supplementary file 4. Table representing blastx researching of assembled genes against *Drosophila* cell-type markers.

• Transparent reporting form

Data availability
Sequencing data have been deposited in DDBJ under accession codes DRA010948, DRA010949, DRA010950, DRA010951, and DRA010952. Digital expression data of Drop-seq from three shrimp were archived in DDBJ under accession code E-GEAD-403. Data code can be accessed at https://github.com/KeiichiroKOIWAI/Drop-seq_on_shrimp (copy archived at https://archive.softwareheritage.org/swh:1:rev:cddd130b4cb841168bf3e29d0fbfef0de5ef2ad7). All data generated or analyzed during this study are included in the manuscript and supporting files. All tables are provided as Supplementary files. Source data files are provided to support the sinaplots in Figure 2, Figure 2-figure supplement 1. Source data files are provided to support thepercentage of cell state in Figure 5. Source data file is provided to support the bar graph in Figure 9F.

The following datasets were generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Koiwai K, Koyama T, Tsuda S, Toyoda A, Kikuchi K, Suzuki H, Kawano R	2020	Direct RNA sequencing data of hemocytes of Marsupenaeus japonicus using MinION	https://trace.ddbj.nig.ac.jp/DRASearch/submission?acc=DRA010948	DDBJ Sequence Read Archive, DRA010948
Koiwai K, Koyama T, Tsuda S, Toyoda A, Kikuchi K, Suzuki H, Kawano R	2020	Raw sequence data of drop-seq on shrimp 1	https://trace.ddbj.nig.ac.jp/DRASearch/submission?acc=DRA010950	DDBJ Sequence Read Archive, DRA010 950
Koiwai K, Koyama T, Tsuda S, Toyoda A, Kikuchi K, Suzuki H, Kawano R	2020	Raw sequence data of drop-seq on shrimp 2	https://trace.ddbj.nig.ac.jp/DRASearch/submission?acc=DRA010951	DDBJ Sequence Read Archive, DRA010 951
Koiwai K, Koyama T, Tsuda S, Toyoda A, Kikuchi K, Suzuki H, Kawano R	2020	Raw sequence data of drop-seq on shrimp 3	https://trace.ddbj.nig.ac.jp/DRASearch/submission?acc=DRA010952	DDBJ Sequence Read Archive, DRA010 952
Koiwai K, Koyama T, Tsuda S, Toyoda A, Kikuchi K, Suzuki H, Kawano R	2020	Digital expression data of Drop-seq from three shrimp	ftp://ftp.ddbj.nig.ac.jp/ddbj_database/gea/experiment/E-GEAD-000/E-GEAD-403	DDBJ Genomic Expression Archive, E-GEAD-403
Koiwai K, Koyama T, Tsuda S, Toyoda A, Kikuchi K, Suzuki H, Kawano R	2020	Raw sequence data of drop-seq on mixture sample of HEK293 and 3T3 cells	https://trace.ddbj.nig.ac.jp/DRASearch/submission?acc=DRA010949	DDBJ Sequence Read Archive, DRA010949

The following previously published dataset was used:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Koiwai K, Alenton RRR, Shiomi R, Nozaki R, Kondo H, Hirono I	2016	Transcriptome analysis on different sub-populations of hemocytes of kuruma shrimp Marsupenaeus japonicus	https://trace.ddbj.nig.ac.jp/DRASearch/submission?acc=DRA004781	DDBJ Sequence Read Archive, DRA004781

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**:403–410. DOI: https://doi.org/10.1016/S0022-2836(05)80360-2, PMID: 2231712

Amparyup P, Charoensapsri W, Tassanakajon A. 2013. Prophenoloxidase system and its role in shrimp immune responses against major pathogens. *Fish & Shellfish Immunology* **34**:990–1001. DOI: https://doi.org/10.1016/j.fsi.2012.08.019, PMID: 22960099

An MY, Gao J, Zhao XF, Wang JX. 2016. A new subfamily of penaeidin with an additional serine-rich region from kuruma shrimp (*Marsupenaeus japonicus*) contributes to antimicrobial and phagocytic activities. *Developmental and Comparative Immunology* **59**:186–198. DOI: https://doi.org/10.1016/j.dci.2016.02.001, PMID: 26855016

Bachère E, Gueguen Y, Gonzalez M, de Lorgeril J, Garnier J, Romestand B. 2004. Insights into the anti-microbial defense of marine invertebrates: the penaeid shrimps and the oyster *Crassostrea gigas. Immunological Reviews* **198**:149–168. DOI: https://doi.org/10.1111/j.0105-2896.2004.00115.x, PMID: 15199961

Bauchau AG. 1981. Cruataceans. *Invertabrate Blood Cells* **2**:385–420.

