bioRxiv preprint doi: https://doi.org/10.1101/2022.05.28.493830; this version posted May 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Single-cell perspectives on the function and development of deep-sea mussel bacteriocytes

Hao Chen[1,3,4], Mengna Li[1,3,4,5], Minxiao Wang[1,3,4,*], Zhaoshan Zhong[1,3,4], Chao Lian[1,3,4], Li Zhou[1,3,4], Huan Zhang[1,3,4], Hao Wang[1,3,4], Lei Cao[1,3,4],Chaolun Li[1,2,3,4,5,*]

[1] Center of Deep Sea Research, and CAS Key Laboratory of Marine Ecology and Environmental Sciences, Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China

[2] South China Sea Institute of Oceanology, Chinese Academy of Sciences, Guangzhou 510301, China

[3] Laboratory for Marine Ecology and Environmental Science, Qingdao National Laboratory for Marine Science and Technology, Qingdao 266071, China

[4] Center for Ocean Mega-Science, Chinese Academy of Sciences, Qingdao 266071, China

[5] University of Chinese Academy of Sciences, Beijing 10049, China

*Correspondence and requests for materials should be addressed to M.W. (email: wangminxiao@qdio.ac.cn) and C.L. (email: lcl@qdio.ac.cn)

WITHDRAWN
see manuscript DOI for details

Abstract

Symbiosis is known to be a major force driving the adaptation and evolution of multicellular organisms. The symbiotic cells and organs, on the meantime, directly provide the mutualist services to the host as the source of phenotypic complexity and outcome of development plasticity. The investigations on the formation and development of symbiotic cells and organs however still remain a challenge in both model and non-model holobionts. Here, by constructing the high-resolution single-cell expression atlas of gill tissue, we have thoroughly surveyed the population census and function atlas of bacteriocytes in deep-sea mussel. Results showed that the bacteriocytes have markedly reshaped its metabolism and are highly coordinated with the endosymbionts in the metabolism of sterol, carbohydrate and ammonia. The

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.28.493830; this version posted May 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

immune process of bacterioytes is also robustly adjusted to facilitate the recognition, engulfment, elimination and transcellular transport of endosymbionts. Noticeably, we recognized that the molecular function of bacteriocytes could be guided synergistically by co-option of several conserved transcription factors of the regulatory networks. Moreover, these conserved transcription factors were also responsible for the differentiation and maturation of bacteriocyte lineages. Our results have for the first time revealed the coordination in the function and development of deep-sea mussel symbiotic cells, which was mediated by an ancestral and intrinsic toolkit that conserved across mollusk and could be influenced by symbionts. The coordination while greatly improved our knowledge on how deep-sea mussel thrive in the extreme environment of deep-sea, also highlighted a promising roadmap in revealing the formation and development of symbiotic cells and organs in all holobionts.

WITHDRAWN
see manuscript DOI for details

Keywords

Host-microbe interaction, spatial transcriptome, single-cell transcriptome, function atlas, development trajectory

Background

Symbiosis is known to be a major force in driving the evolution of eukaryotic organisms, entailing the acquisition of novel adaptive traits, advantaging the expansion of ecological range and thereafter shaping the biodiversity of life (1). To establish symbiotic associations, the host often evolved with highly complex symbiotic cells or organs to harbor the symbionts and provide the mutualist services, which is regarded as the source of phenotypic complexity and outcome of development plasticity (2). The characterization of formation and development of symbiotic cells and organs is therefore crucial in understanding the adaptation and evolution of symbiotic associations. However, while researchers are now able to take a glance at the function and symbiotic interactions within the symbiotic cells or organs, characterizations on the formation and development process is still a challenge in both model and non-model holobionts (3). A more intriguing question

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.28.493830; this version posted May 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

whether and how the symbionts shape the symbiotic cells and organs, which seems to be debatable and might vary greatly across species (4-6).

Being the second most species-rich phylum, mollusk has been found thriving in a wide range of habitats including fresh water, intertidal region and even the deep-sea. While the mollusk are known to possess a vast diversity of body plans, they have also formed diverse and intimate symbiotic association with microbes in adaptation to the harsh environments such as the deep-sea cold seeps and hydrothermal vents (7). In addition, the symbiotic associations of mollusk are quite diverse given to the symbiont location (extracellular or intracellular), type (methanotroph, thiotroph or both) and transmission mode (vertical or horizontal) (8). More noticeably, the symbiotic association could even be lost in some mollusk during tens of million years when organic matters were more accessible (9). The diverse and dynamic traits of symbiosis in mollusk therefore make them a promising model in investigating the formation and development of symbiotic cells and organs as well as the participation of symbionts in such process. Moreover, regardless the diverse forms of symbiotic associations, the symbiotic mollusk were yet estimated to be diverged from the non-symbiotic relatives only a few hundred million years ago and then evolved independently into different species (10). The rapid emergence and divergence of symbiotic association in mollusk have brought a more vital question: are the formation and development of symbiotic cells and organs accomplished by the evolutionary novel genes or co-option of conserved genes. Notwithstanding, the question has long been overlooked due to sufficient information.

As the endemic species in the chemosynthetic ecosystems of cold seep and hydrothermal vent, the deep-sea mussels are featured by the endosymbiotic methanotrophs and/or thiotrophs colonized in their gill bacteriocytes (10). The deep-sea mussels are estimated to be diverged from their non-symbiotic shallow-water relatives at 110 million years ago and could acquire the symbionts horizontally from the local environment all the life span, therefore representing a model species of the endosymbiotic association in mollusk (11). Here, we use the high-resolution single-cell transcriptome and meta-transcriptome data to survey the

WITHDRAWN
see manuscript DOI for details

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.28.493830; this version posted May 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

population census, function atlas and development trajectory of bacteriocytes in the methanotrophic deep-sea mussel *Gigantidas platifrons* and compare with data from the decolonized mussel to clarify the potential influence of symbionts in the function and development of the symbiotic cells. Finally, with the comparison with data from other mollusk, we identified and tested whether the toolkit guiding above process is universal and conserved in mollusk.

Results

Population census and function landscape of bacteriocyte lineages

To investigate the function and development of bacteriocytes, a single-cell resolution of bacteriocyte transcriptome is of necessity. In this study, we used the combination of single-cell transcriptome (scRNA-seq, 10× Genomics Chromium platform) and spatial transcriptomics (ST-seq, 10× Genomics Visium platform) to characterize the molecular atlas of bacteriocytes (Fig. 1A, see "Methods" section). Two deep-sea mussel samples were employed for the assay, among which one was collected from the seepage region (InS group) as indicator of the fully-symbiotic state while the rest one was partially decolonized *in situ* at a low methane region for 604 day (DeC group). To advantage the taxonomy of all gill cells, especially the bacteriocytes, the cross-sectioned middle part of gill (ventral view) was subjected to ST-seq while the dorsal part of gill containing both descending and ascending gill filaments was subjected to scRNA-seq (using the cell nuclei). As a result, over 3,600 barcoded spots with 172,705 mean reads per spot and about 26,707 cells with 37,585 mean reads per cell were obtained for two samples of the ST-seq and scRNA-seq corresponding (Supplementary Table. 1). By profiling the transcripts coverage, it came to our notice that the sequencing depth of both ST-seq and scRNA-seq had saturated all samples in the assay (Supplementary Fig. 1), guaranteeing the reliability of subsequent analysis.

Considering the limited knowledge in the marker genes of mussel gill cells, a morphology-based strategy was employed for the taxonomy of bacteriocyte lineages. As observed in t-distributed Stochastic Neighbor Embedding (t-SNE) result, a total of 9 cell clusters were obtained in the ST-seq data while two types of bacteriocytes

WITHDRAWN
see manuscript DOI for details

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.28.493830; this version posted May 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(cluster 0 and cluster 1) were identified given to the hematoxylin eosin staining and *in situ hybridization* results after projecting all cell clusters on the gill tissue (Fig. 1B, Supplementary Fig. 2). While the ST-seq results have provided necessary information on the identification of bacteriocytes, it came to our notice that the spot size of ST-seq (55 μm in diameter) is much greater than that of majority of gill cells, resulting in a compromised sampling resolution and crude result of taxonomy. On the other hand, we identified a total of 14 cell clusters by the scRNA-seq using tSNE method (Fig. 1C). We therefore combined the scRNA-seq and ST-seq data using Seurat-v3.2 anchor based integration method to improve the resolution of cell taxonomy. Consequently, we found that scRNA-seq data could greatly improve the resolution of ST-seq while the bacteriocytes were now further divided into three different lineages (cluster 1, 8 and 11 of scRNA-seq). By projecting all bacteriocytes on the gill tissue, we found that the cluster 8 bacteriocytes were exclusively distributed at the non-ciliated abfrontal edge of gill filaments while cluster 11 cells were interspersed between cluster 1 and cluster 8 (Fig. 1C, Supplementary Fig. 2). Considering that cells at the non-ciliated abfrontal edge of gill filaments (cluster 8) could bridge the symbiont colonization from mature bacteriocytes to younger gill cells of adjacent gill filament, we therefore termed the cluster 1 bacteriocyte as mature bacteriocyte, cluster 11 as immature bacteriocyte and cluster 8 as naïve bacteriocyte, which occupied only about 27.09%, 1.61% and 1.69% of total gill cells in fully-symbiotic mussels correspondingly (Supplementary Fig. 2).

Besides the distinct anatomical distribution pattern, the three bacteriocyte lineages are also distinguishable at the molecular level, especially in their marker genes (Fig. 1D, Supplementary Table. 2). The mature, immature and naïve bacteriocytes are found transcribing a total of 199, 57 and 82 candidate marker genes (expressed in more than 25% of target cell cluster with at least 2-fold up-regulation than that in other cell cluster, *p*-value less than 0.01) correspondingly. Moreover, a total of 67 marker genes such as excitatory amino acid transporter 1 (SLC1A3), monocarboxylate transporter 13 (SLC16A14), Cathepsin-L (CTL), acid phosphatase type 7 (ACP7) and myoneurin (MYNN), were found constantly expressed in the

WITHDRAWN
see manuscript DOI for details

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.28.493830; this version posted May 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

mature bacteriocytes regardless the fluctuation of symbiont abundance. Meanwhile, about 132 marker genes including solute carrier family 40 member 1 (SLC40A1), ammonium transporter Rh type A (RHBG-A), solute carrier family 26 member 10 (SLC26A2), 24-hydroxycholesterol 7 alpha-hydroxylase (CYP39A1), sugar phosphate exchanger 2 (SLC37A2), zinc finger protein 271 (ZNF271) and ETS-related transcription factor (EHF) could be down-regulated in response to the symbiont decolonization (Fig. 2A, Supplementary Table. 3). Contrarily, the immature and naïve bacteriocytes were found encoding more constant markers. For example, about 50/57 and 44/82 marker genes were permanently expressed in the immature and naïve bacteriocytes correspondingly (Fig. 2B&C, Supplementary Table. 3). By validation via *in situ hybridization* assay (Fig. 2D), we noticed that the mature bacteriocytes were abundantly expressing metabolic genes while the immature and naïve bacteriocytes were massively transcribing signal transducers or transcription/translation regulators, indicating their different roles in symbiosis.

WITHDRAWN
see manuscript DOI for details

In support of our speculation, function enrichment analysis of the marker genes showed that the mature bacteriocytes were massively encoding genes involved in metabolic processes (such as the biosynthesis and transport of carbohydrate, lipid, amino acid and vitamin) and symbiosis-related immune response (Fig. 2E). Comparatively, the immature bacteriocytes has greatly up-regulated cell differentiation and development-related genes (Fig. 2E). As for the naïve bacteriocyte, genes involved in symbiosis-related immune response and cell development (such as cellular component assembly, Notch signaling pathway, and mitotic cell cycle) were greatly up-regulated (Fig. 2E). Considering the population census of different lineages, it's reasonable to conclude that the mature bacteriocytes were the major facility to acquire nutrition and maintain symbiosis for host while the immature and naïve bacteriocytes are paying more attention to cell development and establishment of symbiosis.

Metabolic coordination between the host and symbionts

The chemosynthetic endosymbionts are known to be the major nutrition source for the mussel host, providing more than 90% of total nutrition including sterol

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.28.493830; this version posted May 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

intermediates, essential amino acids, vitamins and some unknown carbohydrates (12). While the metabolism of bacteriocytes is speculated to be rigorously adjusted in adaptation to the symbiotic association, how, and to what extent, has the host reshaped the metabolism yet is largely undiscovered. While the question could be easily addressed by searching the biosynthetic genes and/or transporter genes that transcribed by bacteriocytes and symbionts in a complementary way, such information is a challenge to obtain in non-model holobionts where symbionts were nonculturable and only bulk-seq method is available. With the single-cell atlas of bacteriocytes and the meta-transcriptome data of endosymbionts from the same sample (Fig. 3A), we are now able to survey the metabolic coordination within the deep-sea mussel holobionts. We first focused on the sterol metabolism of both parters, which is a well-known example of the metabolic interaction learned from genome sequencing. We noticed that the mature bacteriocytes are transcribing 24-hydroxycholesterol 7 alpha-hydroxylase (EC 1.14.14.26) gene robustly under normal conditions, suggesting a high turn-over of sterol metabolism and heavy demand of intermediates such as presqualene-PP, squalene, (S)-squalene-2,3-epoxide and 4,4-dimethyl-cholesta-8,14,24-trienol. While the enzymes are absent in the host genome, we noticed that the methanotrophic endosymbionts are cable of transcribing these enzymes (EC2.5.1.21, EC1.14.1417, EC1.14.14154 and EC1.14.1536) in complementation with the host (Fig. 3B, Supplementary Fig. 2AB, Supplementary Table. 4). This finding, while reconfirmed that symbionts could provide sterol intermediates to host (13), certificated the reliability of scRNA-seq in studying the metabolic coordination in deep-sea mussels.

We then focused on the carbohydrate metabolism to identify any possible interactions between both partners. While it is believed that the endosymbionts could provide the host with carbohydrates to meet the nutrition need in the oligotrophic habitats, what and how the carbohydrates were exchanged remained completely unknown. With the meta-transcriptome data of endosymbionts, we noted that genes involved in gluconeogenesis and glycogen biosynthesis could be massively transcribed (Fig. 3B, Supplementary Fig. 2C, Supplementary Table. 4), suggesting the

WITHDRAWN
see manuscript DOI for details

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.28.493830; this version posted May 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

robust production of glycogen in preparation to the supply. However, we also noticed that multiple enzymes in the gluconeogenesis pathway of endosymbionts were either insufficiently expressed (EC5.3.1.9) or completely absent in genome (EC3.1.3.11, EC2.7.1.11 and EC4.1.2.13), resulting in the overproduction of fructose-6P and deficiency of glucose-6P, fructose-1,6P_2, glyceraldehyde-3P and glucose. On the other hand, we found that all these enzymes could be encoded by the mature bacteriocytes while a sugar phosphate exchanger gene was massively transcribed simultaneously in mature bacteriocytes, which is responsible for the transmembrane transport of fructose-6P, glucose-6P, fructose-1,6P_2 or glyceraldehyde-3P (Fig. 3B, Supplementary Fig. 2D). The complementation in carbohydrate metabolism strongly suggested that the symbionts could also supply fructose-6P directly to host in exchange for intermediates of the gluconeogenesis. While the lack of gluconeogenesis-related enzymes is not detrimental to the symbiont, the coordination of both partners in exchange for sugar phosphate yet could benefit host greatly by providing a more efficient and diverse way to acquire carbohydrates.

