Haematopoiesis is an essential process that evolved in multicellular animals. At the heart of this process are haematopoietic stem cells (HSCs), which are multipotent and self-renewing, and generate the entire repertoire of blood and immune cells throughout an animal’s life. Although there have been comprehensive studies on self-renewal, differentiation, physiological regulation and niche occupation in vertebrate HSCs, relatively little is known about the evolutionary origin and niches of these cells. Here we describe the haematopoietic system of Botryllus schlosseri, a colonial tunicate that has a vasculature and circulating blood cells, and investigating stem-cell biology and immunity characteristics. Self-recognition between genetically compatible B. schlosseri colonies leads to the formation of natural parabions with shared circulation, whereas incompatible colonies reject each other. Using flow cytometry, whole-transcriptome sequencing of defined cell populations and diverse functional assays, we identify HSCs, progenitors, immune effector cells and an HSC niche, and demonstrate that self-recognition inhibits allo-specific cytotoxic reactions. Our results show that HSC and myeloid lineage immune cells emerged in a common ancestor of tunicates and vertebrates, and also suggest that haematopoietic bone marrow and the B. schlosseri endostyle niche evolved from a common origin.

Charles Darwin recognized that the study of tunicates is critical to understanding the evolution of vertebrates, and tunicates were later discovered to be a sister group of vertebrates. To gain insight into the evolution of the mammalian haematopoietic system, we characterized the haematopoietic and immune system in the colonial tunicate B. schlosseri.

B. schlosseri colonies produce genetically identical individuals (zooids) through stem-cell-mediated cyclical budding. Each week, developing buds replace their parent zooids, which then undergo synchronized programmed cell death. When colonies touch, their extracorporeal vasculatures either fuse or reject. This self–nonself recognition process is controlled by the highly polymorphic histocompatibility gene BHF, and at least one shared BHF allele is required for fusion to take place. We adapted fluorescence-activated cell sorting (FACS) to separate B. schlosseri cells and isolated 34 cell populations using size, granularity, natural auto-fluorescence, reagents such as antibodies that bind differentially to live cells (CD49d, CD57 and BHF), concanavalin-A, and alkaline phosphatase expression. We sequenced the transcriptomes of 23 sorted cell populations, the hierarchical endpoint populations of our FACS gating strategy, and found correlations between gene expression profiles, morphology and marker expression. In the cluster of cell populations CP25, CP33 and CP34, there were 235 genes that were differentially upregulated and are known to be expressed in vertebrate blood and haematopoietic systems.

To measure the ability of these candidate HSC populations (cHSCs) to differentiate into other cell types, cells were transplanted from orange colonies of genetically compatible B. schlosseri into red colonies of genetically incompatible colonies. Arrows point to fused vasculature and points of rejection (bottom).
To identify the cHSC niches, the cHSC population and a control population (CP3) were isolated, labelled with a lipophilic dye (DiD) and injected into compatible colonies labelled with carboxyfluorescein succinimidyl ester (CFSE). Five to ten days after transplantation, DiD-labelled cHSC populations migrated into the recipient colony and aggregated in two known B. schlosseri stem-cell niches: the endostyle and the cell islands. The endostyle niche, located at the anterior subendostylar sinus (endo-niche), has been identified as a somatic stem-cell niche, and the cell islands niche is considered a germline stem-cell niche.\(^6,16\) (Fig. 2g, Extended Data Fig. 5b). Proliferating cells and haemoblasts have been identified near the endostyle and around the branchial sac's stigmata in juvenile Ciona intestinalis.\(^17\) Indeed, blood cells with a haemoblast (cHSC) morphology and proliferating cells are abundant in the B. schlosseri endo-niche.\(^6\) (Fig. 2h, Extended Data Fig. 5c, d). The control population (CP3) does not express a gene signature of the haematopoietic system, but does have a germline gene expression signature and is localized to a known germline stem-cell niche,\(^16\) the cell islands (Extended Data Figs. 4c, 5b). The B. schlosseri CHSCs localized towards the endo-niche, similar to the homing process of mammalian HSCs to the bone marrow.\(^18-20\) The endostyle is a complex tissue with defined anatomical structures and molecular features.\(^5,16,21-24\) We sequenced the transcriptomes of ten endostyles and compared them to the transcriptomes of whole colonies (n = 34). Homologous genes that were
significantly upregulated (P < 0.05) in the endostyle were analysed by GeneAnalytics, revealing shared expression of 327 genes between the endostyle and human hematopoietic bone marrow (Fig. 2i, left, Extended Data Fig. 6a, Supplementary Tables 2, 4). This finding was further supported by Gene Expression Commons (Extended Data Fig. 6b). We queried previously obtained Ciona robusta in situ expression data for endostyle-associated genes and found them to be similar (Fig. 2i, right, Extended Data Fig. 6c).