Biočanin M, Bues J, Dainese R, Amstad E, Deplancke B. 2019. Simplified Drop-seq workflow with minimized bead loss using a bead capture and processing microfluidic chip. *Lab on a Chip* **19**:1610–1620. DOI: https://doi.org/10.1039/C9LC00014C, PMID: 30920557

Boonchuen P, Jaree P, Somboonviwat K, Somboonwiwat K. 2021. Regulation of shrimp prophenoloxidase activating system by lva-miR-4850 during bacterial infection. *Scientific Reports* **11**:3821. DOI: https://doi.org/10.1038/s41598-021-82881-2, PMID: 33589707

Broseus L, Thomas A, Oldfield AJ, Severac D, Dubois E, Ritchie W. 2020a. TALC: transcript-level aware Long-read correction. *Bioinformatics* **36**:5000–5006. DOI: https://doi.org/10.1093/bioinformatics/btaa634, PMID: 32910174

Broseus L, Thomas A, Oldfield AJ, Severac D, Dubois E, Ritchie W. 2020b. TALC: Transcription-Aware Long Read Correction. *GitLab.* v1.01.https://gitlab.igh.cnrs.fr/lbroseus/TALC

Bushmanova E, Antipov D, Lapidus A, Prjibelski AD. 2019a. rnaSPAdes: a *de novo* transcriptome assembler and its application to RNA-Seq data. *GigaScience* **8**:1–13. DOI: https://doi.org/10.1093/gigascience/giz100

Bushmanova E, Antipov D, Lapidus A, Prjibelski AD. 2019b. rnaSPAdes . *Center for Algorithmic Biotechnology, St.Petersburg State University.* v3.14.1. https://cab.spbu.ru/software/rnaspades/,

Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. 2018. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature Biotechnology* **36**:411–420. DOI: https://doi.org/10.1038/nbt.4096, PMID: 29608179

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* **10**:421. DOI: https://doi.org/10.1186/1471-2105-10-421, PMID: 20003500

Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. 2021. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *EMBL.* v2. http://eggnog-mapper.embl.de/

Carmona SJ, Teichmann SA, Ferreira L, Macaulay IC, Stubbington MJ, Cvejic A, Gfeller D. 2017. Single-cell transcriptome analysis of fish immune cells provides insight into the evolution of vertebrate immune cell types. *Genome Research* **27**:451–461. DOI: https://doi.org/10.1101/gr.207704.116, PMID: 28087841

Castillo J, Brown MR, Strand MR. 2011. Blood feeding and insulin-like peptide 3 stimulate proliferation of hemocytes in the mosquito *aedes aegypti. PLOS Pathogens* **7**:e1002274. DOI: https://doi.org/10.1371/journal.ppat.1002274, PMID: 21998579

Cattenoz PB, Sakr R, Pavlidaki A, Delaporte C, Riba A, Molina N, Hariharan N, Mukherjee T, Giangrande A. 2020. Temporal specificity and heterogeneity of *Drosophila* immune cells. *The EMBO Journal* **39**:e104486. DOI: https://doi.org/10.15252/embj.2020104486, PMID: 32162708

Chaikeeratisak V, Somboonwiwat K, Tassanakajon A. 2012. Shrimp alpha-2-macroglobulin prevents the bacterial escape by inhibiting fibrinolysis of blood clots. *PLOS ONE* **7**:e47384. DOI: https://doi.org/10.1371/journal.pone.0047384, PMID: 23082160

Charoensapsri W, Sangsuriya P, Lertwimol T, Gangnonngiw W, Phiwsaiya K, Senapin S. 2015. Laminin receptor protein is implicated in hemocyte homeostasis for the whiteleg shrimp *Penaeus (Litopenaeus) vannamei. Developmental & Comparative Immunology* **51**:39–47. DOI: https://doi.org/10.1016/j.dci.2015.02.012, PMID: 25720979

Chen MY, Hu KY, Huang CC, Song YL. 2005. More than one type of transglutaminase in invertebrates? A second type of transglutaminase is involved in shrimp coagulation. *Developmental & Comparative Immunology* **29**:1003–1016. DOI: https://doi.org/10.1016/j.dci.2005.03.012, PMID: 15985293

Cheung HC, Hai T, Zhu W, Baggerly KA, Tsavachidis S, Krahe R, Cote GJ. 2009. Splicing factors PTBP1 and PTBP2 promote proliferation and migration of glioma cell lines. *Brain* **132**:2277–2288. DOI: https://doi.org/10.1093/brain/awp153, PMID: 19506066

Cho B, Yoon S-H, Lee D, Koranteng F, Tattikota SG, Cha N, Shin M, Do H, Hu Y, Oh SY, Lee D, Vipin Menon A, Moon SJ, Perrimon N, Nam J-W, Shim J. 2020. Single-cell transcriptome maps of myeloid blood cell lineages in *Drosophila*. *Nature Communications* **11**:4483. DOI: https://doi.org/10.1038/s41467-020-18135-y

Dannenberg JH, David G, Zhong S, van der Torre J, Wong WH, Depinho RA. 2005. mSin3A corepressor regulates diverse transcriptional networks governing normal and neoplastic growth and survival. *Genes & Development* **19**:1581–1595. DOI: https://doi.org/10.1101/gad.1286905, PMID: 15998811

Dantas-Lima JJ, Tuan VV, Corteel M, Grauwet K, An NTT, Sorgeloos P, Nauwynck HJ. 2013. Separation of *Penaeus vannamei* haemocyte subpopulations by iodixanol density gradient centrifugation. *Aquaculture* **408-409**:128–135. DOI: https://doi.org/10.1016/j.aquaculture.2013.04.031

De Lucia F, Ni JQ, Vaillant C, Sun FL. 2005. HP1 modulates the transcription of cell-cycle regulators in *Drosophila melanogaster*. *Nucleic Acids Research* **33**:2852–2858. DOI: https://doi.org/10.1093/nar/gki584, PMID: 15905474