WITHDRAWN
see manuscript DOI for details

When examining the TEM images of bacteriocytes, it is clearly showed that the endosymbionts were contained in separate vacuoles that termed as symbiosomes. The presence of symbiosomes, while creates suitable micro-niche for symbionts, also limits their access to environmental nutrients. While nutrients such as oxygen, carbon dioxide or methane could pass through the membrane freely, other nutrients such as ammonia yet rely completely on the host for supply. What's more, we have observed a high consumption of ammonia in endosymbionts while the production of ammonia is relatively low (Fig. 3B, Supplementary Fig. 2E, Supplementary Table. 4), suggesting a great demand of ammonia supply from host. The ammonia, while important for symbiont, yet is detrimental for host as by-products of protein digestion and nucleotide metabolism. To facilitate the transport of ammonia into symbiosomes, we found that the mature bacteriocytes are massively encoding the ammonium transporter genes (Fig. 3B). While the supplement of ammonia by host is a common scenario in other symbiotic associations via the glutamine synthetase/glutamate synthase cycle and translocating of polyamines (14), the supplement of ammonia by ammonium

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.28.493830; this version posted May 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

transporters could satisfy the nutritional needs of symbionts more directly and efficiently (15).

While above examples collectively demonstrated a strong coordination in the metabolism of both partners, we also noticed that the intimate cooperation would be compromised when methane supply was limited and symbiont abundance dropped. In detail, we found that the overall metabolic level of both partners could be greatly repressed when the relative abundance of endosymbionts decreased due to insufficient methane supply (Fig. 3CD, Supplementary Fig. 3C). Noticeably, several key genes in methane oxidation and carbohydrate/lipid biosynthesis were significantly down-regulated in symbionts, suggesting a divestment in the carbon fixation (Fig. 3C, Supplementary Table. 4). In congruent with the austerity of symbionts, the expression of genes involved in the metabolic interaction, such as 24-hydroxycholesterol 7 alpha-hydroxylase, ammonium transporter and sugar phosphate exchanger, were also robustly repressed in mature bacteriocytes, demonstrating a reciprocal modulation in the metabolic coordination (Fig. 3D).

WITHDRAWN
see manuscript DOI for details

Symbiosis-related immune process in bacteriocytes

While the metabolic coordination endows the necessity of symbiosis, the host has to rely on its immune system to harness the association. We therefore screened the possible genes and pathways that participate in the establishment and maintenance of symbiosis in bacteriocyte lineages. Noticeably, we found that only a few pattern recognition receptors (PRRs) were abundantly expressed in either mature or naïve bacteriocytes in fully-symbiotic mussels or decolonized mussels while a large amount of PRRs were retained at a relatively low expressional level or even failed to be characterized (e.g. about 55 out of 146 TLRs were not characterized). Of interest, one TLR2 gene and three C1q genes were annotated as the markers for mature and naïve bacteriocytes correspondingly (Fig. 3E, Supplementary Table. 2). In addition, two (rhamnose-binding lectin and TLR2) and four (including two C1q, two PGRP) PRRs in mature and naïve bacteriocytes were found increased after the decolonization correspondingly (Fig. 3F, Supplementary Table. 3). Considering that PRRs such as TLRs and PGRPs could play vital roles in the recognition of symbionts in other

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.28.493830; this version posted May 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

holobionts, the aforesaid PRRs were therefore speculated to be the main effectors in the establishment of symbiotic association for deep-sea mussels.

Besides these PRRs, we have also detected an overwhelming expression of endocytosis-related genes and lysosome-related genes in mature and naïve bacteriocytes of fully-symbiotic mussels (Fig. 3E, Supplementary Table. 2). The coordination of these genes could facilitate the phagocytosis and lysosome-mediated digestion of bacteria (Supplementary Fig. 4), advantaging the establishment of symbiosis and the control symbiont population. In addition, we noticed that the endocytosis and lysosome-mediated digestion could also be dynamically regulated by the host given to the symbiont abundance (Fig. 3F, Supplementary Table. 3), therefore providing a flexible toolkit to maintain the association.

As reported, the naïve bacteriocytes could bridge the colonization of symbionts from symbiotic bacteriocytes to aposymbiotic bacteriocytes. In consistency, we have observed the enrichment of cell adhesion- and exocytosis-related genes (such as neuroglian, filaggrin-2, dipeptidyl peptidase DPP4, Ras-related protein Rab-3) in naïve bacteriocytes (Fig. 3E, Supplementary Table. 2), which therefore facilitated the transcellular transport of endosymbionts.

WITHDRAWN
see manuscript DOI for details

Coordinated regulatory networks guiding the molecular functions of bacteriocytes

While above results collectively showed the host is adjusting its metabolism and immune process to facilitate symbiosis, we wondered if there were coordinated regulatory networks guiding these adaptions. We hypothesized that the regulatory networks should contain majority of aforesaid genes, which shared an orchestrated expression pattern across all cell types or samples. We then constructed the regulatory network via the weighted gene co-expression network analysis (WGCNA) using the scRNA-seq data of all cell clusters in fully-symbiotic and decolonized mussels. As a result, we obtained six co-expression modules that positively associated with three bacteriocytes subpopulations (Fig. 4A, Supplementary Table. 5). Among them, the brown, light green and dark orange module were of interest as they contained 112/199, 8/57 and 14/82 marker genes of mature, immature and naïve bacteriocyte correspondingly (Supplementary Table. 6). While the marker gene number is limited

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.28.493830; this version posted May 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

in light green and dark orange module, the functional enrichment results of the 112 element maker genes in brown module were yet largely identical with that of all 199 maker genes of mature bacteriocytes (Supplementary Fig. 5A, Supplementary Table. 7), suggesting the markers in brown module could be the core elements in the regulatory networks of mature bacteriocytes.

We then screened the highly expressed transcription factors or signal transducers in the regulatory networks, which could be the hub regulator of bacteriocyte function. Within the regulatory networks of mature bacteriocytes, we identified seven transcription factors (zinc finger protein ZNF271, ETS-related transcription factor ELF-3, autism susceptibility gene AUST2, histone-lysine N-methyltransferase EZH2, GATA zinc finger domain-containing protein GATAD14, hepatocyte growth factor receptor MET, myoneurin) that abundantly expressed in mature bacteriocytes (as marker genes) and correlated with the expression of 107/199 marker genes, including the aforesaid metabolic genes such as sugar phosphate exchanger, ammonium transporter, and immune genes such as TLR2, rabenosyn-5, lysosomal-trafficking regulator, cathepsins (Fig. 4B). Among these seven transcription factors, four (ZNF271, ELF-3, AUST2, EZH2) could also be significantly down-regulated during decolonization, in convergence with the differential expression of another 63 marker genes (Fig. 4C). These findings highlighted their regulatory role in the expression of mature bacteriocyte marker genes. Moreover, a strong interconnectivity (weight higher than 0.3) could be observed within these transcription factors, showing that the hub genes of mature bacteriocytes could work coordinately to advantage the efficiency of the regulatory network. Similarly, we presumed serine/threonine-protein kinase ULK4 and histone-lysine N-methyltransferase PRDM9 as the hub genes of immature and naïve bacteriocytes correspondingly, by connecting other 4 and 12 markers in the regulatory networks (Fig. 4D).

The successive trajectory for the development of bacteriocyte lineages

In comparison with the non-symbiotic mussels, the gill tissue of deep-sea mussel is featured by the dense bacteriocytes along the gill filaments. How these specialized epidermal cells developed is yet largely undiscovered. With the updated information

WITHDRAWN
see manuscript DOI for details

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.28.493830; this version posted May 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

on the expression atlas and molecular function of different bacteriocyte lineages, we are now able to take a glance at it. We firstly have to identify the stem-like cells in gill tissue. By the EdU-labeling assay, we found that the stem-like cells were mainly located at the gill base (dorsal part of the gill, connected with mantle) with the rest dispersed along the gill filaments (Fig. 4E), in a similar way with that in fish gill (16). While the stem-like cells were failed to be included in the ST-seq sample, we noted that the cell cluster 6 and 10 in scRNA-seq could be the stem-like cells as they were abundantly transcribing cycle-related genes and cell proliferation- and differentiation-related genes, including some marker genes of stem-like cells, such as G1/S-specific cyclin-E gene (17, 18), calcium/calmodulin-dependent protein kinase type II (19, 20), clathrin heavy chain 1 (21, 22), transcriptional coactivator YAP1 (23), protein sprouty homolog 2 (24) and discoidin domain-containing receptor 2 (25) (Supplementary Fig. 5B, Supplementary Table. 2). With the expression atlas of stem-like cells and bacteriocyte lineages, we then characterized the development trajectory of bacteriocytes by the pseudo-time analysis using Monocle. As a result, the bacteriocytes and stem-like cells were found distributed at three distinct pseudo-time states, where majority of the stem-like cells, being the starting point of differentiation, were in state 3 of the trajectory while and immature bacteriocytes dispersed evenly in state 3 and state 2 (naïve BC: 50.48% state 3, 39.09% state 2, immature BC: 45.71% state 3, 50.48% state 2). In comparison, as the endpoint of differentiation, about 80.74% of mature bacteriocytes were in state 1 while 17.25% in state 2 and 2.01% in state 3 (Fig.4F). The distinct pseudo-time trajectory states, while confirmed the developmental heterogeneity within the bacteriocytes, also suggested an successive process in the differentiation (cell type transitions, e.g. transitions from stem-like cell to naïve/immature/mature bacteriocytes) and maturation (subtype transitions, e.g. transitions from state 3 to state 2 or state 2 to state 1).

To further identify the differentiation processes of bacteriocyte lineages, we then examined the abundantly expressed genes in different pseudo-time states (Supplementary Table. 8). As the indicators of each state, we speculate that the progenitor states would massive encode some hub marker genes (especially the

WITHDRAWN
see manuscript DOI for details

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.28.493830; this version posted May 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

transcription factors) of the descendent cells while the expression level of marker genes could increse during further maturation, maximizing at the function-matured state. It turned out that the state 1 of mature bacteriocytes and state 3 of immature/naïve bacteriocytes could be the function-matured state, transcribing more marker genes (including the hub marker genes) in comparison with the rest states (Fig.4G, Supplementary Table. 8). Meanwhile, the state 2 of mature bacteriocytes could be the function-maturing state with moderate expression of marker genes. Interestingly, we also noted that majority of the abundantly expressed genes in the state 1 of stem-like cells (16/18) and immature bacteriocytes (12/23) were the marker genes of mature bacteriocytes while 12/77 of abundantly expressed genes in the state 3 of naïve bacteriocytes were the marker genes of immature bacteriocytes (Fig.4H, Supplementary Table. 8), suggesting that the state 1 stem-like cells and immature bacteriocytes could be the progenitor of mature bacteriocytes while immature bacteriocytes might come from the state 3 naïve bacteriocytes. It also came to our notice that the state 3 mature bacteriocytes were encoding multiple DNA replication and cell circle-related genes (Fig. 4I, Supplementary Table. 8), indicating that the mature bacteriocyte could proliferate directly. In support of this hypothesis, we have observed an intensive proliferation signal in some mature bacteriocytes by EdU-labeling assay (Fig. 4J) (26).

WITHDRAWN
see manuscript DOI for details

The co-option of conserved transcription factors in the development of bacteriocytes

While the development of symbiotic cells and organs could be a highly organized process, the regulatory networks guiding the developmental process are yet difficult to identify in either model or non-model associations. Here, with the development trajectory of bacteriocytes, we then explored the mechanisms beneath the differentiation and maturation, focusing on the abundantly expressed genes in the progenitor state. As a result, we found that transcripts of histone-lysine N-methyltransferase EZH2, one of the hub marker genes in the regulatory network of mature bacteriocytes, were also abundantly expressed in the state 1 stem-like cells (Fig.5A, Supplementary Table. 8). Meanwhile, the CD22 gene and prosaposin gene, two element marker genes in the regulatory network of mature bacteriocytes and

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.28.493830; this version posted May 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

crucial regulator in cell differentiation and lysosomal function (27-29), were also highly expressed in the state 1 stem-like cells (Fig.5A, Supplementary Table. 8). These three genes were speculated to promote the differentiation of state 1 stem-like cells into mature bacteriocytes. In addition with these three genes, ETS-related transcription factor ELF-3, another hub gene in the regulatory network of mature bacteriocytes, was also found significantly promoted in state 1 of immature bacteriocytes (Fig.5B, Supplementary Table. 8), collectively contributing to the differentiation of immature bacteriocytes into mature bacteriocytes. Similarly, we observed a robust expression of the Ras-specific guanine nucleotide-releasing factor 2 gene (RASGRF2), a marker of immature bacteriocytes and activator of Ras signaling pathway (30), in the state 3 naïve bacteriocytes (Fig.5C, Supplementary Table. 8). These genes were speculated to facilitate the transition from naïve bacteriocytes to immature bacteriocytes by transducing intracellular signals and activating the expression of immature bacteriocyte-specific genes subsequently.

WITHDRAWN
see manuscript DOI for details

During the maturation process of mature bacteriocytes, we have observed a gradual increase of four hub marker genes (ZNF271, EZH2, ELF-3, AUST2) from state 2 to state 1 (Fig.5D, Supplementary Table. 8). Meanwhile, about 15 transcription factors (or activator/repressor) along with 3 nuclear receptors were also found intensively expressed at the state 2 of mature bacteriocytes (Supplementary Fig.5C, Supplementary Table. 8). These genes were speculated to be the pioneer molecules collaboratively promoting the function maturation of bacteriocytes. Interestingly, we also noted that 24-hydroxycholesterol 7alpha-hydroxylase (EC 1.14.14.26), another marker gene of mature bacteriocyte as well as a sterol metabolism gene, were abundantly expressed at state 2 (Supplementary Fig.5C, Supplementary Table. 8). The sterol metabolic gene could produce ligands for nuclear receptors such as 7α,24(S)-diihydroxycholesterol and primary bile acids by coordination with symbiont sterol metabolism (31), suggesting the participation of symbionts in the maturation of bacteriocytes.

While the aforesaid hub transcription factors and signaling transducers were suggested to be the pioneer molecules promoting the development of bacteriocytes,

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.28.493830; this version posted May 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

we wondered whether they were evolutionarily novel genes diverged from the non-symbiotic ancestors or conserved across evolution. It turned out that, by conducting the phylogenetic analysis using the protein sequences of their homologues from lophotrochozoa, all these genes were highly conserved across the mollusk (Supplementary Fig. 6), concluding that the development of bacteriocytes was achieved by the co-option of conserved genes rather than the evolutionarily novel genes (Fig. 5E).

Discussion

Being an evolutionary novelty, the symbiotic cells and organs have provided a suitable niche for the symbiosis and entailed the holobionts unique adaptation to the extreme environments. Here, with the integrated information from single-cell transcriptome and spatial transcriptome, we have successfully uncovered the function and development of symbiotic cells in deep-sea mussels, which is guided by an ancestral and intrinsic toolkit with co-option of conserved transcription factors and modulated by the chemosynthetic endosymbionts. These findings, while demonstrated the how the holobionts adapt to extreme environment of deep-sea via the intimate coordination, have also proposed a promising roadmap in revealing the formation and development of symbiotic cells and organs in all holobionts.