In mammals, innate cellular immune responses are mediated in part by phagocytosis (the engulfment of target cells), and both adaptive and innate immune responses in part by cellular cytotoxicity (the direct killing of target cells). We used diverse ex vivo phagocytosis assays to identify phagocytic cells and track their cell populations (Extended Data Fig. 7). These assays revealed three major phagocytic populations, including two previously described phagocytic cells: amoebocytes (within CP4 and CP18), large phagocytes (within CP13) and a previously undescribed population, the candidate myeloid cell population (within CP7 and CP10) (Extended Data Fig. 7). The myeloid cells were the main contributors to phagocytosis (they contributed more than 40% to each of the phagocytosis assays), whereas the large phagocytes contributed mainly to allogeneic phagocytosis (Extended Data Fig. 7a).

We did not find a cell population with a clear mammalian cytotoxic gene expression signature. Morula cells, which contain phenoloxidase, accumulate at rejection points and have been proposed to be cytotoxic cells that mediate rejection. We detected morula cells in population CP18, but in lower levels than expected, prompting us to look for a candidate precursor cell. We studied the large granular lymphocyte-like (LGL) cells (enriched in CP31) as a potential candidate, because their morphology resembles that of natural killer cells, which mediate cytotoxicity as part of the mammalian innate immune response. Ex vivo experiments revealed that purified LGL cells transitioned into morula cells after two days (Fig. 3a, labelled 1–4), suggesting that LGL cells are cytotoxic morula cells that become pigmented granular morula cells following activation.

Assays that compared the ability of morula and LGL cells to induce cytotoxicity showed that isolated LGL cells were significantly (P < 0.005) more cytotoxic than isolated morula cells or the control population (CP3; Extended Data Fig. 8a). Overnight incubations of isolated LGL cells in either a syngeneic (self) or an allogeneic (nonself) challenge led to a transition of 60% of the LGL cells that were incubated with allogeneic cells and an increase in morula cells. Only about 10% of the LGL cells that were incubated with syngeneic cells became morula cells (Extended Data Fig. 8b). Upon activation, LGL cells change their morphology, develop granularity and pigmentation, presumably owing to phenoloxidase activation, and become morula cells (Fig. 3a, Extended Data Fig. 8b, c). This set of experiments demonstrates that the LGL cells are the cytotoxic cell population, and therefore we call LGL cells cytotoxic morula cells from here onwards.

Analysis of the genes that were differentially upregulated by the highly enriched cytotoxic morula cell population (CP31) revealed 52 unannotated genes with no human or mouse homologues. Twenty-one of these genes carry domains that are associated with functions of cytotoxic cells such as recognition or lysis, and C-type lectin, a domain contained in human natural killer cell receptors (Supplementary Table 5). Among the 18 differentially upregulated genes that have sequence homology to genes in vertebrates, 14 are associated with at least one of the following functions: cellular recognition, cytotoxicity or peptidase activity, leukocyte homing and general immune response (Fig. 3b). CP31 cells also express tyrosinase-associated gene (TYRP2), the vertebrate parologue gene to phenoloxidase, which is one of the main proteins found in morula cells (present at sevenfold-higher levels than in other cells). Although we did not find a B. schlosseri cell population that had a significant lymphoid lineage signature, GeneSet Activity Analysis suggested that population CP19—which is enriched in morula cells—resembles mouse T cells, B cells, and immature natural killer cells in gene expression (Extended Data Fig. 4d). This list does not include any gene that is associated with known adaptive immunity function (Supplementary Table 2).