Destoumieux D, Bulet P, Loew D, Van Dorsselaer A, Rodriguez J, Bachère E. 1997. Penaeidins, a new family of antimicrobial peptides isolated from the shrimp *Penaeus vannamei* (Decapoda). *Journal of Biological Chemistry* **272**:28398–28406. DOI: https://doi.org/10.1074/jbc.272.45.28398, PMID: 9353298

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**:15–21. DOI: https://doi.org/10.1093/bioinformatics/bts635, PMID: 23104886

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2021. STAR 2.7.8a. *Github*. v2.7.8a. https://github.com/alexdobin/STAR

Elbahnaswy S, Koiwai K, Zaki VH, Shaheen AA, Kondo H, Hirono I. 2017. A novel viral responsive protein (MjVRP) from *Marsupenaeus japonicus* haemocytes is involved in white spot syndrome virus infection. *Fish & Shellfish Immunology* **70**:638–647. DOI: https://doi.org/10.1016/j.fsi.2017.09.045, PMID: 28935599

FAO. 2020. *Fishery and Aquaculture Statistics 2018/FAO Annuaire*. FAO.

Flaherty MS, Salis P, Evans CJ, Ekas LA, Marouf A, Zavadil J, Banerjee U, Bach EA. 2010. Chinmo is a functional effector of the JAK/STAT pathway that regulates eye development, tumor formation, and stem cell self-renewal in *Drosophila*. *Developmental Cell* **18**:556–568. DOI: https://doi.org/10.1016/j.devcel.2010.02.006, PMID: 20412771

Flegel TW. 2019. A future vision for disease control in shrimp aquaculture. *Journal of the World Aquaculture Society* **50**:249–266. DOI: https://doi.org/10.1111/jwas.12589

Fu Y, Huang X, Zhang P, van de Leemput J, Han Z. 2020. Single-cell RNA sequencing identifies novel cell types in *Drosophila* blood. *Journal of Genetics and Genomics* **47**:175–186. DOI: https://doi.org/10.1016/j.jgg.2020.02.004, PMID: 32487456

Gao J, Wang JX, Wang XW. 2019. MD-2 homologue recognizes the white spot syndrome virus lipid component and induces antiviral molecule expression in shrimp. *The Journal of Immunology* **203**:1131–1141. DOI: https://doi.org/10.4049/jimmunol.1900268, PMID: 31331974

Gilbert D. 2019. EvidentialGene. *Genome Informatics Lab of Indiana University Biology Department*. v2022.01.20. http://arthropods.eugenes.org/EvidentialGene/

Giulianini PG, Bierti M, Lorenzon S, Battistella S, Ferrero EA. 2007. Ultrastructural and functional characterization of circulating hemocytes from the freshwater crayfish *Astacus leptodactylus*: cell types and their role after in vivo artificial non-self challenge. *Micron* **38**:49–57. DOI: https://doi.org/10.1016/j.micron.2006.03.019, PMID: 16839768

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* **29**:644–652. DOI: https://doi.org/10.1038/nbt.1883, PMID: 21572440

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, et al. 2018. trinityrnaseq. *GitHub*. v1.01. https://github.com/trinityrnaseq/trinityrnaseq/

Hafemeister C, Satija R. 2019. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biology* **20**:296. DOI: https://doi.org/10.1186/s13059-019-1874-1, PMID: 31870423

He Y, Yao W, Liu P, Li J, Wang Q. 2018. Expression profiles of the p38 MAPK signaling pathway from chinese shrimp *Fenneropenaeus chinensis* in response to viral and bacterial infections. *Gene* **642**:381–388. DOI: https://doi.org/10.1016/j.gene.2017.11.050, PMID: 29155327

Hikima S, Hikima J, Rojtinnakorn J, Hirono I, Aoki T. 2003. Characterization and function of kuruma shrimp lysozyme possessing lytic activity against Vibrio species. *Gene* **316**:187–195. DOI: https://doi.org/10.1016/S0378-1119(03)00761-3, PMID: 14563565

Hipolito SG, Shitara A, Kondo H, Hirono I. 2014. Role of Marsupenaeus japonicus crustin-like peptide against Vibrio penaeicida and white spot syndrome virus infection. *Developmental & Comparative Immunology* **46**:461–469. DOI: https://doi.org/10.1016/j.dci.2014.06.001, PMID: 24929027

Huang CC, Sritunyalucksana K, Söderhäll K, Song YL. 2004. Molecular cloning and characterization of tiger shrimp (*Penaeus monodon*) transglutaminase. *Developmental & Comparative Immunology* **28**:279–294. DOI: https://doi.org/10.1016/j.dci.2003.08.005, PMID: 14698215

Hung MN, Shiomi R, Nozaki R, Kondo H, Hirono I. 2014. Identification of novel copper/zinc superoxide dismutase (Cu/ZnSOD) genes in Kuruma shrimp Marsupenaeus japonicus. *Fish & Shellfish Immunology* **40**:472–477. DOI: https://doi.org/10.1016/j.fsi.2014.07.030, PMID: 25107696

Jiao H, Dong P, Yan L, Yang Z, Lv X, Li Q, Zong X, Fan J, Fu X, Liu X, Xiao R. 2016. TGF-β1 induces polypyrimidine Tract-Binding protein to alter fibroblasts proliferation and fibronectin deposition in keloid. *Scientific Reports* **6**:38033. DOI: https://doi.org/10.1038/srep38033, PMID: 27897224