Metabolic remodeling of bacteriocytes to maximize the symbiotic profits

It is widely recognized that the formation of symbiotic associations in mollusk are nutrition-driven. However, whether and how the molluscan symbiotic cells adjust the metabolism to acquire nutrition from symbionts remained to be fully addressed. More importantly, while studies have revealed the nutritional contribution of symbionts, the way that molluscan host supports the metabolism of symbionts is largely in obscure (11, 13, 32-34). With the state-of-the-art single-cell transcriptome, it is now clear that the deep-sea mussels have reshaped the bacteriocytes metabolism remarkably to maximize the symbiotic profits for both partners. For example, the bacteriocytes are encoding dozens of genes involved in the biosynthesis and transport of carbohydrate, lipid, amino acid and vitamin to better the acquirement of nutrition from symbionts. Among them, it was for the first time demonstrated that the bacteriocytes would

WITHDRAWN
see manuscript DOI for details

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.28.493830; this version posted May 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

retrieve fructose-6P directly with the methanotrophic symbionts by sugar phosphate exchanger genes and converted it into glucose-6P, fructose-1,6P_2 and glyceraldehyde-3P before redistributing back to both the host and symbionts as supplements of gluconeogenesis and TCA cycle. We have also demonstrated the direct supply of ammonia from the host to symbionts via the ammonium transporters. The bioconversion and exchanges of sugar phosphate and ammonia between host and symbionts provided an efficient and direct way to coordinate the metabolism of both partners, which have greatly promoted the efficiency and profits of symbiosis.

Adapted immune process in harnessing the symbiotic associations

The establishment and maintenance of endosymbiosis with exogenous bacteria is always a challenge for multicellular organisms in the presence of host immune system, which therefore require an extensive adjustment to entail the inoculation and proliferation of symbionts inside the cell (35). With the single-cell expression atlas, we have unexpectedly characterized the different immune strategies adopted by three bacteriocyte lineages and noted that only a few PRRs were abundantly expressed in mature and naïve bacteriocytes disregard their expansion in genome (11). On the other hand, we for the first time confirmed that the mature bacteriocytes were capable of phagocytizing exogenous bacteria, which directly facilitate the entrance of endosymbionts. In comparison with mature bacteriocytes, the naïve bacteriocytes were found encoding multiple cell adhesion- and exocytosis-related genes, which are responsible for the colonization of symbionts in these newborn cells. Although the phagocytosis and exocytosis of bacteriocytes has not been reported, similar phenomena were observed in non-symbiotic molluscan gill cells (36-38), which therefore is believed to be a conserved trait across mollusk and a necessity in the formation of symbiosis.

A more interesting finding is that we could rarely see digested symbionts in the lysosomes of normal mussels as well as any sign of proliferating endosymbionts in the EdU-labeling assay, which implied that lysosome-mediated symbiont digestion would be the secondary option in obtaining nutrition regardless of the abundant expression of lysosome-related genes in all three lineages. In convergence with the

WITHDRAWN
see manuscript DOI for details

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.28.493830; this version posted May 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

speculation, we have noted a significant repression of lysosome activity in mussels containing a limited number of symbionts after the long-term methane recession. Collectively, it was suggested that the lysosome could be more vital in controlling the population of symbionts rather than satisfying the nutritional needs.

Unification of the function and development via the ancestral toolkit

Being an exquisite and stable niche for the endosymbiosis, the development of symbiotic cells is intriguing and crucial to understand the formation and evolution of symbiosis (2, 3). With the single-cell transcriptome data, we have for the first time recognized the successive development trajectory of bacteriocytes in deep-sea mussel, which was controlled by several conserved transcription factors that also responsible for the molecular function of bacteriocytes. The synergistic modulation by these genes confirmed that the formation of symbiotic cells in deep-sea mussel was accomplished by the co-option of conserved genes instead of evolutionary novel genes. Interestingly, similar phenomena could also be observed in deep-sea scaly-foot snail where the formation of biomineralised skeletons is also driven by an ancestral and intrinsic toolkit that conserved across mollusk (39), collectively certificating the robust developmental plasticity of mollusk. It is also noticeable that the function and development of mussel bacteriocytes could be influences by symbiont via the sterol metabolism. The participation of symbionts in the development of symbiotic cells and organs has attracted much attention for years yet remained largely in obscure at the mechanistic level (40, 41). The sterol metabolism of symbionts yet could be a common and promising way in such process since majority of animal hosts are relying on their symbionts for the sterol intermediates (42-44). In support of the speculation, studies have also proposed the manipulation of host development and reproduction by symbiotic *Wolbachia* via the steroid-nuclear receptor signaling pathway (45). While the regulatory role of symbionts in host development remained to be fully addressed, our finding has exemplified the robust plasticity of molluscan symbiotic cells, which entailed the wide distribution of mollusk in a wide range of habitats including the deep-sea.

WITHDRAWN
see manuscript DOI for details

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.28.493830; this version posted May 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Conclusion

In conclusion, with the high-resolution single-cell transcriptome, we have successfully elucidated the population census, function atlas and development trajectory of bacteriocyte lineages in deep-sea mussels. The function and development of bacteriocytes are highly coordinated by the co-option of conserved genes and could be dynamically modulated by the symbionts. The intimate symbiotic association and robust plasticity of symbiotic cells have certificated the mollusk as one of the most successful organism in the deep-sea. Further works focusing on the function and development of symbiotic cells and organs in more model and non-model symbiotic associations are needed to clarify the formation and evolution of symbiosis from the ecological evolutionary development aspects. .

WITHDRAWN
see manuscript DOI for details

Methods

in situ transplantation and animal collection

The *G. platifrons* mussels were collected from cold seeps (site F, 22°06′N, 119°17′E) of the South China Sea during the cruise of 2018 and 2020. The mussel fauna inhabit at 1,120 m beneath the surface with temperature of approximately 3.35□, salinity of about 35.54 psu, dissolved oxygen of 2.98–3.17 mg/L, and methane concentrations up to 31,227 ppm (seepage region) (46). To shelter the specimens from temperature and pressure alternations during sampling, all mussel samples were collected with an isothermal isobaric sampler and a manually controlled macrofauna *in situ* sampling device as described previously (47). In brief, several mussels inhabit closely with the active fluid seepages were collected using the isothermal isobaric sampler for 10×Genomics single cell transcriptome sequencing (scRNA-seq) and 10×Visium spatial transcriptome sequencing (ST-seq) (designated as InS group). Other mussels were also collected with the multipurpose *in situ* sampling device and treated with 4% paraformaldehyde *in situ* for the *in situ* hybridization (ISH) assay and transmission electron microscopy (TEM) imaging. During the 2018 cruise, an *in situ* transplantation assay was conducted in which dozens of mussels in the seepage region were translocated to an authigenic carbonate region with low concentration of CH_4

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.28.493830; this version posted May 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(about 100 m away from the seepage with methane concentration dropped to 800 ppm). One of the transplanted mussels was retrieved 604 days later during the 2020 cruise using isothermal isobaric sampler for scRNA-seq and ST-seq (designated as DeC group). After the retrieval, the isothermal isobaric sampler was depressurized instantly while mussels were quickly dissected into two half to remove excessive seawater and frozen in isopentane bath (Macklin) with liquid nitrogen for scRNA-seq and ST-seq.

ScRNA-seq

Two individual mussels with similar size (about 80 mm in length) from InS and DeC group were employed for the scRNA-seq. Considering the difficulties in isolation and *in vitro* culture of gill cells as well as the potential influences on gene expression during mussel sampling and single cell preparation, scRNA-seq of *G. platiforns* gill tissue were conducted with the nuclei isolated from fresh frozen samples that collected in isothermal isobaric way. The isolated nuclei were subjected to the 10×Genomics single cell protocol immediately by generating the Single-cell Gel Bead-In-EMlusion (GEM) with the help of GemCode single-cell instrument. The full-length cDNA were synthesized using the Chromium Next GEM Single Cell 3' Reagent Kits v3.1 and sequenced by Illumina HiSeq X Ten platform by Gene Denovo Biotechnology Co. (Guangzhou, China).

Cryosection and ST-seq

Gill tissue employed for the 10×Visium ST-seq was first incubated with precooled methanol (4□) for 15 min to reduce potential damage by the microtome blade during cryosection. The tissue was then embedded in OCT (Sakura) and cross sectioned at the middle part of gill in a cryostat (Leica CM1950). Morphology of the gill sections (10 μm thickness) were quick checked under light microscopy with 4,6-diamidino-2-phenylindole (DAPI, Thermo Fisher) and hematoxylin-eosin staining (Sangon Biotech). A successive section was then obtained from the suitable embedded tissue block and placed on the Visium Spatial Tissue Optimization Slide within the capture area for the tissue optimization assay. For the sequencing assay, appropriate gill sections were first mounted on the Visium Spatial Gene Expression Slide and

WITHDRAWN
see manuscript DOI for details

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.28.493830; this version posted May 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

permeabilized according to the aforesaid parameters (permeabilization time 9 min) after HE staining and microscope imaging. After the reverse transcription and library preparation, all samples were sequenced on the Illumina HiSeq X Ten platform by Gene Denovo Biotechnology Co. (Guangzhou, China).

Cell clustering, cell type annotation and differentially expressed genes analysis

Single-cell clustering of *G. platifrons* gill tissue was first conducted using data from scRNA-seq. After the quality control, all raw reads were mapped to the reference genome of *G. platifrons.* The genome was first reported by Sun et al (11) and updated by us (unpublished data). Unique molecular identifier (UMI) in each sequenced read was then counted and corrected for sequencing errors. With the information of valid barcodes that identified based on the EmptyDrops method, the gene matrices of all cells were then produced and imported into Seurat (version 3.1.1) for cell clustering using a graph-based clustering approach.

Single-cell clustering of *G. platifrons* gill tissue was also conducted with ST-seq data advantaging the cell type annotation. The slide image obtained before the permeabilization procedure was first imported into Space Ranger software for the fiducial alignment and Visium barcoded spot alignment. After decoding the correlation between tissue and capture spots (as well as the barcodes), a splicing-aware alignment of sequencing reads to the *G. platifrons* genome was conducted by the STAR program inside the Space Ranger software.

An integrated analysis using scRNA-seq data, ST-seq data and *in situ* hybridization assay was finally employed for the cell type annotation in InS group. In short, an anchor-based integration method was first used for the integration of ST-seq data with scRNA-seq data using FindIntegrationAnchors command in Seurat-v3.2. All cell type labels of the scRNA-seq were then transferred to the spatial data using TransferData command. Cell type predication scores indicating the similarity between the ST-seq spots and scRNA-seq cell clusters were calculated simultaneously and only the spot-cluster pair with highest score was considered for further cell type annotation. For the further annotation and verification of cell types, a brief inspection of cell types in gill tissue were first conducted according to the HE staining image of gill

WITHDRAWN
see manuscript DOI for details

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.28.493830; this version posted May 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

cryosection and the tissue plot with spots colored by clustering from ST-seq. Several cell markers obtained using the FindAllMarkers function in Seurat were also cloned and subjected to *in situ* hybridization using gill tissues of InS group.

To further investigate the biological function of all identified cell clusters, differentially expressed genes (DEGs) were surveyed in both scRNA-seq and ST-seq. For scRNA-seq, expression values of all identified genes in given cluster were compared against the rest of cell clusters using Wilcoxon rank sum test. Genes expressed mainly in the target cluster (more than 25% of the cells are the target cluster) with at least 1.28-fold up-regulation and *p*-value less than 0.01 were considered as DEGs per cell cluster. For ST-seq, R package named as MAST (short for Model-based Analysis of Single cell Transcriptomics) were employed in determining the DEGs for a single cell cluster using the same criterion with scRNA-seq. All the genes in ST-seq were further subjected to the analysis of spatially DEGs using markvariogram in Seurat R package and mark-segregation hypothesis testing in trendsceek R package. Genes with r.metric.5 parameter less than0.8 and *p*-value less than 0.01 were suggested to be spatially DEGs.

WITHDRAWN
see manuscript DOI for details

Meta-transcriptome sequencing

To explore the expression atlas of symbionts in InS and DeC group, we also conducted the meta-transcriptome sequencing using the same samples in the scRNA-seq. In addition, six more samples belonging to the InS group were also employed to advance the reliability of results. The extraction of total RNA, removal of eukaryotic and prokaryotic rRNA and the synthesis of cDNA library of the bulk-seq were conducted using the same protocol of previous reports. A total of eight libraries belonging to the InS and DeC group was finally sequenced using the Illumina HiSeq 2500 platform with paired-end reads by Novogene. After the quality control, the filtered reads were then aligned against the genomes of both the endosymbiont and *G. platifrons* using HISAT (v2.0.4). The genome of methanotrophic symbionts was first reported by Takishita et al., (2017) and updated by us (48). The expression level of host and symbiont gene was calculated by HTSeq (v0.6.1) while the significant difference between groups was determined by DESeq

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.28.493830; this version posted May 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(v1.10.1). Additionally, we also defined the top 10% genes with the most abundant transcripts in InS group as the abundantly expressed genes to survey the biological processes that actively occurred *in situ* under normal condition.

GO/KEGG/GSEA analysis

Gene Ontology (GO) annotation of *G. platifrons* genes was obtained by Blast2GO software (version 5.2) and employed for GO enrichment analysis using homemade scripts. Enriched GO terms were determined from the gene numbers of the DEGs against the genome background and verified using a hypergeometric test (FDR less than 0.05). Similarly, KEGG pathway enrichment analysis was conducted using homemade scripts with KEGG annotations obtained from the KEGG database (http://www.genome.jp/kegg/pathway.html).

To determine whether an a priori defined set of genes of the selected cell cluster was significantly modulated between InS and DeC group, we also introduced Gene Set Enrichment Analysis (GSEA) in this study In comparison with conventional GO/KEGG enrichment analysis, genes with subtle expression changes (suggested as the minor genes) were also taken into consideration by the GESA analysis considering the number of median genes per cell was far less than that in conventional bulk RNA-seq. The GSEA analysis was conducted with the GSVA R package (v1.26) while the background information was obtained and modified from Molecular Signatures Database (http://www.gsea-msigdb.org/gsea/index.jsp).

Cell trajectory construction and WGCNA analysis of bacteriocytes

The single cell trajectory analysis was conducted by Monocle (Version 2.6.4) using the reversed graph embedding algorithm. Gene expression matrix of stem-like cells and bacteriocytes were employed in the analysis and visualized by the orderCells function (sigma = 0.001, lambda = NULL, param.gamma = 10, tol = 0.001). The weighted gene co-expression network analysis (WGCNA) is conducted using the WGCNA package (v1.47) in R software using the gene expression values (power=7, minModuleSize=50) and visualized using Cytoscape (v3.8.2).

EdU-labeling assay, phagocytosis assay, ISH assay and phylogenetic analysis

The EdU-labeling assay was carried out *in situ* using mussels collected from the

WITHDRAWN
see manuscript DOI for details

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.28.493830; this version posted May 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

seepage region. In brief, mussels were incubated with 5-ethynyl 2'-deoxyuridine (EdU, final concentration at 40 μM) in a manually controlled macrofauna *in situ* experiment device for about 18 h and retrieved afterward using the isothermal isobaric sampler. The Click-iT Plus EdU Imaging Kit (ThermoFisher) was then used to visualize the EdU signal.