To test whether allogeneic cells interact in vivo during rejection, whole colonies were differentially labelled, set near each other and monitored by live imaging (Fig. 3d, Supplementary Video 3). We detected direct contact between allogeneic cells during rejection within the points of rejection (PORs; Fig. 3d, lower right). Next, we performed allogeneic phagocytosis and cytotoxicity assays. Assays
cells by the major histocompatibility complex, self-BHF recognition inhibits cytotoxicity. These results—and the observation that fusion occurs when colonies share at least one BHF allele—demonstrate that, as in natural killer cell recognition, the cellular cytotoxicity mechanism in *B. schlosseri* is based on ’missing self’.

The haematopoietic and immune systems of *B. schlosseri* combine features of vertebrates and invertebrates (Fig. 4). *B. schlosseri* has cells that share morphological and molecular characteristics with the vertebrate HSC and myeloid lineages, including cells that take part in phagocytosis. It also has amoebocytes and large phagocytes with morphologies resembling invertebrate cell types. The cytotoxic morula cells of *B. schlosseri* carry imprints reminiscent of vertebrate lymphocytes, but mainly express tunicate-specific genes (Fig. 3b, Extended Data Table 4d). Several studies have described morula-like cells with phenoloxidase activity in other invertebrate species. Our analysis of gene sets that are expressed by the cytotoxic morula cells and identification of the BHF inhibition pathway will be likely to reveal novel mechanisms to delimit self from non-self and target pathogens.

### Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-018-0783-x.

Received: 1 December 2017; Accepted: 15 October 2018; Published online 5 December 2018.

---

Fig. 4 | Evolution of the cellular immune system. Information regarding the cellular immune systems of invertebrate and vertebrate species. The table to the right shows the types of immune-associated cell found in each animal. Colonial tunicates contain immune system cells found in both invertebrate and vertebrate species. Whereas amoebocytes and large phagocytes are found in *B. schlosseri* and other invertebrate species (yellow and orange), myeloid lineage phagocytic cells have so far been found only in *B. schlosseri* and vertebrates (red). Cytotoxic morula cells have been identified in *B. schlosseri* and are likely to exist in a few other invertebrate species (turquoise), but have not been identified in vertebrates. Although a classic lymphoid lineage has been found only in vertebrates (blue), there is a cellular population and additional molecular analysis suggesting the existence of mainly undifferentiated lymphoid cells in *B. schlosseri*. The cellular allorecognition molecules identified in each species are shown in the right-hand column: immunoglobulin superfamily major histocompatibility complex (MHC), killer inhibitory receptors (KIRs) and T cell receptors (TCRs), leucine-rich repeats receptors of the variable lymphocyte receptors (VLRs), and BHF. Question marks represent missing functional or molecular validation. Data source summarized in Extended Data Table 2b.
22. Cañestro, C., Bassham, S. & Postlethwait, J. H. Evolution of the thyroid: anterior-posterior regionalization of the Oikopleura endostyle revealed by Otx, Pax2/5/8, and Hox1 expression. Dev. Dyn. 237, 1490–1499 (2008).

23. Burighel, P. & Cloney, R. A. in Microscopic Anatomy of Invertebrates (eds Harrison, F. W. & Ruppert, E. E.) 221–347 (Alan R. Liss, 1997).

24. Ogasawara, M., Lauro, R. D. & Satoh, N. Ascidian homologs of mammalian thyroid transcription factor-1 gene are expressed in the endostyle. Zool. Sci. 16, 559–565 (1999).

25. Chen, J. Y. et al. Hoxb5 marks long-term haematopoietic stem cells and reveals a homogenous perivascular niche. Nature 530, 223–227 (2016).

26. Ballarin, L. & Cima, F. Cytochemical properties of Botryllus schlosseri haemocytes: indications for morpho-functional characterisation. Eur. J. Histochem. 49, 255–264 (2005).

27. Lauzon, R. J., Brown, C., Kerr, L. & Tiozzo, S. Phagocytic dynamics in a highly regenerative urochordate: insights into development and host defense. Dev. Biol. 374, 357–373 (2013).

28. Franchi, N. & Ballarin, L. Immunity in protochordates: the tunicate perspective. Front. Immunol. 8, 674 (2017).

29. Oren, M. et al. Marine invertebrates cross phyla comparisons reveal highly conserved immune machinery. Immunobiology 218, 484–495 (2013).

30. Timonen, T., Ortaldo, J. R. & Herberman, R. B. Characteristics of human large granular lymphocytes and relationship to natural killer and K cells. J. Exp. Med. 153, 569–582 (1981).