Jiravanichpaisal P, Lee BL, Söderhäll K. 2006. Cell-mediated immunity in arthropods: hematopoiesis, coagulation, melanization and opsonization. *Immunobiology* **211**:213–236. DOI: https://doi.org/10.1016/j.imbio.2005.10.015, PMID: 16697916

Johansson MW, Keyser P, Sritunyalucksana K, Söderhäll K. 2000. Crustacean haemocytes and haematopoiesis. *Aquaculture* **191**:45–52. DOI: https://doi.org/10.1016/S0044-8486(00)00418-X

Junkunlo K, Söderhäll K, Noonin C, Söderhäll I. 2017. PDGF/VEGF-Related receptor affects transglutaminase activity to control cell migration during crustacean hematopoiesis. *Stem Cells and Development* **26**:1449–1459. DOI: https://doi.org/10.1089/scd.2017.0086, PMID: 28805145

Junkunlo K, Söderhäll K, Söderhäll I. 2019. Transglutaminase inhibition stimulates hematopoiesis and reduces aggressive behavior of crayfish, *Pacifastacus leniusculus*. *Journal of Biological Chemistry* **294**:708–715. DOI: https://doi.org/10.1074/jbc.RA118.005489, PMID: 30425101

Junkunlo K, Söderhäll K, Söderhäll I. 2020. Transglutaminase 1 and 2 are localized in different blood cells in the freshwater crayfish *Pacifastacus leniusculus*. *Fish & Shellfish Immunology* **104**:83–91. DOI: https://doi.org/10.1016/j.fsi.2020.05.062, PMID: 32479868

Kim JM, Yamada M, Masai H. 2003. Functions of mammalian Cdc7 kinase in initiation/monitoring of DNA replication and development. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **532**:29–40. DOI: https://doi.org/10.1016/j.mrfmmm.2003.08.008, PMID: 14643427

Kirchner R, Barrera V. 2019. Tinyatlas. *GitHub*. 18d88df. https://github.com/hbc/tinyatlas

Koiwai K, Alenton RR, Shiomi R, Nozaki R, Kondo H, Hirono I. 2017. Two hemocyte sub-populations of kuruma shrimp *Marsupenaeus japonicus*. *Molecular Immunology* **85**:1–8. DOI: https://doi.org/10.1016/j.molimm.2017.01.024, PMID: 28167202

Koiwai K, Kondo H, Hirono I. 2018. RNA-seq identifies integrin alpha of kuruma shrimp *Marsupenaeus japonicus* as a candidate molecular marker for phagocytic hemocytes. *Developmental & Comparative Immunology* **81**:271–278. DOI: https://doi.org/10.1016/j.dci.2017.12.014, PMID: 29258750

Koiwai K, Kondo H, Hirono I. 2019. Isolation and molecular characterization of hemocyte sub-populations in Kuruma shrimp *Marsupenaeus japonicus*. *Fisheries Science* **85**:521–532. DOI: https://doi.org/10.1007/s12562-019-01311-5

Koiwai K. 2021a. Drop-seq_on_shrimp. *GitHub*. cddd130. https://github.com/KeiichiroKOIWAI/Drop-seq_on_shrimp

Koiwai K. 2021b. Drop-seqonshrimp. *Software Heritage*. swh:1:rev:cddd130b4cb841168bf3e29d0fbfef0de5ef2ad7. https://archive.softwareheritage.org/swh:1:rev:cddd130b4cb841168bf3e29d0fbfef0de5ef2ad7

Kondo M, Tomonaga S, Takahashi Y. 2012. Granulocytes with cytoplasmic deposits of kuruma prawn. *Aquaculture Science* **60**:151–152. DOI: https://doi.org/10.11233/aquaculturesci.60.151

Kotani E, Yamakawa M, Iwamoto S, Tashiro M, Mori H, Sumida M, Matsubara F, Taniai K, Kadono-Okuda K, Kato Y, Mori H. 1995. Cloning and expression of the gene of Hemocytin, an insect humoral lectin which is homologous with the mammalian von willebrand factor. *Biochimica Et Biophysica Acta (BBA) - Gene Structure and Expression* **1260**:245–258. DOI: https://doi.org/10.1016/0167-4781(94)00202-E

Lavine MD, Strand MR. 2002. Insect hemocytes and their role in immunity. *Insect Biochemistry and Molecular Biology* **32**:1295–1309. DOI: https://doi.org/10.1016/S0965-1748(02)00092-9, PMID: 12225920

Lei H, Kazlauskas A. 2009. Growth factors outside of the platelet-derived growth factor (PDGF) family employ reactive oxygen species/Src family kinases to activate PDGF receptor alpha and thereby promote proliferation and survival of cells. *Journal of Biological Chemistry* **284**:6329–6336. DOI: https://doi.org/10.1074/jbc.M808426200, PMID: 19126548

Li C, Chai J, Li H, Zuo H, Wang S, Qiu W, Weng S, He J, Xu X. 2014. Pellino protein from Pacific white shrimp *Litopenaeus vannamei* positively regulates NF-κB activation. *Developmental & Comparative Immunology* **44**:341–350. DOI: https://doi.org/10.1016/j.dci.2014.01.012, PMID: 24463313