The phagocytosis assay was conducted with primary gill cells of fresh collected mussels using methods modified from previous reports (36, 49). Considering that endosymbiotic methanotrophs was nonculturable, *Vibrio alginolyticus*, an environmental bacteria isolated in cold seep, was employed in this assay and labeled with fluorescein isothiocyanate (FITC, sigma) before use. To collect the primary gill cells, gill tissue were first treated with 1% trypsin (diluted by sterilized seawater) at 4□ for 30 min and then centrifuged at 500 rpm (4□, 5 min) and 3000 rpm (4□, 5 min) successively. Cell pellets were finally resuspended in modified L15 medium (Gibco, supplemented with 0.54 g/L KCl, 0.6 g/L $CaCl_2$, 1 g/L $MgSO_4$, 3.9 g/L $MgCl_2$, 20.2 g/L NaCl) to a final concentration of 1×10^6 cells mL^{-1} and incubated with the same volume of FITC-labeled *V. splendidus* (1×10^8 cells mL^{-1}) for 30 min at 4□ in dark. After the incubation, primary cells were washed three times with modified L15 medium to remove extracellular bacteria. Finally, cell samples were imaged on the laser scanning confocal microscope (Zeiss LSM710) after treated stained with DAPI and DiI perchlorate.

The *in situ* hybridization assay (ISH) was conducted as previously described (50). For the verification of cell types, cell markers were cloned with gene-specific primers to synthesize the digoxigenin labelled ISH probes. Fluorescent ISH for symbionts were conducted using Cyanine 3 (Cy3)-labeled Eub338 eubacteria probe (5'-GCTGCCTCCCGTAGGAGT-3') or FITC-labeled pmoB (methanotroph-specific gene) probe (5'-CGAGATATTATCCTCGCCTG-3').

For the phylogenetic analysis of hub transcription factors, homologue proteins were first obtained by NCBI blastp program and aligned using Seaview. The maximum likelihood phylogenetic trees for these proteins were then constructed by Mega software (v11) in jones-taylor-thornton model with bootstrap of 100.

WITHDRAWN
see manuscript DOI for details

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.28.493830; this version posted May 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Author contributions

HC participated in the study design and data interpretation. ML, MW, ZS and Chao Lian helped in the sample collection. HC and ML conducted the majority of the experimental work and data analysis. HC drafted the manuscript, while MW, ZS, LZ, HZ, and HW assisted with the interpretation of the results. LC surveyed environmental parameters during the cruise. MW and Chaolun Li conceived the study, coordinated the experiments, and helped draft the manuscript. All authors gave their final approval for publication.

Conflict of interests

The authors declare no competing interests.

WITHDRAWN
see manuscript DOI for details

Acknowledgments

We thank all the crew members onboard the R/V Kexue for their assistance in sample collection and all the laboratory staff for continuous technical advice and helpful discussions. The present study was supported by the National Natural Science Foundation of China (42106134, 42030407 and 42076091), the Key Research Program of Frontier Sciences (ZDBS-LY-DQC032) and the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA22050303, XDB42020401), Key Deployment Project of Centre for Ocean Mega-Research of Science, CAS (COMS2020Q02) and Marine S&T Fund of Shandong Province for Pilot National Laboratory for Marine Science and Technology(Qingdao) (No. 2022QNLM030004).

Data Accessibility

All sequencing data were deposited in the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/) with accession NO. PRJNA838712.

References

1. Douglas AE (2014) Symbiosis as a General Principle in Eukaryotic Evolution. *Cold Spring Harbor perspectives in biology* 6(2).

2. Gilbert SF, Bosch TCG, & Ledón-Rettig C (2015) Eco-Evo-Devo: developmental

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.28.493830; this version posted May 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

symbiosis and developmental plasticity as evolutionary agents. *Nature Reviews Genetics* 16(10):611-622.

3. Douglas AE (2020) Housing microbial symbionts: evolutionary origins and diversification of symbiotic organs in animals. *Philosophical Transactions of the Royal Society B: Biological Sciences* 375(1808):20190603.

4. McFall-Ngai Margaret J & Ruby Edward G (1991) Symbiont Recognition and Subsequent Morphogenesis as Early Events in an Animal-Bacterial Mutualism. *Science* 254(5037):1491-1494.

5. Braendle C, *et al.* (2003) Developmental Origin and Evolution of Bacteriocytes in the Aphid–Buchnera Symbiosis. *PLoS biology* 1(1):e21.

6. Matsuura Y, Kikuchi Y, Miura T, & Fukatsu T (2015) Ultrabithorax is essential for bacteriocyte development. *Proceedings of the National Academy of Sciences* 112(30):9376-9381.

7. Dubilier N, Bergin C, & Lott C (2008) Symbiotic diversity in marine animals: the art of harnessing chemosynthesis. *Nature Reviews Microbiology* 6:725.

8. Duperron S, *et al.* (2013) An overview of chemosynthetic symbioses in bivalves from the North Atlantic and Mediterranean Sea. *Biogeosciences* 10(5):3241-3267.

9. Rodrigues CF, *et al.* (2015) A sad tale: has the small mussel Idas argenteus lost its symbionts. *Biological Journal of the Linnean Society* 114:389-405.

10. Miyazaki J-I, Martins LdO, Fujita Y, Matsumoto H, & Fujiwara Y (2010) Evolutionary Process of Deep-Sea Bathymodiolus Mussels. *PloS one* 5(4):e10363.

11. Sun J, *et al.* (2017) Adaptation to deep-sea chemosynthetic environments as revealed by mussel genomes. *Nature Ecology & Evolution* 1(5):121.

12. Zhao Y, *et al.* (2020) Ecological characterization of cold-seep epifauna in the South China Sea. *Deep Sea Research Part I: Oceanographic Research Papers* 163:103361.

13. Takishita K, *et al.* (2017) Genomic Evidence that Methanotrophic Endosymbionts Likely Provide Deep-Sea Bathymodiolus Mussels with a Sterol Intermediate in Cholesterol Biosynthesis. *Genome Biol Evol* 9(5):1148-1160.

14. Cui G, *et al.* (2019) Host-dependent nitrogen recycling as a mechanism of

WITHDRAWN
see manuscript DOI for details

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.28.493830; this version posted May 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

symbiont control in Aiptasia. *PLoS genetics* 15(6):e1008189.

15. Fochi V, *et al.* (2017) Fungal and plant gene expression in the Tulasnella calospora–Serapias vomeracea symbiosis provides clues about nitrogen pathways in orchid mycorrhizas. *New Phytologist* 213(1):365-379.

16. Stolper J, *et al.* (2019) Stem cell topography splits growth and homeostatic functions in the fish gill. *eLife* 8:e43747.

17. Gonnot F, Langer D, Bourillot P-Y, Doerflinger N, & Savatier P (2019) Regulation of Cyclin E by transcription factors of the naïve pluripotency network in mouse embryonic stem cells. *Cell cycle* 18(20):2697-2712.

18. Ables ET & Drummond-Barbosa D (2013) Cyclin E controls Drosophila female germline stem cell maintenance independently of its role in proliferation by modulating responsiveness to niche signals. *Development* 140(3):530-540.

19. Aji K, *et al.* (2016) Differentiation of Human Adipose Derived Stem Cells into Smooth Muscle Cells Is Modulated by CaMKII γ. *Stem Cells International* 2016:1267480.

20. Gu Y, *et al.* (2012) CaMKII γ, a critical regulator of CML stem/progenitor cells, is a target of the natural product berbamine. *Blood* 120(24):4829-4839.

21. Narayana YV, Gadgil C, Mote RD, Rajan R, & Subramanyam D (2019) Clathrin-Mediated Endocytosis Regulates a Balance between Opposing Signals to Maintain the Pluripotent State of Embryonic Stem Cells. *Stem Cell Reports* 12(1):152-164.

22. Mote RD, *et al.* (2020) Loss of clathrin heavy chain enhances actin-dependent stiffness of mouse embryonic stem cells. *bioRxiv*:2020.2005.2010.086579.

23. Camargo FD, *et al.* (2007) YAP1 Increases Organ Size and Expands Undifferentiated Progenitor Cells. *Current Biology* 17(23):2054-2060.

24. Mendenhall EM, Eckfeldt CE, Ekker SC, & Verfaillie CM (2004) The Role of Sprouty Family Members in Hematopiesis in Zebrafish and Mammals. *Blood* 104(11):137.

25. Yang H, *et al.* (2020) DDR2, a discoidin domain receptor, is a marker of periosteal osteoblast and osteoblast progenitors. *Journal of Bone and Mineral*

WITHDRAWN
see manuscript DOI for details

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.28.493830; this version posted May 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Metabolism 38(5):670-677.

26. Piquet B, *et al.* (2020) Regionalized cell proliferation in the symbiont-bearing gill of the hydrothermal vent mussel Bathymodiolus azoricus. *Symbiosis* 82(3):225-233.

27. Samardzic T, *et al.* (2002) CD22 regulates early B cell development in BOB.1/OBF.1-deficient mice. *European journal of immunology* 32(9):2481-2489.

28. O'Brien JS, Carson GS, Seo HC, Hiraiwa M, & Kishimoto Y (1994) Identification of prosaposin as a neurotrophic factor. *Proceedings of the National Academy of Sciences* 91(20):9593-9596.

29. O'Brien JS & Kishimoto Y (1991) Saposin proteins: structure, function, and role in human lysosomal storage disorders. *The FASEB Journal* 5(3):301-308.

30. Fernández-Medarde A & Santos E (2011) The RasGrf family of mammalian guanine nucleotide exchange factors. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* 1815(2):170-188.

31. Weatherman RV, Fletterick RJ, & Scanlan TS (1999) Nuclear-Receptor Ligands and Ligand-Binding Domains. *Annual review of biochemistry* 68(1):559-581.

32. Ponnudurai R, *et al.* (2016) Metabolic and physiological interdependencies in the Bathymodiolus azoricus symbiosis. *The ISME journal*.

33. Ponnudurai R, *et al.* (2017) Genome sequence of the sulfur-oxidizing Bathymodiolus thermophilus gill endosymbiont. *Standards in genomic sciences* 12:50.

34. Geier B, *et al.* (2020) Spatial metabolomics of in situ host–microbe interactions at the micrometre scale. *Nature Microbiology*:1-13.

35. Gross R, *et al.* (2009) Immunity and symbiosis. *Molecular microbiology* 73(5):751-759.

36. Li Y, *et al.* (2017) The hematopoiesis in gill and its role in the immune response of Pacific oyster Crassostrea gigas against secondary challenge with Vibrio splendidus. *Developmental & Comparative Immunology* 71:59-69.

37. Pichon D, *et al.* (2013) Characterization of abalone Haliotis tuberculata–Vibrio harveyi interactions in gill primary cultures. *Cytotechnology* 65(5):759-772.

38. Tame A, Maruyama T, & Yoshida T (2022) Phagocytosis of exogenous bacteria

WITHDRAWN
see manuscript DOI for details

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.28.493830; this version posted May 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

by gill epithelial cells in the deep-sea symbiotic mussel Bathymodiolus japonicus. *Royal Society Open Science* 9(5):211384.

39. Sun J, *et al.* (2020) The Scaly-foot Snail genome and implications for the origins of biomineralised armour. *Nature communications* 11(1):1-12.

40. Altura MA, Stabb E, Goldman W, Apicella M, & McFall-Ngai MJ (2011) Attenuation of host NO production by MAMPs potentiates development of the host in the squid-vibrio symbiosis. *Cellular microbiology* 13(4):527-537.

41. Foster JS, Apicella MA, & McFall-Ngai MJ (2000) Vibrio fischeri lipopolysaccharide induces developmental apoptosis, but not complete morphogenesis, of the Euprymna scolopes symbiotic light organ. *Dev Biol* 226(2):242-254.

42. Bien T, Hambleton EA, Dreisewerd K, & Soltwisch J (2021) Molecular insights into symbiosis—mapping sterols in a marine flatworm-algae-system using high spatial resolution MALDI-2-MS imaging with ion mobility separation. *Analytical and Bioanalytical Chemistry* 413(10):2767-2777.

43. Kok LT, Norris DM, & Chu HM (1970) Sterol Metabolism as a Basis for a Mutualistic Symbiosis. *Nature* 225(5233):661-662.

44. Hambleton EA, *et al.* (2019) Sterol transfer by atypical cholesterol-binding NPC2 proteins in coral-algal symbiosis. *eLife* 8.

45. Negri I (2012) Wolbachia as an “Infectious” Extrinsic Factor Manipulating Host Signaling Pathways. *Frontiers in Endocrinology* 2.

46. Cao L, *et al.* (2021) In situ detection of the fine scale heterogeneity of active cold seep environment of the Formosa Ridge, the South China Sea. *Journal of Marine Systems* 218:103530.

47. Chen H, *et al.* (2021) A glimpse of deep-sea adaptation in chemosynthetic holobionts: depressurization causes DNA fragmentation and cell death of methanotrophic endosymbionts rather than their deep-sea Bathymodiolinae host. *Molecular ecology*.

48. Sun Y, *et al.* (2022) Adaption to hydrogen sulfide-rich environments: Strategies for active detoxification in deep-sea symbiotic mussels, Gigantidas platifrons. *Science of The Total Environment* 804:150054.

WITHDRAWN
see manuscript DOI for details

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.28.493830; this version posted May 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

49. Chen H, *et al.* (2016) An invertebrate-specific and immune-responsive microRNA augments oyster haemocyte phagocytosis by targeting CgIkappaB2. *Scientific reports* 6:29591.

50. Wang H, *et al.* (2021) Molecular analyses of the gill symbiosis of the bathymodiolin mussel Gigantidas platifrons. *iScience* 24(1):101894.

WITHDRAWN
see manuscript DOI for details

Figure legend

Fig.1 Spatially resolved single-cell transcriptomic atlas of deep-sea mussel gill

(A) Experimental workflow and analysis for single-cell transcriptome (scRNA-seq) and spatial transcriptome (ST-seq) of gill tissue in fully symbiotic (InS group) and partially decolonized (DeC) deep-sea mussels.

(B) t-SNE projection of spatial transcriptome clustered by gene expression in InS group and colored by cell type (Left). Projection of the cell clusters in InS group on the spatial transcriptomes barcoded spots (Right). Two successive sections were employed in the same capture region of spatial transcriptome.

(C) t-SNE projection of single-cell transcriptome by gene expression in InS group and colored by cell type (Left). Spatial transcriptome barcoded spots labeled by scRNA-seq cell type with maximum prediction score (Right).

(D) Expression pattern of cell cluster-specific markers. The size of the dot represents the percentage of cells in the clusters transcribing the marker and the color intensity represents the average expression level of the marker gene in that cluster.

Fig.2 Spatially resolved single-cell transcriptomic atlas of deep-sea mussel gill

(A-C) Expression pattern of the representative marker genes of mature (A), immature (B) and naive (C) bacteriocytes shown in the color-coded t-SNE plots. The color intensity represents the expression level of the marker gene in a single cell.

(D) *in situ hybridization* verification of marker genes of mature bacteriocytes.

(E) Gene Ontology (GO) enrichment analysis of the marker genes of mature, immature and naive bacteriocytes. The size of the dot represents the number of enriched genes in the pathway and the color represents the q-value of enrichment

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.28.493830; this version posted May 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

result.