31. Ljunggren, H. G. & Kärre, K. In search of the ‘missing self’: MHC molecules and NK cell recognition. Immunol. Today 11, 237–244 (1990).

32. Smith, L. C. et al. in Invertebrate Immunity (ed. Söderhäll, K.) 260–301 (Springer, 2010).

33. Kawabata, S. in Invertebrate Immunity (ed. Söderhäll, K.) 122–136 (Springer, 2010).

34. Ballarin, L. Ascidian cytotoxic cells: state of the art and research perspectives. Inv. Surv. J. 9, 1–6 (2012).

Acknowledgements We thank C. Lowe, C. Anselmi, I. Dimov, S. Karten, C. Patton, J. Thompson, P. Lovelace, R. Veskoboynik, N. Fernhoff, W.-J. Lu, P. Chu, K. Weiskopf, M. Oren, B. Wang, J. Lee, B. Compton, K. Uhlinger, T. Naik and T. Storm for technical advice and help. This study was supported by NIH grants R56AI089968, R01AG037968 and RO1GM100315 (to I.L.W., S.R.Q., and A.V.), the Virginia and D. K. Ludwig Fund for Cancer Research, a grant from the Sibbel Stem Cell Institute and a Stinehart-Reed grant (to I.L.W.). L.M. was supported by PRIN - Prot. 2015NSFHXF, B.R. was supported by a Postdoctoral Fellowship of the Human Frontier Science Program Organization LT000591-2014-L, NIH Immunology training grant 5T32AI07290-28 and NIH Hematology training grant T32 HL120824-03.

Reviewer information Nature thanks M. D. Cooper, W. Jeffery, G. Litman and J. Pascual-Anaya for their contribution to the peer review of this work.

Author contributions Conception and design: B.R., A.V., M.K., S.R.Q. and I.L.W.; mariculture: K.J.I. and K.J.P.; flow cytometry and sorting: B.R.; CyTOF screening and cluster analysis: B.R., S.-Y.C. and G.P.N.; RNA isolation and library preparation: B.R., K.J.P., R.S. and A.V.; sequencing: J.O., G.M. and N.F.N.; sequencing analysis and development of analytical tools: M.K., J.S, A.M.N. and S.R.Q.; immunological assays: B.R.; microscopy and experimental design: B.R., D.M.C., K.J.I., D.N.C. and A.V.; electron microscopy: L.M.; BHF localization to the membrane and serum validation: J.M.T., B.R. and A.V.; transplantation: B.R.; writing of manuscript: B.R., A.V., M.K., T.R., K.J.P. and I.L.W.; technical support and conceptual advice: N.F.N., D.M.C., A.M.N., T.R., G.P.N., S.R.Q., A.V. and I.L.W.

Competing interests The authors declare no competing interests.

Additional information Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0783-x.

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0783-x.

Reprints and permissions information is available at http://www.nature.com/reprints.

Correspondence and requests for materials should be addressed to B.R., I.L.W. or A.V.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Colony development, labelling and allorecognition assays. The life cycle of B. schlosseri includes both sexual and asexual reproduction pathways. Sexual reproduction starts with fertilization and progresses through classic embryonic stages into a tadpole larva with chordate characters. Embryonic development commences when the ovary and testis are formed.