Li C, Li H, Chen Y, Chen Y, Wang S, Weng SP, Xu X, He J. 2015. Activation of vago by interferon regulatory factor (IRF) suggests an interferon system-like antiviral mechanism in shrimp. *Scientific Reports* **5**:15078. DOI: https://doi.org/10.1038/srep15078, PMID: 26459861

Lin Y, Zhan W, Li Q, Zhang Z, Wei X, Sheng X. 2007. Ontogenesis of haemocytes in shrimp (*Fenneropenaeus chinensis*) studied with probes of monoclonal antibody. *Developmental & Comparative Immunology* **31**:1073–1081. DOI: https://doi.org/10.1016/j.dci.2007.02.001, PMID: 17428538

Lin X, Söderhäll K, Söderhäll I. 2008. Transglutaminase activity in the hematopoietic tissue of a crustacean, *Pacifastacus leniusculus*, importance in hemocyte homeostasis. *BMC Immunology* **9**:58. DOI: https://doi.org/10.1186/1471-2172-9-58

Lin X, Söderhäll K, Söderhäll I. 2011. Invertebrate hematopoiesis: an astakine-dependent novel hematopoietic factor. *The Journal of Immunology* **186**:2073–2079. DOI: https://doi.org/10.4049/jimmunol.1001229, PMID: 21220699

Lin X, Söderhäll I. 2011. Crustacean hematopoiesis and the astakine cytokines. *Blood* **117**:6417–6424. DOI: https://doi.org/10.1182/blood-2010-11-320614

Liu H, Jiravanichpaisal P, Söderhäll I, Cerenius L, Söderhäll K. 2006. Antilipopolysaccharide factor interferes with white spot syndrome virus replication in vitro and in vivo in the crayfish *Pacifastacus leniusculus*. *Journal of Virology* **80**:10365–10371. DOI: https://doi.org/10.1128/JVI.01101-06, PMID: 17041217

Liu H, Wang J, Mao Y, Liu M, Niu S, Qiao Y, Su Y, Wang C, Zheng Z. 2015. Identification and expression analysis of a novel stylicin antimicrobial peptide from kuruma shrimp (Marsupenaeus japonicus). *Fish & Shellfish Immunology* **47**:817–823. DOI: https://doi.org/10.1016/j.fsi.2015.09.044

Liu S, Zheng SC, Li YL, Li J, Liu HP. 2020. Hemocyte-Mediated phagocytosis in crustaceans. *Frontiers in Immunology* **11**:268. DOI: https://doi.org/10.3389/fimmu.2020.00268, PMID: 32194551

Luo Z, Huang W, Wang G, Sun H, Chen X, Luo P, Liu J, Hu C, Li H, Shu H. 2020. Identification and characterization of p38MAPK in response to acute cold stress in the gill of Pacific white shrimp (*Litopenaeus vannamei*). *Aquaculture Reports* **17**:100365. DOI: https://doi.org/10.1016/j.aqrep.2020.100365

Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, Trombetta JJ, Weitz DA, Sanes JR, Shalek AK, Regev A, McCarroll SA. 2015. Highly parallel Genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**:1202–1214. DOI: https://doi.org/10.1016/j.cell.2015.05.002, PMID: 26000488

Martínez R, Carpio Y, Arenal A, Lugo JM, Morales R, Martín L, Rodríguez RF, Acosta J, Morales A, Duconge J, Estrada MP. 2017. Significant improvement of shrimp growth performance by growth hormone-releasing peptide-6 immersion treatments. *Aquaculture Research* **48**:4632–4645. DOI: https://doi.org/10.1111/are.13286

McDonel P, Demmers J, Tan DW, Watt F, Hendrich BD. 2012. Sin3a is essential for the genome integrity and viability of pluripotent cells. *Developmental Biology* **363**:62–73. DOI: https://doi.org/10.1016/j.ydbio.2011.12.019, PMID: 22206758

Meister M, Lagueux M. 2003. *Drosophila* blood cells. *Cellular Microbiology* **5**:573–580. DOI: https://doi.org/10.1046/j.1462-5822.2003.00302.x, PMID: 12925127

Mekata T, Sudhakaran R, Okugawa S, Kono T, Sakai M, Itami T. 2010. Molecular cloning and transcriptional analysis of a newly identified anti-lipopolysaccharide factor gene in Kuruma shrimp, *Marsupenaeus japonicus*. *Letters in Applied Microbiology* **50**:112–119. DOI: https://doi.org/10.1111/j.1472-765X.2009.02763.x

Mohan M, Bartkuhn M, Herold M, Philippen A, Heinl N, Bardenhagen I, Leers J, White RAH, Renkawitz-Pohl R, Saumweber H, Renkawitz R. 2007. The *Drosophila* insulator proteins CTCF and CP190 link enhancer blocking to body patterning. *The EMBO Journal* **26**:4203–4214. DOI: https://doi.org/10.1038/sj.emboj.7601851

Monzón-Casanova E, Matheson LS, Tabbada K, Zarnack K, Smith CW, Turner M. 2020. Polypyrimidine tract-binding proteins are essential for B cell development. *eLife* **9**:e53557. DOI: https://doi.org/10.7554/eLife.53557, PMID: 32081131