Fig.3 Metabolic and immune genes and processes involved in the symbiosis

(A) GO enrichment analysis of the abundantly expressed genes (Top 10%) of endosymbionts by meta-transcriptome of the sample of InS group. Illustrated in the circular map where the first (inner) circle represents the enrichment factor, the second circle represents the enriched gene number, the third circle represents the background gene number and q-value, and the fourth circle represents the GO pathway.

(B) Schematic drawing showing the metabolic coordination between host and symbionts. The key enzymes were illustrated by the colored rhombus where grey represents enzymes not detected in either bulk-seq (meta-transcriptome) or genome, green represents enzymes detected in low expression in bulk-seq, red and claret represent high expression in scRNA-seq or bulk-seq (Top 10%).

(C) Differentially expressed genes of endosymbionts in decolonized mussels against fully symbiotic mussels. A total of 135 symbiont genes were differentially expressed (Left) while multiple metabolic pathway were found enriched by the Gene Set Enrichment Analysis (GSEA, Right). Among them, three key genes in methane oxidation and carbohydrate/lipid biosynthesis were significantly down-regulated (Right). For the GSEA results, key statistics including enrichment score and ranked list matric of each gene in the given GO pathway were evaluated and significance were determined considering both the normalized enrichment score (NES) and q-value (P*adj*).

(D) Differentially expressed genes of mature bacteriocytes in fully symbiotic mussels against decolonized mussels. A total of 1343 genes were differentially expressed (Left) involved in multiple metabolic pathways by the GO enrichment analysis (Middle, dot size representing the overall number of differentially expressed genes, up-down normalization = [number of up-regulated genes] / [number of down-regulated genes] - 1). Noticeably, genes involved in the carbohydrate, vitamin, lipid and amino acid biosynthesis were found mostly down-regulated in decolonized mussels by the GO enrichment analysis (Right).

WITHDRAWN
see manuscript DOI for details

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.28.493830; this version posted May 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(E) A schematic drawing of the abundantly expressed immune-related marker genes of the mature, immature and naïve bacteriocytes.

(F) Expression pattern of immune-related genes in bacteriocytes of decolonized mussels. Two and four PRRs were found up-regulated in decolonized mussels (upper). A diversity of endocytosis- and lysosome-related genes were also found differentially expressed as showed in radar chart. The size of blue/yellow colored apical circles represent the absolute value of log_2(fold changes) where blue indicated down-regulation in partially decolonized mussels and yellow indicated up-regulation. Green and dark green colored inner diagram represent the abundance of given gene in decolonized and fully symbiotic mussels respectively.

WITHDRAWN
see manuscript DOI for details

Fig.4 Development trajectory of deep-sea mussel bacteriocyte lineages

(A) Clustering of the module eigengenes in the weighted gene co-expression network analysis (WGCNA) of scRNA-seq data for all cell types and samples. The eigengenes represent the overall expression level of genes in each module in the given cell type and sample with blue indicating a negative correlation and red indicating a positive correlation.

(B) Genetic regulatory network of mature bacteriocytes constructed by WGCNA. Marker genes with similar function or in same biological process were in labeled with same color. Blue represents the hub genes (transcription factors and signal tranducers) of the network, green represents lysosome-related genes, light green represent genes involved in metabolism of carbohydrate, amino acids and fatty acids, cyan represents transporters, sand yellow represent sterol metabolism related genes, red represents genes involved in cell differentiation and cytoskeleton, pink represent genes involved in cell circle and apoptosis, grey represents the rest genes.

(C) Expression pattern of the genetic regulatory network elements of mature bacteriocytes during decolonization. Hub transcription factors such as zinc finger protein ZNF271, ETS-related transcription factor ELF-3, autism susceptibility gene AUST2 and histone-lysine N-methyltransferase EZH2 were found down-regulated, accompanying with other 63 element markers.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.28.493830; this version posted May 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(D) Genetic regulatory network of immature (Left) and naïve (Right) bacteriocytes constructed by WGCNA.

(E) EdU-labelling assay of deep-sea mussels. New synthesized DNA of proliferating cells were incorporated with 5-ethynyl 2'-deoxyuridine (EdU) and stained with green fluorescence. Cell nuclei were stained by DAPI while symbionts were detected by Cy3-labeled Eub338 probe.

(F) Development trajectory of bacteriocyte lineages. Pseudo-time analysis by Monocle showed an successive development process for the stem-like cells and bacteriocytes, which could be further divided into three distinct states (Upper two images). Majority of stem-like cells were found in state 3 while the naïve and immature bacteriocytes were distributed evenly in state 3 and state 2. Majority of mature bacteriocytes were in state 1 (Lower six images).

(G) Veen diagrams of the marker genes of mature (Left), immature (Middle), and naïve (Right) bacteriocytes with the state 1-, 2- and 3-related high expression genes (HEG) of corresponding cell types.

(H) Veen diagrams of the marker genes of mature with the state 1-related high expression genes (HEG) of naïve bacteriocyte, immature bacteriocyte and stem-like cells (Left). Veen diagrams of the marker genes of immature with the state 1-, 2- and 3-related high expression genes (HEG) of naïve bacteriocytes (Right).

(I) Expression level of probable serine/threonine-protein kinase PATS1 and calcium-dependent protein kinase C (PRKC1) in different states of mature bacteriocytes. Each dot represents the expression value of a single mature bacteriocyte.

(J) Magnified image for the EdU-labelling assay of deep-sea mussels. EdU signals could be observed in some mature bacteriocytes (white triangle).

Fig.5 Co-option of conserved transcription factors in the function and development of bacteriocytes

(A) Expression pattern of the hub transcription factor EZH2, CD22 and prosaposin in the color-coded t-SNE plots (Left) and different states of stem-like cells.

WITHDRAWN
see manuscript DOI for details

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.28.493830; this version posted May 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(B) Expression pattern of the hub transcription factor ELF-3 in the color-coded t-SNE plots (Left) and different states of immature bacteriocytes.

(C) Expression pattern of the hub signal transducer Ras-specific guanine nucleotide-releasing factor 2 gene in the color-coded t-SNE plots (Left) and different states of naïve bacteriocytes.

(D) Expression level of ZNF271 and AUST2 in different states of mature bacteriocytes.

(E) Schematic diagram on the function and development of bacteriocyte lineages guided by the co-option of conserved transcription factors identified from the genetic regulatory networks. Arrows represent the deduced development trajectory (including cell type transition and maturation) of bacteriocytes.

WITHDRAWN
see manuscript DOI for details

Supplementary. Fig. 1 Saturation analysis of ST-seq and scRNA-seq data.

(A) Saturation analysis of ST-seq data of InS group. (B) Saturation analysis of ST-seq data of DeC group. (C) Saturation analysis of scRNA-seq data of InS group. (B) Saturation analysis of scRNA-seq data of DeC group.

Supplementary. Fig. 2 Projection of ST-seq and scRNA-seq data of InS group on the gill tissue

(A) The overall projection results of ST-seq and scRNA-seq data with cryosection (bright field) and FISH (merged image by DAPI, Cy3-Eub338 and FITC-pmoB) of gill tissue. (B-F) Magnified results of the projection. (G) Population census of gill cells in InS group from the scRNA-seq data

Supplementary. Fig. 3 Expression pattern of metabolic pathways in methanotrophic symbiont and deep-sea mussel host.

(A) Metabolic pathway of steroid biosynthesis in symbiont revealed by the genome (Left) and meta-transcriptome (Right). For the genome information, all identified genes were colored in blue, yellow or orange. For the transcriptome information, all identified genes were colored in green while the abundantly expressed genes (Top 10%) were colored in pink. (B) Metabolic pathway of steroid biosynthesis in host revealed by the genome (Left) and meta-transcriptome (Right). For the transcriptome

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.28.493830; this version posted May 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

information, only the abundantly expressed genes (Top 10%) were colored in green. (C) Metabolic pathway of carbohydrate metabolism in symbiont revealed by the genome and meta-transcriptome. (D) Metabolic pathway of carbohydrate metabolism in host revealed by the genome and meta-transcriptome. (E) Metabolic pathway of nitrogen metabolism in symbiont revealed by the genome and meta-transcriptome.

Supplementary. Fig. 4 Phagocytosis of deep-sea mussel bactericytes.

(A) TEM image showing the engulfment of symbionts (labeled with star) by bacteriocytes. (B) Phagocytosis assay of bacteriocytes using FITC-labeled *Vibrio alginolyticus*. Cell membrane was stained by DiI (red) and cell nuclei were stained by DAPI (blue). (C) Phagocytosis assay of bacteriocytes using FITC-labeled beads.

Supplementary. Fig. 5 Phagocytosis of deep-sea mussel bactericytes.

(A) GO enrichment result of the 112 element maker genes in brown module. (B) Expression pattern of cycle-related genes and cell proliferation- and differentiation-related genes in stem-like cells. (C) Expression pattern of transcription factors and nuclear receptors in different states of mature bacteriocytes.

WITHDRAWN
see manuscript DOI for details

Supplementary. Fig. 6 Phylogenetic analysis of transcription factors.

(A) Phylogenetic analysis of EZH2, CD22 and prosaposin proteins. (B) Phylogenetic analysis of ELF-3 proteins. (C) Phylogenetic analysis of OASGRF2 proteins. (D) Phylogenetic analysis of ZNF271 and AUST2 proteins.

Supplementary. Table. 1 Stastics of ST-seq and scRNA-seq

Supplementary. Table. 2 Cell markers of the scRNA-seq

Supplementary. Table. 3 Differentially expressed markers of the scRNA-seq between DeC and InS group

Supplementary. Table. 4 Meta-transcriptome of host and symbiont

Supplementary. Table. 5 WGCNA sample expression pattern

Supplementary. Table. 6 WGCNA networks for bacteriocyte lineages

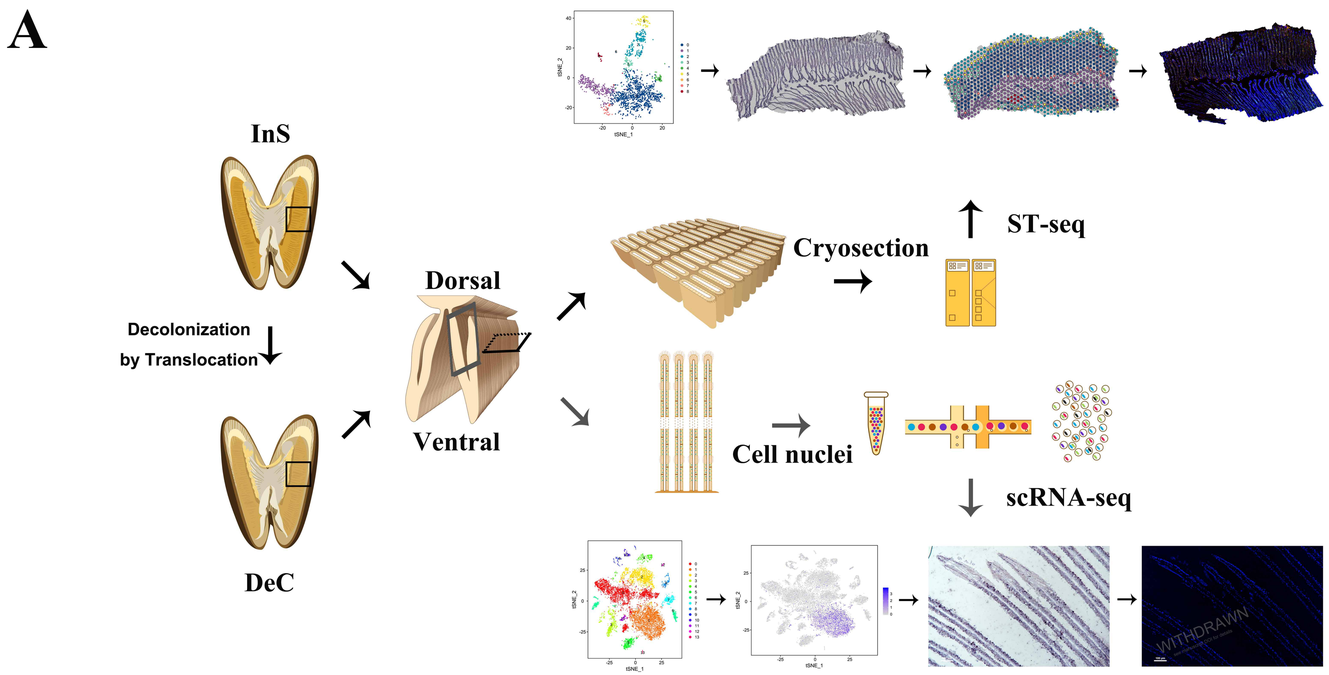
Supplementary. Table. 7 GO enrichment of WGCNA markers and all markers of mature bacteriocyte

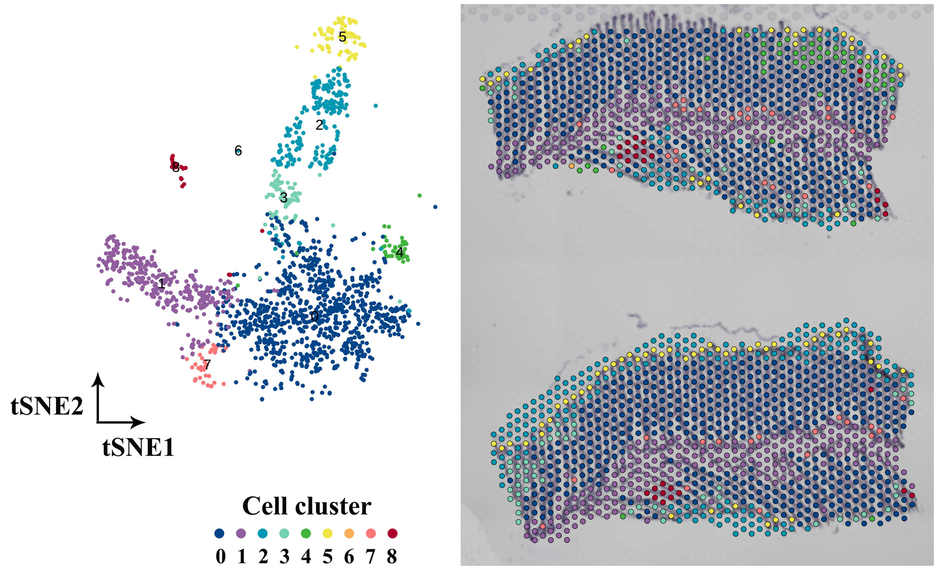
Supplementary. Table. 8 State-related differentially expressed genes of the

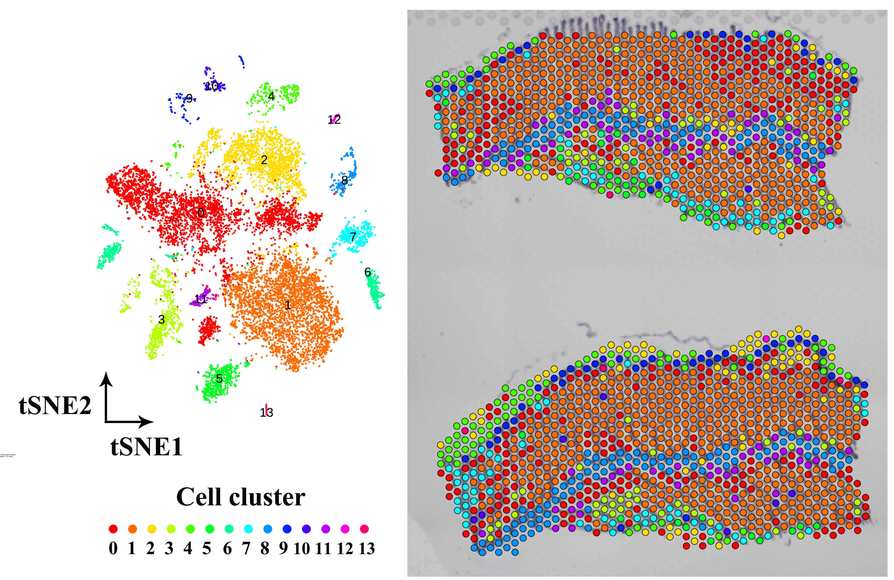
bioRxiv preprint doi: https://doi.org/10.1101/2022.05.28.493830; this version posted May 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

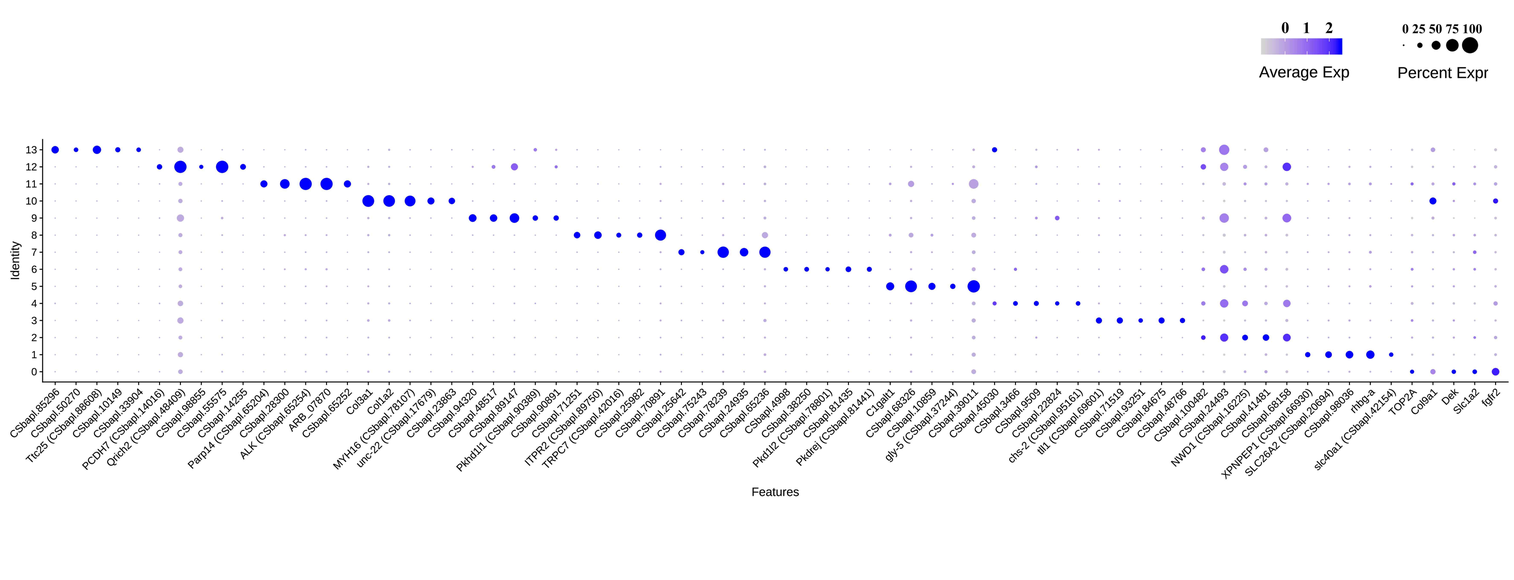
stem-like cell and bacteriocyte lineages

WITHDRAWN
see manuscript DOI for details

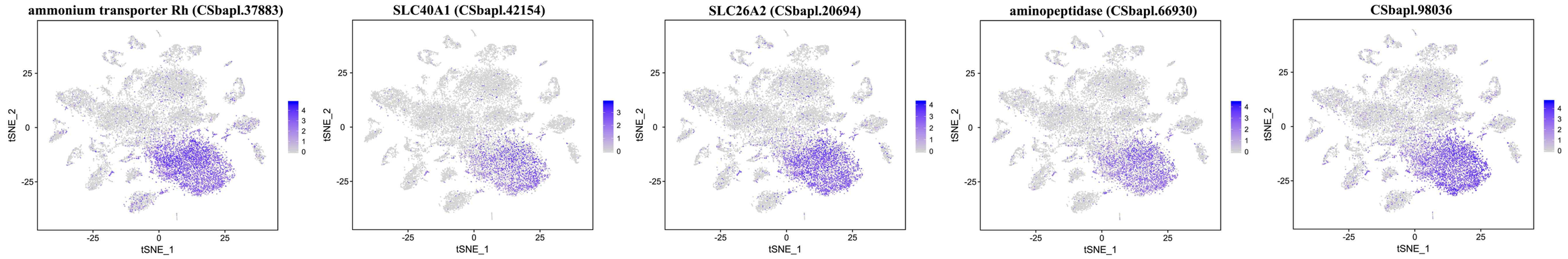

A
InS
DeC
Decolonization
by Translocation
Dorsal
Ventral
Cryosection
ST-seq
Cell nuclei
scRNA-seq
WITHDRAWN


B
tSNE2
tSNE1
Cell cluster
0 1 2 3 4 5 6 7 8

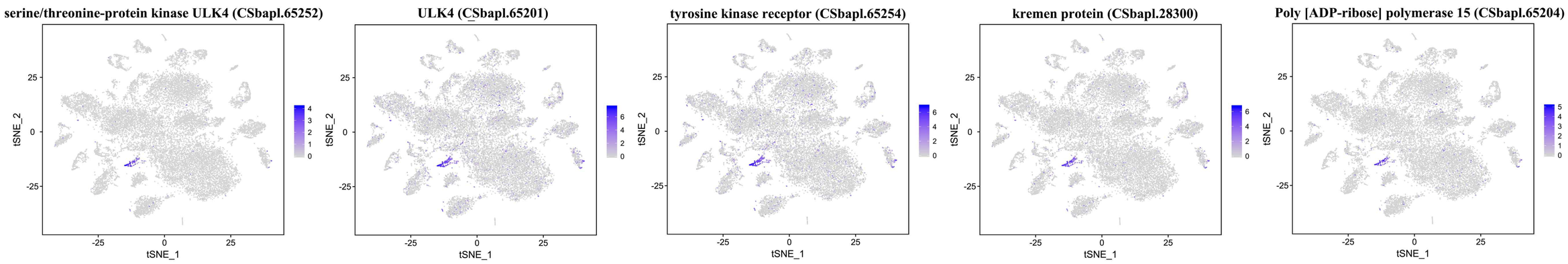

C
tSNE2
tSNE1
Cell cluster
0 1 2 3 4 5 6 7 8 9 10 11 12 13


D
Average Exp
Percent Expr
Identity
Features

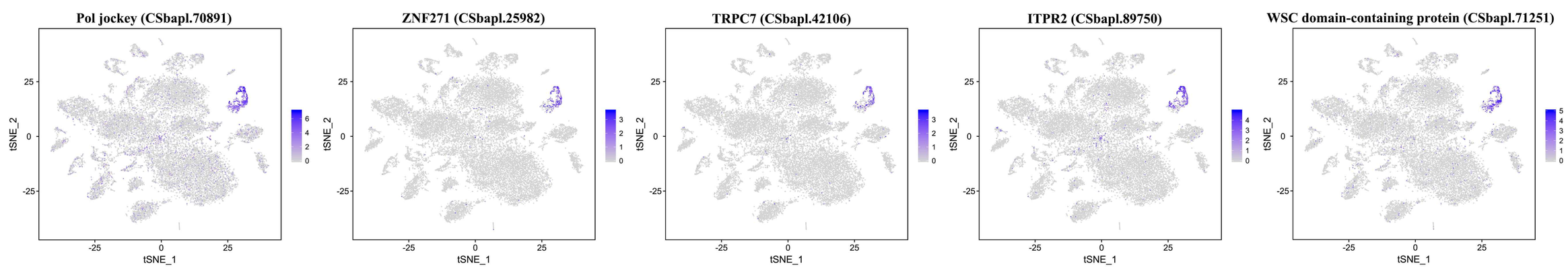
A


ammonium transporter Rh (CSbapl.37883)
SLC40A1 (CSbapl.42154)
SLC26A2 (CSbapl.20694)
aminopeptidase (CSbapl.66930)
CSbapl.98036
tSNE_1
tSNE_2

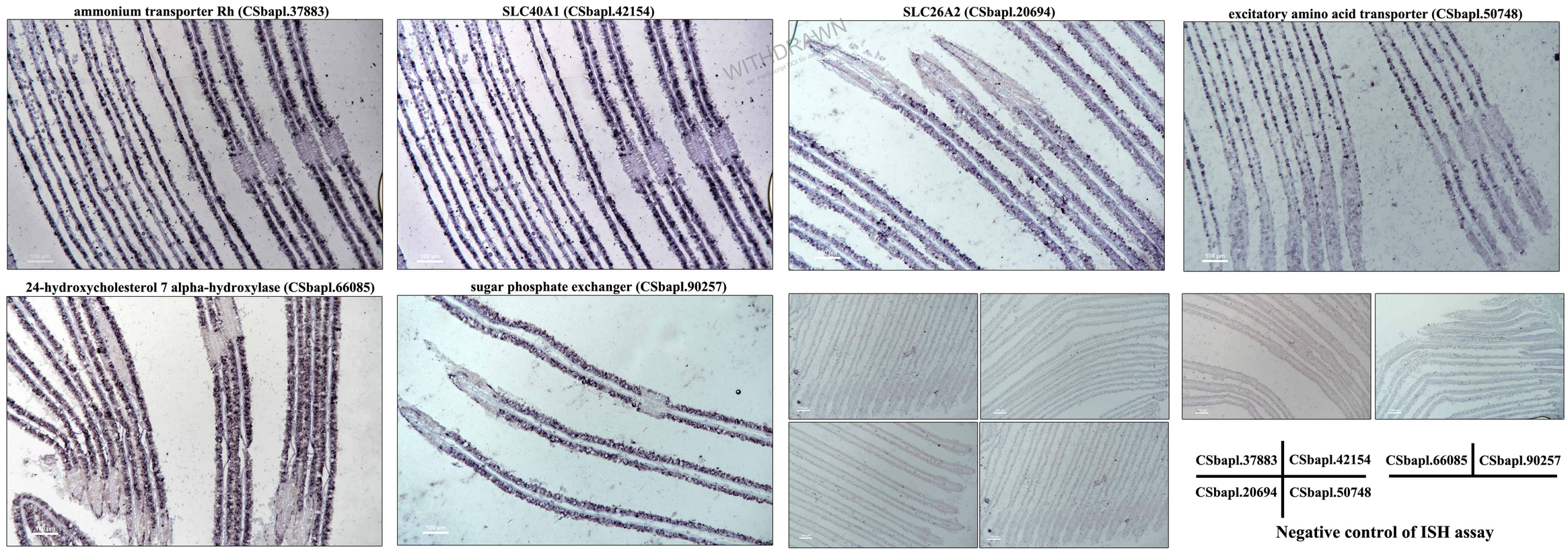
B


serine/threonine-protein kinase ULK4 (CSbapl.65252)
ULK4 (CSbapl.65201)
tyrosine kinase receptor (CSbapl.65254)
kremen protein (CSbapl.28300)
Poly [ADP-ribose] polymerase 15 (CSbapl.65204)
tSNE_1
tSNE_2