Munier AI, Doucet D, Perrodou E, Zachary D, Meister M, Hoffmann JA, Janeway CA, Lagueux M. 2002. PVF2, a PDGF/VEGF-like growth factor, induces hemocyte proliferation in *Drosophila* larvae. *EMBO Reports* **3**:1195–1200. DOI: https://doi.org/10.1093/embo-reports/kvf242, PMID: 12446570

Neufeld G, Cohen T, Gengrinovitch S, Poltorak Z. 1999. Vascular endothelial growth factor (VEGF) and its receptors. *The FASEB Journal* **13**:9–22. DOI: https://doi.org/10.1096/fasebj.13.1.9, PMID: 9872925

Ning P, Zheng Z, Aweya JJ, Yao D, Li S, Ma H, Wang F, Zhang Y. 2018. *Litopenaeus vannamei* notch affects lipopolysaccharides induced reactive oxygen species. *Developmental & Comparative Immunology* **81**:74–82. DOI: https://doi.org/10.1016/j.dci.2017.11.010, PMID: 29155012

Noonin C, Lin X, Jiravanichpaisal P, Söderhäll K, Söderhäll I. 2012. Invertebrate hematopoiesis: an anterior proliferation center as a link between the hematopoietic tissue and the brain. *Stem Cells and Development* **21**:3173–3186. DOI: https://doi.org/10.1089/scd.2012.0077, PMID: 22564088

Paro R, Hogness DS. 1991. The polycomb protein shares a homologous domain with a heterochromatin-associated protein of *Drosophila*. *PNAS* **88**:263–267. DOI: https://doi.org/10.1073/pnas.88.1.263, PMID: 1898775

Peichev M, Naiyer AJ, Pereira D, Zhu Z, Lane WJ, Williams M, Oz MC, Hicklin DJ, Witte L, Moore MA, Rafii S. 2000. Expression of VEGFR-2 and AC133 by circulating human CD34(+) cells identifies a population of functional endothelial precursors. *Blood* **95**:952–958. DOI: https://doi.org/10.1182/blood.V95.3.952.003k27_952_958, PMID: 10648408

Picard Toolkit. 2019. Broad Institute. *GitHub*. 2.0.1. https://github.com/broadinstitute/picard

Picelli S, Björklund AK, Reinius B, Sagasser S, Winberg G, Sandberg R. 2014. Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Research* **24**:2033–2040. DOI: https://doi.org/10.1101/gr.177881.114, PMID: 25079858

Raddi G, Barletta ABF, Efremova M, Ramirez JL, Cantera R, Teichmann SA, Barillas-Mury C, Billker O. 2020. Mosquito cellular immunity at single-cell resolution. *Science* **369**:1128–1132. DOI: https://doi.org/10.1126/science.abc0322, PMID: 32855340

Rasko JE, Klenova EM, Leon J, Filippova GN, Loukinov DI, Vatolin S, Robinson AF, Hu YJ, Ulmer J, Ward MD, Pugacheva EM, Neiman PE, Morse HC, Collins SJ, Lobanenkov VV. 2001. Cell growth inhibition by the multifunctional multivalent zinc-finger factor CTCF. *Cancer Research* **61**:6002–6007. PMID: 11507042

Rodriguez J, Boulo V, Mialhe E, Bachere E. 1995. Characterisation of shrimp haemocytes and plasma components by monoclonal antibodies. *Journal of Cell Science* **108**:1043–1050. DOI: https://doi.org/10.1242/jcs.108.3.1043, PMID: 7622592

Rosa RD, Vergnes A, de Lorgeril J, Goncalves P, Perazzolo LM, Sauné L, Romestand B, Fievet J, Gueguen Y, Bachère E, Destoumieux-Garzón D. 2013. Functional divergence in shrimp anti-lipopolysaccharide factors (ALFs): from recognition of cell wall components to antimicrobial activity. *PLOS ONE* **8**:e67937. DOI: https://doi.org/10.1371/journal.pone.0067937, PMID: 23861837

Rosa RD, Barracco MA. 2010. Antimicrobial peptides in crustaceans. *Invertebrate Survival Journal* **7**:262–284.

Rota-Stabelli O, Daley AC, Pisani D. 2013. Molecular timetrees reveal a cambrian colonization of land and a new scenario for ecdysozoan evolution. *Current Biology* **23**:392–398. DOI: https://doi.org/10.1016/j.cub.2013.01.026, PMID: 23375891

Scott CL, Gil J, Hernando E, Teruya-Feldstein J, Narita M, Martínez D, Visakorpi T, Mu D, Cordon-Cardo C, Peters G, Beach D, Lowe SW. 2007. Role of the chromobox protein CBX7 in lymphomagenesis. *PNAS* **104**: 5389–5394. DOI: https://doi.org/10.1073/pnas.0608721104, PMID: 17374722

Seppey M, Manni M, Zdobnov EM. 1962. BUSCO: assessing genome assembly and annotation completeness. *Methods in Molecular Biology* **2019**:227–245.

Sequeira T, Tavares D, Arala-Chaves M. 1996. Evidence for circulating hemocyte proliferation in the shrimp *Penaeus japonicus*. *Developmental & Comparative Immunology* **20**:97–104. DOI: https://doi.org/10.1016/0145-305X(96)00001-8, PMID: 8799615

Shibayama M, Ohno S, Osaka T, Sakamoto R, Tokunaga A, Nakatake Y, Sato M, Yoshida N. 2009. Polypyrimidine tract-binding protein is essential for early mouse development and embryonic stem cell proliferation. *FEBS Journal* **276**:6658–6668. DOI: https://doi.org/10.1111/j.1742-4658.2009.07380.x, PMID: 1 9843185

Smith RG, Leonard R, Bailey ART, Palyha O, Feighner S, Tan C, McKee KK, Pong S-S, Griffin P, Howard A. 2001. Growth hormone secretagogue receptor family members and ligands. *Endocrine* **14**:009–014. DOI: https://doi.org/10.1385/ENDO:14:1:009

Söderhäll I, Kim YA, Jiravanichpaisal P, Lee SY, Söderhäll K. 2005. An ancient role for a prokineticin domain in invertebrate hematopoiesis. *The Journal of Immunology* **174**:6153–6160. DOI: https://doi.org/10.4049/jimmunol.174.10.6153, PMID: 15879111

Söderhäll I. 2013. Recent advances in crayfish hematopoietic stem cell culture: a model for studies of hemocyte differentiation and immunity. *Cytotechnology* **65**:691–695. DOI: https://doi.org/10.1007/s10616-013-9578-y, PMID: 23686548

Söderhäll I. 2016. Crustacean hematopoiesis. *Developmental & Comparative Immunology* **58**:129–141. DOI: https://doi.org/10.1016/j.dci.2015.12.009, PMID: 26721583

Söderhäll K, Smith VJ. 1983. Separation of the haemocyte populations of *Carcinus maenas* and other marine decapods, and prophenoloxidase distribution. *Developmental & Comparative Immunology* **7**:229–239. DOI: https://doi.org/10.1016/0145-305X(83)90004-6, PMID: 6409683

Soneson C, Robinson MD. 2018. Bias, robustness and scalability in single-cell differential expression analysis. *Nature Methods* **15**:255–261. DOI: https://doi.org/10.1038/nmeth.4612, PMID: 29481549

Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, Hao Y, Stoeckius M, Smibert P, Satija R. 2019. Comprehensive integration of Single-Cell data. *Cell* **177**:1888–1902. DOI: https://doi.org/10.1016/j.cell.2019.05.031, PMID: 31178118

Su TT, Parry DH, Donahoe B, Chien CT, O'Farrell PH, Purdy A. 2001. Cell cycle roles for two 14-3-3 proteins during *Drosophila* development. *Journal of Cell Science* **114**:3445–3454. PMID: 11682604

Sung H-H, Wu P-Y, Song Y-L. 1999. Characterisation of monoclonal antibodies to haemocyte subpopulations of tiger shrimp (*Penaeus monodon*): immunochemical differentiation of three major haemocyte types. *Fish & Shellfish Immunology* **9**:167–179. DOI: https://doi.org/10.1006/fsim.1998.0185

Sung HH, Sun R. 2002. Use of monoclonal antibodies to classify hemocyte subpopulations of tiger shrimp (*Penaeus monodon*). *Journal of Crustacean Biology* **22**:337–344. DOI: https://doi.org/10.1163/20021975-99990240

Tassanakajon A, Somboonwiwat K, Supungul P, Tang S. 2013. Discovery of immune molecules and their crucial functions in shrimp immunity. *Fish & Shellfish Immunology* **34**:954–967. DOI: https://doi.org/10.1016/j.fsi.2012.09.021, PMID: 23059654

Tattikota SG, Cho B, Liu Y, Hu Y, Barrera V, Steinbaugh MJ, Yoon SH, Comjean A, Li F, Dervis F, Hung RJ, Nam JW, Ho Sui S, Shim J, Perrimon N. 2020. A single-cell survey of *Drosophila* blood. *eLife* **9**:e54818. DOI: https://doi.org/10.7554/eLife.54818, PMID: 32396065

Thomas GWC, Dohmen E, Hughes DST, Murali SC, Poelchau M, Glastad K, Anstead CA, Ayoub NA, Batterham P, Bellair M, Binford GJ, Chao H, Chen YH, Childers C, Dinh H, Doddapaneni HV, Duan JJ, Dugan S, Esposito LA, Friedrich M, et al. 2020. Gene content evolution in the arthropods. *Genome Biology* **21**:15. DOI: https://doi.org/10.1186/s13059-019-1925-7, PMID: 31969194

Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, Rinn JL. 2014. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature Biotechnology* **32**:381–386. DOI: https://doi.org/10.1038/nbt.2859, PMID: 24658644

Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, Rinn JL, Qiu X, Pliner H. 2021. MONOCLE 3. *GitHub*. v0.2.3.0. https://github.com/cole-trapnell-lab/monocle3

van de Braak CB, Taverne N, Botterblom MH, van der Knaap WP, Rombout JH. 2000. Characterisation of different morphological features of black tiger shrimp (*Penaeus monodon*) haemocytes using monoclonal antibodies. *Fish & Shellfish Immunology* **10**:515–530. DOI: https://doi.org/10.1006/fsim.2000.0269, PMID: 11016586

van de Braak CB, Botterblom MH, Liu W, Taverne N, van der Knaap WP, Rombout JH. 2002. The role of the haematopoietic tissue in haemocyte production and maturation in the black tiger shrimp (*Penaeus monodon*). *Fish & Shellfish Immunology* **12**:253–272. DOI: https://doi.org/10.1006/fsim.2001.0369, PMID: 11931020