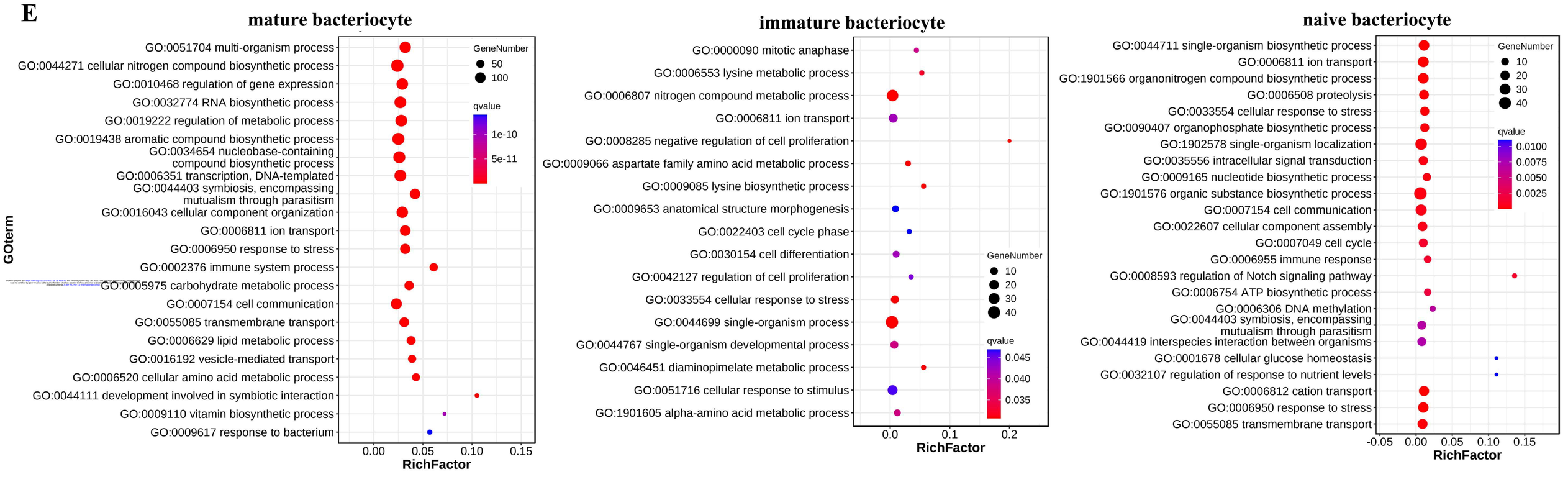
C

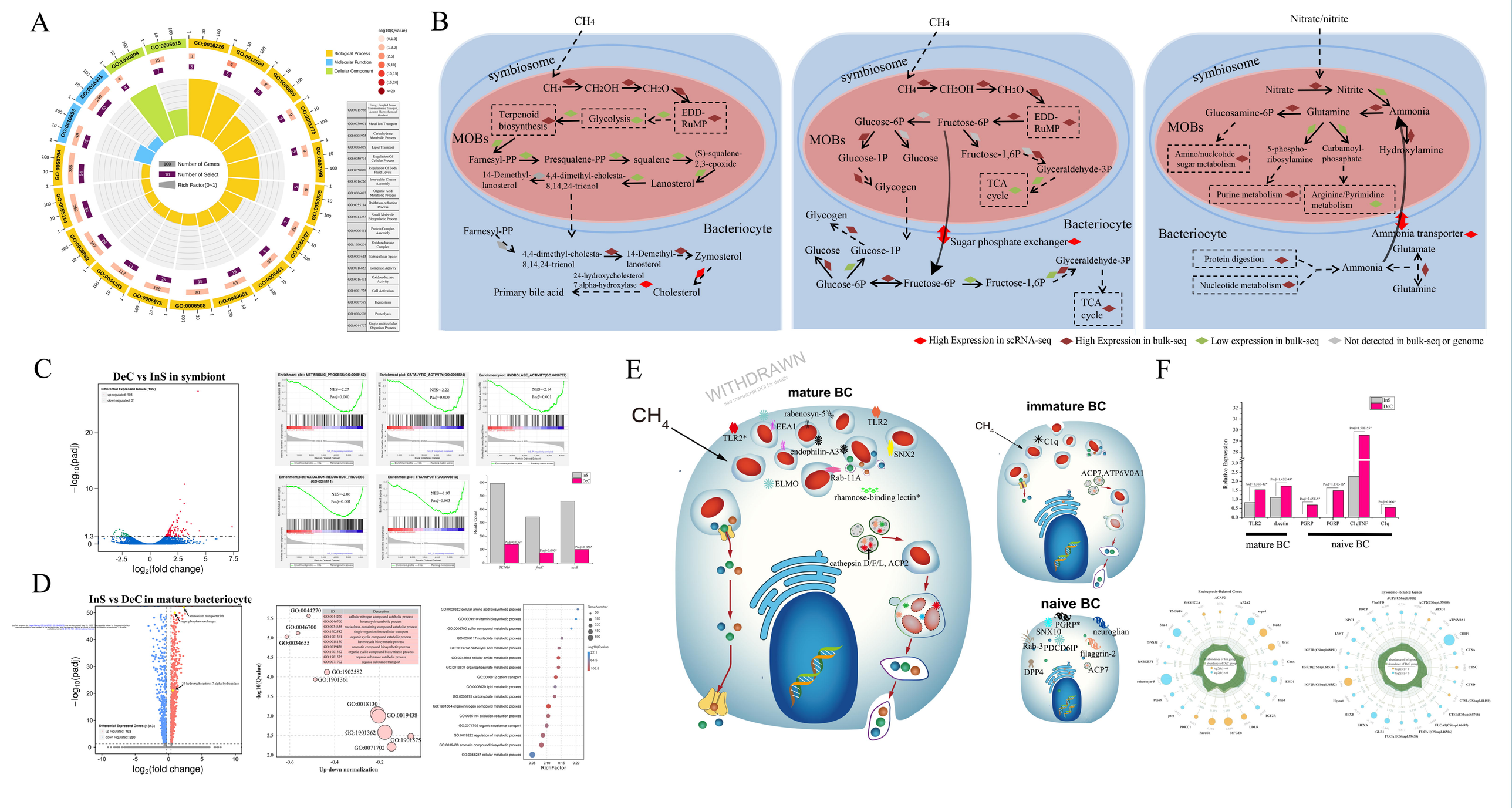

Pol jockey (CSbapl.70891)
ZNF271 (CSbapl.25982)
TRPC7 (CSbapl.42106)
ITPR2 (CSbapl.89750)
WSC domain-containing protein (CSbapl.71251)
tSNE_1
tSNE_2

D

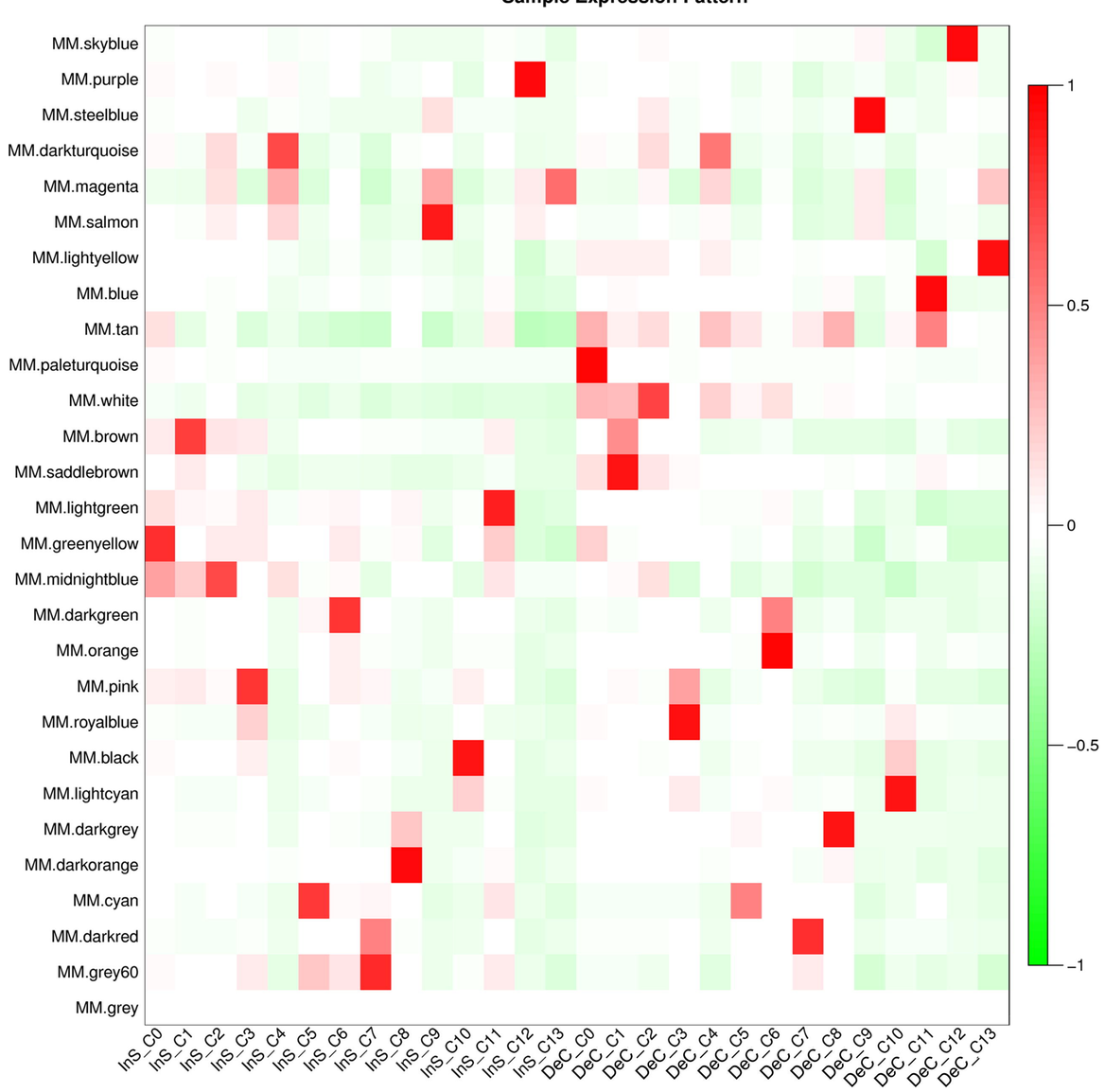

ammonium transporter Rh (CSbapl.37883)
SLC40A1 (CSbapl.42154)
SLC26A2 (CSbapl.20694)
excitatory amino acid transporter (CSbapl.50748)
24-hydroxycholesterol 7 alpha-hydroxylase (CSbapl.66085)
sugar phosphate exchanger (CSbapl.90257)
CSbapl.37883
CSbapl.42154
CSbapl.66085
CSbapl.90257
CSbapl.20694
CSbapl.50748
Negative control of ISH assay

E

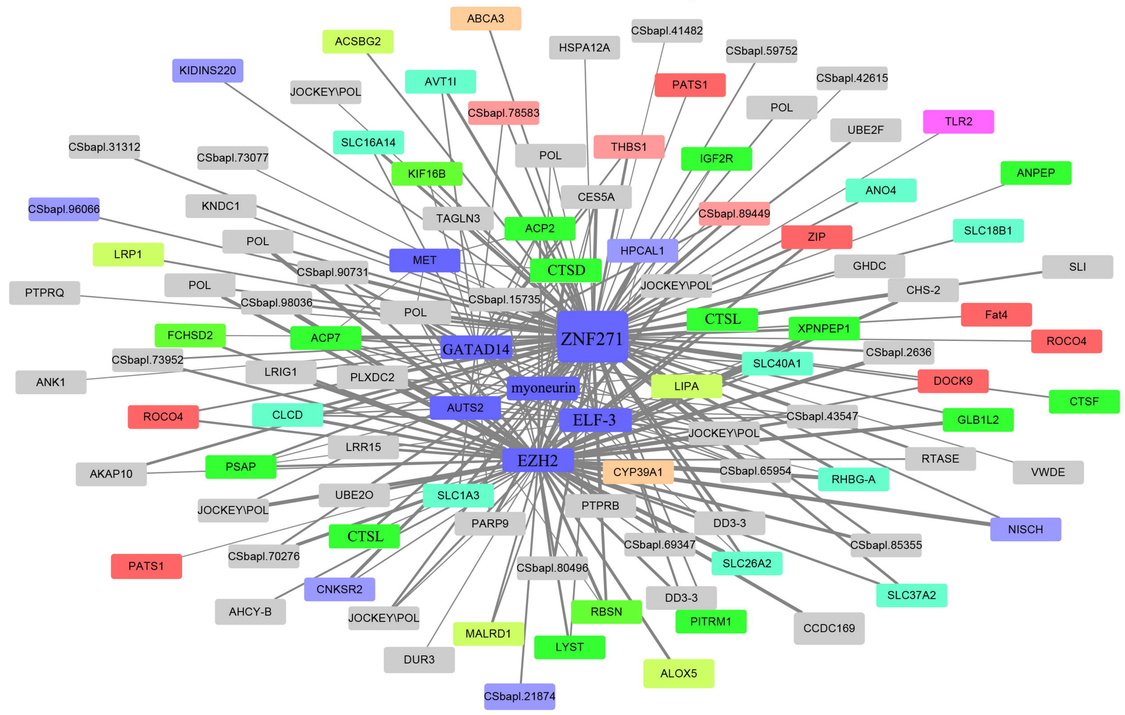

mature bacteriocyte
GO:0051704 multi-organism process
GO:0044271 cellular nitrogen compound biosynthetic process
GO:0010468 regulation of gene expression
GO:0032774 RNA biosynthetic process
GO:0019222 regulation of metabolic process
GO:0019438 aromatic compound biosynthetic process
GO:0034654 nucleobase-containing compound biosynthetic process
GO:0006351 transcription, DNA-templated
GO:0044403 symbiosis, encompassing mutualism through parasitism
GO:0016043 cellular component organization
GO:0006811 ion transport
GO:0006950 response to stress
GO:0002376 immune system process
GO:0005975 carbohydrate metabolic process
GO:0007154 cell communication
GO:0055085 transmembrane transport
GO:0006629 lipid metabolic process
GO:0016192 vesicle-mediated transport
GO:0006520 cellular amino acid metabolic process
GO:0044111 development involved in symbiotic interaction
GO:0009110 vitamin biosynthetic process
GO:0009617 response to bacterium
GeneNumber
50
100
qvalue
1e-10
5e-11
GOterm
RichFactor
immature bacteriocyte
GO:0000090 mitotic anaphase
GO:0006553 lysine metabolic process
GO:0006807 nitrogen compound metabolic process
GO:0006811 ion transport
GO:0008285 negative regulation of cell proliferation
GO:0009066 aspartate family amino acid metabolic process
GO:0009085 lysine biosynthetic process
GO:0009653 anatomical structure morphogenesis
GO:0022403 cell cycle phase
GO:0030154 cell differentiation
GO:0042127 regulation of cell proliferation
GO:0033554 cellular response to stress
GO:0044699 single-organism process
GO:0044767 single-organism developmental process
GO:0046451 diaminopimelate metabolic process
GO:0051716 cellular response to stimulus
GO:1901605 alpha-amino acid metabolic process
GeneNumber
10
20
30
40
qvalue
0.045
0.040
0.035
RichFactor
naive bacteriocyte
GO:0044711 single-organism biosynthetic process
GO:0006811 ion transport
GO:1901566 organonitrogen compound biosynthetic process
GO:0006508 proteolysis
GO:0033554 cellular response to stress
GO:0090407 organophosphate biosynthetic process
GO:1902578 single-organism localization
GO:0035556 intracellular signal transduction
GO:0009165 nucleotide biosynthetic process
GO:1901576 organic substance biosynthetic process
GO:0007154 cell communication
GO:0022607 cellular component assembly
GO:0007049 cell cycle
GO:0006955 immune response
GO:0008593 regulation of Notch signaling pathway
GO:0006754 ATP biosynthetic process
GO:0006306 DNA methylation
GO:0044403 symbiosis, encompassing mutualism through parasitism
GO:0044419 interspecies interaction between organisms
GO:0001678 cellular glucose homeostasis
GO:0032107 regulation of response to nutrient levels
GO:0006812 cation transport
GO:0006950 response to stress
GO:0055085 transmembrane transport
GeneNumber
10
20
30
40
qvalue
0.0100
0.0075
0.0050
0.0025
RichFactor


A
-log10(Qvalue)
Biological Process
Molecular Function
Cellular Component
Number of Genes
Number of Select
Rich Factor(0~1)
B
CH4
symbiosome
CH4 → CH2OH → CH2O
Terpenoid biosynthesis
Glycolysis
EDD-RuMP
MOBs
Farnesyl-PP → Presqualene-PP → squalene → (S)-squalene-2,3-epoxide
14-Demethyl-lanosterol
4,4-dimethyl-cholesta-8,14,24-trienol
Lanosterol
Bacteriocyte
Farnesyl-PP
4,4-dimethyl-cholesta-8,14,24-trienol
14-Demethyl-lanosterol
Zymosterol
24-hydroxycholesterol
7 alpha-hydroxylase
Primary bile acid
Cholesterol
Glucose-6P
Fructose-6P
Glucose-1P
Glucose
Fructose-1,6P
Glyceraldehyde-3P
Glycogen
TCA cycle
Sugar phosphate exchanger
Nitrate/nitrite
Nitrate
Nitrite
Glucosamine-6P
Glutamine
Ammonia
Amino/nucleotide sugar metabolism
5-phospho-ribosylamine
Carbamoyl-phosphate
Hydroxylamine
Purine metabolism
Arginine/Pyrimidine metabolism
Ammonia transporter
Glutamate
Protein digestion
Nucleotide metabolism
Glutamine
High Expression in scRNA-seq
High Expression in bulk-seq
Low expression in bulk-seq
Not detected in bulk-seq or genome
C
DeC vs InS in symbiont
Differential Expressed Genes (135)
up regulated: 104
down regulated: 31
log2(fold change)
Enrichment plot: METABOLIC_PROCESS(GO:0008152)
NES=-2.27
Padj=0.000
Enrichment plot: CATALYTIC_ACTIVITY(GO:0003824)
NES=-2.22
Padj=0.000
Enrichment plot: HYDROLASE_ACTIVITY(GO:0016787)
NES=-2.14
Padj=0.001
Enrichment plot: OXIDATION-REDUCTION_PROCESS (GO:0055114)
NES=-2.06
Padj=0.001
Enrichment plot: TRANSPORT(GO:0006810)
NES=-1.97
Padj=0.003
InS
DeC
Reads Count
Padj=0.026*
Padj=0.040*
Padj=0.026*
TK1436
fusC
accB
D
InS vs DeC in mature bacteriocyte
Differential Expressed Genes (1343)
up regulated: 793
down regulated: 550
Ammonium transporter Rh
Sugar phosphate exchanger
24-hydroxycholesterol 7 alpha-hydroxylase
log2(fold change)
GO:0044270
GO:0046700
GO:0034655
GO:1902582
GO:1901361
GO:0018130
GO:0019438
GO:1901362
GO:1901575
GO:0071702
Up-down normalization
GO:0008652 cellular amino acid biosynthetic process
GO:0009110 vitamin biosynthetic process
GO:0006790 sulfur compound metabolic process
GO:0009117 nucleotide metabolic process
GO:0019752 carboxylic acid metabolic process
GO:0043603 cellular amide metabolic process
GO:0019637 organophosphate metabolic process
GO:0006812 cation transport
GO:0006629 lipid metabolic process
GO:0005975 carbohydrate metabolic process
GO:1901564 organonitrogen compound metabolic process
GO:0055114 oxidation-reduction process
GO:0071702 organic substance transport
GO:0019222 regulation of metabolic process
GO:0019438 aromatic compound biosynthetic process
GO:0044237 cellular metabolic process
GeneNumber
RichFactor
E
WITHDRAWN
see manuscript DOI for details
mature BC
CH4
TLR2*
rabenosyn-5
EEA1
TLR2
endophilin-A3
SNX2
ELMO
Rab-11A
rhamnose-binding lectin*
cathepsin D/F/L, ACP2
immature BC
CH4
C1q
ACP7,ATP6V0A1
naive BC
PGRP*
SNX10
neuroglian
Rab-3
PDCD6IP
filaggrin-2
DPP4
ACP7
F
InS
DeC
Relative Expression
Padj=4.36E-12*
Padj=1.43E-43*
Padj=2.65E-5*
Padj=1.15E-16*
Padj=1.50E-55*
Padj=0.006*
TLR2
rl.ectin
PGRP
PGRP
C1qTNF
C1q
mature BC
naive BC
Endocytosis-Related Genes
Lysosome-Related Genes