Varma Shrivastav S, Bhardwaj A, Pathak KA, Shrivastav A. 2020. Insulin-Like growth factor binding Protein-3 (IGFBP-3): Unraveling the role in mediating IGF-Independent effects within the cell. *Frontiers in Cell and Developmental Biology* **8**:286. DOI: https://doi.org/10.3389/fcell.2020.00286, PMID: 32478064

Ventura-López C, Galindo-Torres PE, Arcos FG, Galindo-Sánchez C, Racotta IS, Escobedo-Fregoso C, Llera-Herrera R, Ibarra AM. 2017. Transcriptomic information from Pacific white shrimp (*Litopenaeus vannamei*) ovary and Eyestalk, and expression patterns for genes putatively involved in the reproductive process. *General and Comparative Endocrinology* **246**:164–182. DOI: https://doi.org/10.1016/j.ygcen.2016.12.005, PMID: 27964922

Wang G, Li N, Zhang L, Zhang L, Zhang Z, Wang Y. 2015. IGFBP7 promotes hemocyte proliferation in small abalone *Haliotis diversicolor*, proved by dsRNA and cap mRNA exposure. *Gene* **571**:65–70. DOI: https://doi.org/10.1016/j.gene.2015.06.051, PMID: 26115770

Winotaphan P, Sithigorngul P, Muenpol O, Longyant S, Rukpratanporn S, Chaivisuthangkura P, Sithigorngul W, Petsom A, Menasveta P. 2005. Monoclonal antibodies specific to haemocytes of black tiger prawn *Penaeus monodon*. *Fish & Shellfish Immunology* **18**:189–198. DOI: https://doi.org/10.1016/j.fsi.2004.05.007, PMID: 15519539

Wolfe JM, Breinholt JW, Crandall KA, Lemmon AR, Lemmon EM, Timm LE. 1901. A phylogenomic framework, evolutionary timeline and genomic resources for comparative studies of decapod crustaceans. *Proceedings. Biological Sciences* **2019**:20190079. DOI: https://doi.org/10.1098/rspb.2019.0079

Wysoker A, Nemesh J, Kashin S, Gould J, cpudan, Wendt J. 2020. Drop-seq. *GitHub.* 2.3.0. https://github.com/broadinstitute/Drop-seq

Xing J, Chang Y, Tang X, Sheng X, Zhan W. 2017. Separation of haemocyte subpopulations in shrimp *Fenneropenaeus chinensis* by immunomagnetic bead using monoclonal antibody against granulocytes. *Fish & Shellfish Immunology* **60**:114–118. DOI: https://doi.org/10.1016/j.fsi.2016.11.034, PMID: 27847341

Yang L, Liu X, Huang J, Yang Q, Qiu L, Liu W, Jiang S. 2012. Molecular characterization and expression profile of MAP2K1ip1/MP1 gene from tiger shrimp, *Penaeus monodon*. *Molecular Biology Reports* **39**:5811–5818. DOI: https://doi.org/10.1007/s11033-011-1391-0, PMID: 22209950

Yeh M-S, Kao L-R, Huang C-J, Tsai I-H. 2006. Biochemical characterization and cloning of transglutaminases responsible for hemolymph clotting in Penaeus monodon and Marsupenaeus japonicus. *Biochimica Et Biophysica Acta (BBA) - Proteins and Proteomics* **1764**:1167–1178. DOI: https://doi.org/10.1016/j.bbapap.2006.04.005

Zettervall CJ, Anderl I, Williams MJ, Palmer R, Kurucz E, Ando I, Hultmark D. 2004. A directed screen for genes involved in *Drosophila* blood cell activation. *PNAS* **101**:14192–14197. DOI: https://doi.org/10.1073/pnas.0403789101, PMID: 15381778

Zhang X, Yuan J, Sun Y, Li S, Gao Y, Yu Y, Liu C, Wang Q, Lv X, Zhang X, Ma KY, Wang X, Lin W, Wang L, Zhu X, Zhang C, Zhang J, Jin S, Yu K, Kong J, et al. 2019. Penaeid shrimp genome provides insights into benthic adaptation and frequent molting. *Nature Communications* **10**:356. DOI: https://doi.org/10.1038/s41467-018-08197-4, PMID: 30664654

Zhang H, Cheng W, Zheng J, Wang P, Liu Q, Li Z, Shi T, Zhou Y, Mao Y, Yu X. 2020. Identification and molecular characterization of a pellino protein in Kuruma prawn (*Marsupenaeus japonicus*) in response to white spot syndrome virus and Vibrio parahaemolyticus infection. *International Journal of Molecular Sciences* **21**:1243. DOI: https://doi.org/10.3390/ijms21041243

Zhao YZ, Chen XL, Zeng DG, Yang CL, Peng M, Chen XH. 2015. Molecular cloning, characterization, and expression of Rab5B, Rab6A, and Rab7 from *Litopenaeus vannamei* (Penaeidae). *Genetics and Molecular Research* **14**:7740–7750. DOI: https://doi.org/10.4238/2015.July.13.20, PMID: 26214455

Zimmermann HB, Crusoe MR, MacManes M, Plessy C. 2018. TransDecoder. *GitHub.* 5.5.0. https://github.com/TransDecoder/TransDecoder