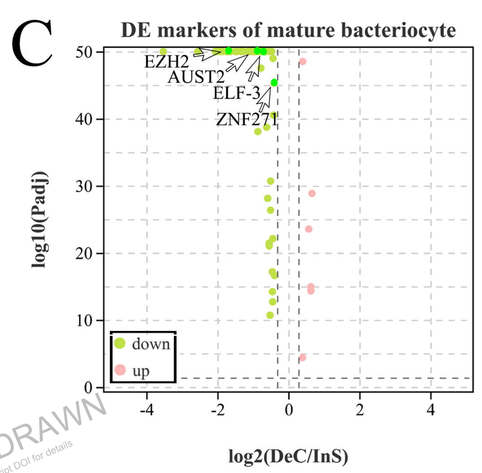
A



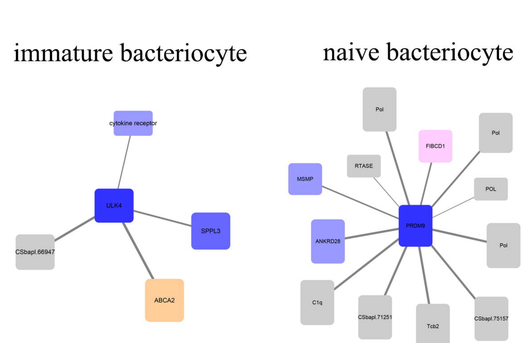
B



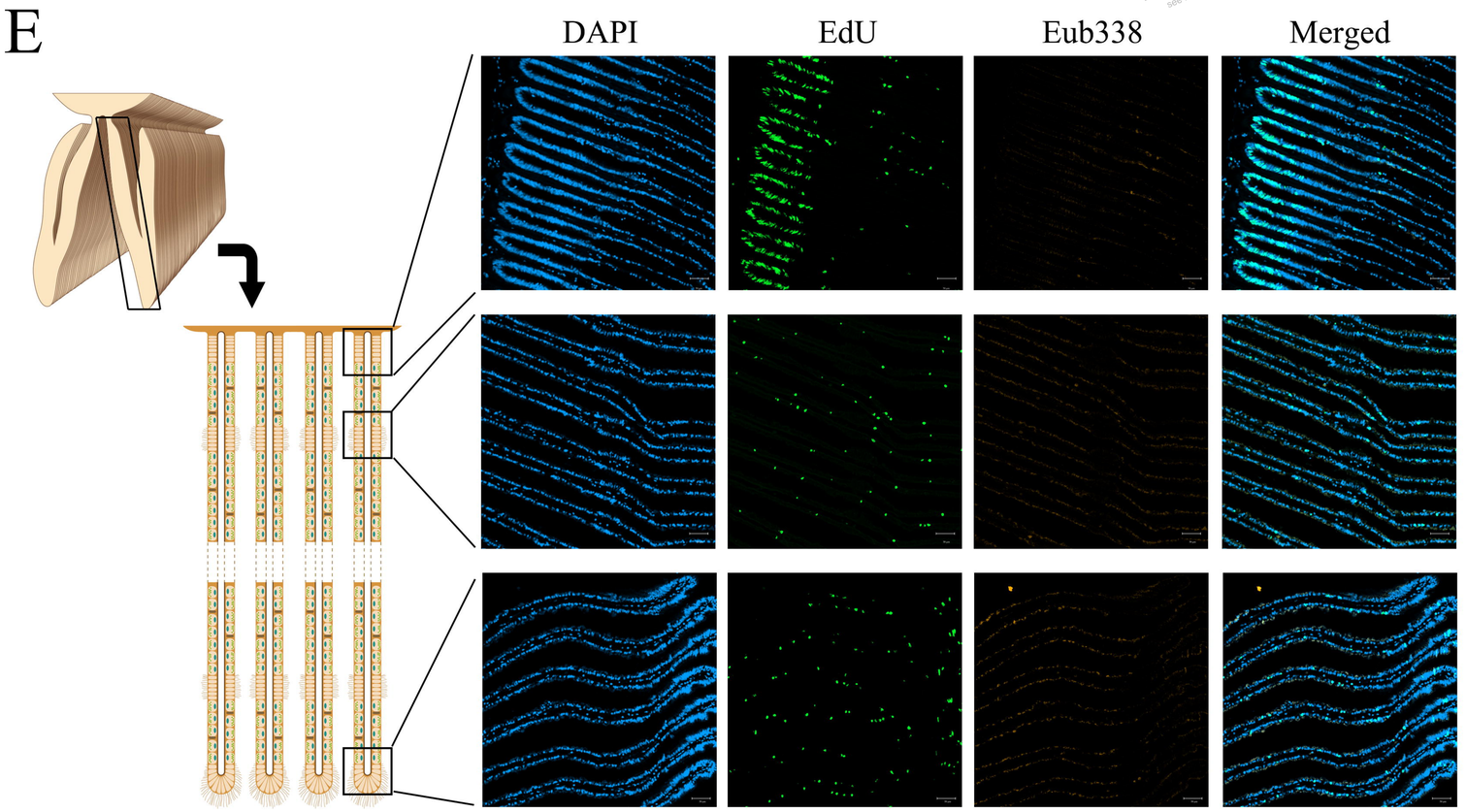
C



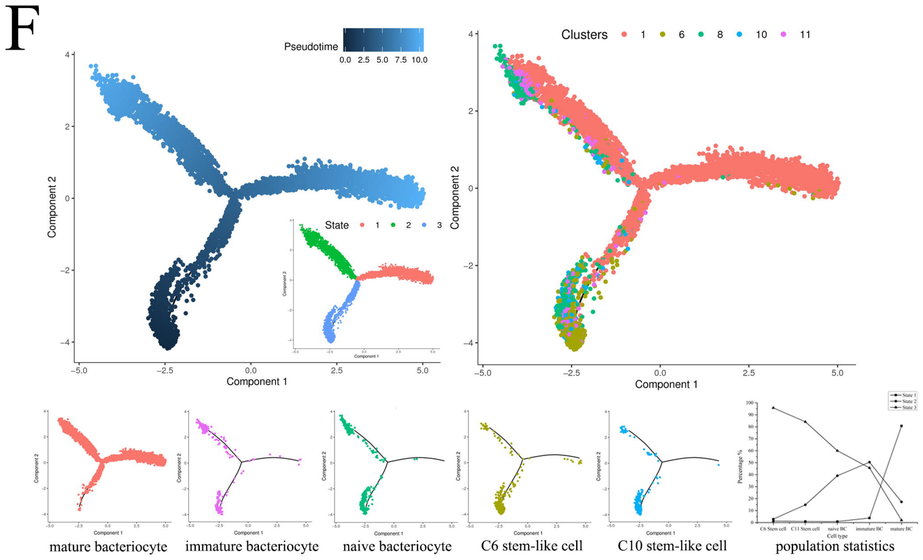
D



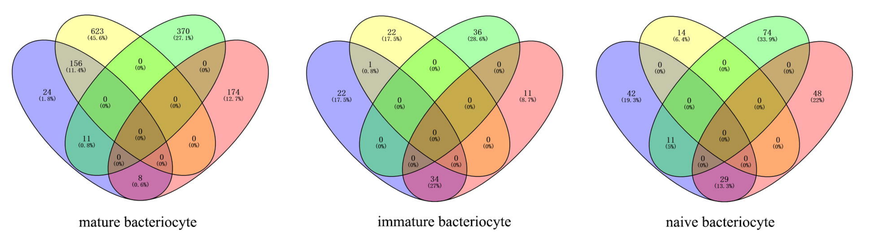
E



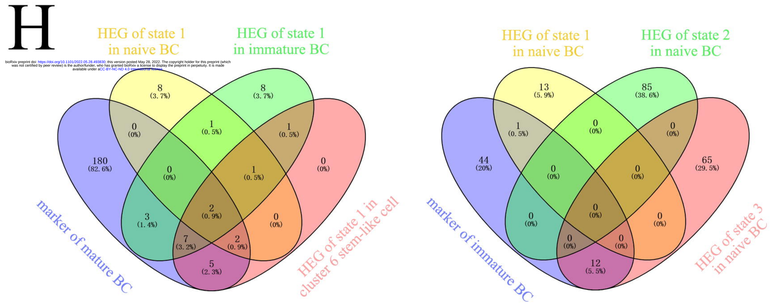
F



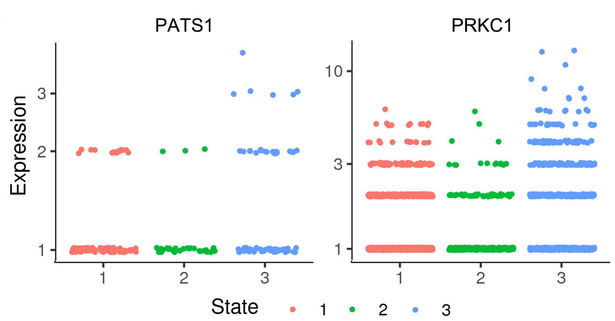
G



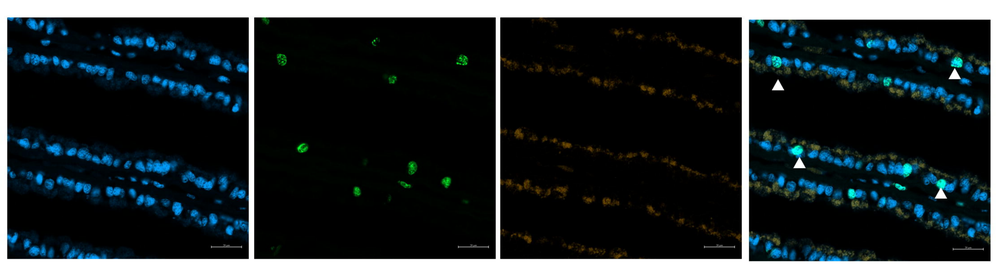
H



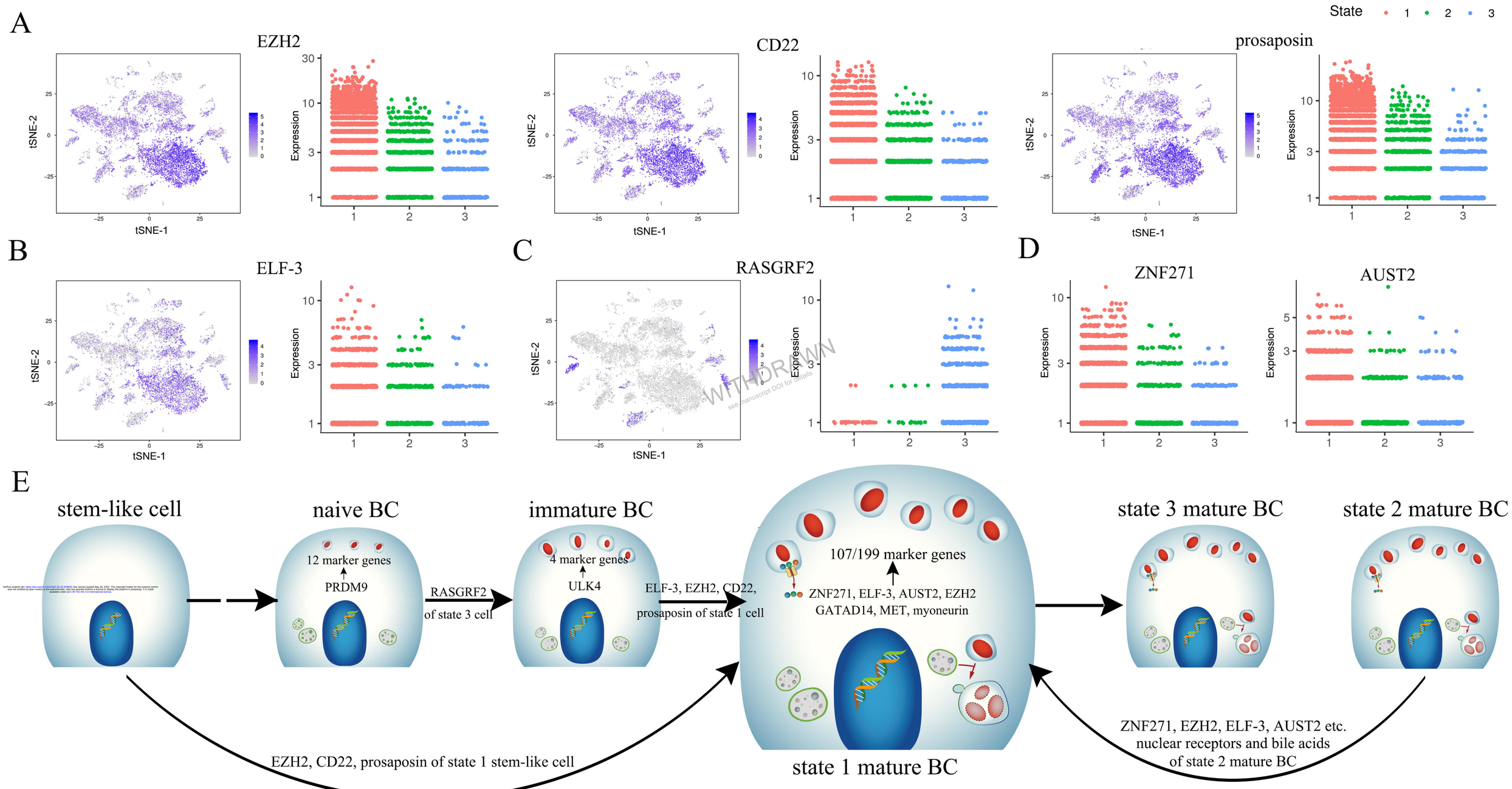
I



J



WITHDRAWN



WITHDRAWN
see manuscript DOI for details