Juvenile colonies were labelled with distinct fluorescent stains (CellTrace CFSE Green, Invitrogen #C34554; CellTrace FarRed, Invitrogen #C34572) and positioned in controlled fusion/rejection reactions (Fig. 3d). For allorecognition assays, naive colonies were tied to 3 mm mesh using a sterile 1-ml syringe pump (a similar procedure to murine transplantation reactions). The fusion requires at least one shared allele: upon fusion, the resultant screened markers (Extended Data Fig. 1). This self–nonself recognition process is controlled by BHF. Fusion requires at least one shared BHF allele: upon fusion, the resultant screened markers (Extended Data Fig. 1). This self–nonself recognition process is controlled by BHF. Fusion requires at least one shared allele: upon fusion, the anti-BHF serum shows that BHF is expressed in live cells (propidium iodide negative) by the anti-BHF serum indicates that BHF is expressed in live cells (propidium iodide negative) by the anti-BHF serum. We used a published protocol38 to extract RNA from B. schlosseri colonies (n = 24). Twenty-three endpoint populations were sequenced, in addition to CP2, which was composed of all live cells. CP17 was composed of dead cells and cell debris, and therefore was excluded from the population analysis. Twenty thousand cells of each sequenced population were sorted in 750 l of staining medium before FACS analysis: alkaline phosphatase (AP) Live Stain (Life Technologies A14353) with staining medium (3.3% FCS, 0.1% NaN3) and within a few days the sexually reproduced tadpoles hatched, swam to the setonlics and metamorphoses into an oozoid, with a sessile body plan (Fig. 1a). The oozoid begins a cyclical budding process of asexual reproduction, forming a colony of genetically identical zooids and buds (Fig. 1b). The colonial individuals are united under a single gelatinous tunic by a network of blood vessels, which terminate in sausage-shaped protrusions (ampullae; Fig. 1a–c). Throughout adult life, B. schlosseri regenerates its entire body every two weeks. This stem-cell-mediated cycle of development includes the formation of all body organs including the heart, respiratory system, digestive system and neural complex. Sexual reproduction starts with fertilization and progresses through classic embryonic stages into a tadpole larva with chordate characters. Embryonic development commences when the ovary and testis are formed.

Juvenile colonies were labelled with distinct fluorescent stains (CellTrace CFSE Green, Invitrogen #C34554; CellTrace FarRed, Invitrogen #C34572) and positioned in controlled fusion/rejection reactions (Fig. 3d). For allorecognition assays, naive colonies were tied to 3 mm mesh using a sterile 1-ml syringe pump (a similar procedure to murine transplantation reactions). The fusion requires at least one shared allele: upon fusion, the resultant screened markers (Extended Data Fig. 1). This self–nonself recognition process is controlled by BHF. Fusion requires at least one shared allele: upon fusion, the anti-BHF serum shows that BHF is expressed in live cells (propidium iodide negative) by the anti-BHF serum. We used a published protocol38 to extract RNA from B. schlosseri colonies (n = 24). Twenty-three endpoint populations were sequenced, in addition to CP2, which was composed of all live cells. CP17 was composed of dead cells and cell debris, and therefore was excluded from the population analysis. Twenty thousand cells of each sequenced population were sorted in 750 l of staining medium before FACS analysis: alkaline phosphatase (AP) Live Stain (Life Technologies A14353) with staining medium (3.3% FCS, 0.1% NaN3) and within a few days the sexually reproduced tadpoles hatched, swam to the setonlics and metamorphoses into an oozoid, with a sessile body plan (Fig. 1a). The oozoid begins a cyclical budding process of asexual reproduction, forming a colony of genetically identical zooids and buds (Fig. 1b). The colonial individuals are united under a single gelatinous tunic by a network of blood vessels, which terminate in sausage-shaped protrusions (ampullae; Fig. 1a–c). Throughout adult life, B. schlosseri regenerates its entire body every two weeks. This stem-cell-mediated cycle of development includes the formation of all body organs including the heart, respiratory system, digestive system and neural complex. Sexual reproduction starts with fertilization and progresses through classic embryonic stages into a tadpole larva with chordate characters. Embryonic development commences when the ovary and testis are formed.

Juvenile colonies were labelled with distinct fluorescent stains (CellTrace CFSE Green, Invitrogen #C34554; CellTrace FarRed, Invitrogen #C34572) and positioned in controlled fusion/rejection reactions (Fig. 3d). For allorecognition assays, naive colonies were tied to 3 mm mesh using a sterile 1-ml syringe pump (a similar procedure to murine transplantation reactions). The fusion requires at least one shared allele: upon fusion, the resultant screened markers (Extended Data Fig. 1). This self–nonself recognition process is controlled by BHF. Fusion requires at least one shared allele: upon fusion, the anti-BHF serum shows that BHF is expressed in live cells (propidium iodide negative) by the anti-BHF serum. We used a published protocol38 to extract RNA from B. schlosseri colonies (n = 24). Twenty-three endpoint populations were sequenced, in addition to CP2, which was composed of all live cells. CP17 was composed of dead cells and cell debris, and therefore was excluded from the population analysis. Twenty thousand cells of each sequenced population were sorted in 750 l of staining medium before FACS analysis: alkaline phosphatase (AP) Live Stain (Life Technologies A14353) with staining medium (3.3% FCS, 0.1% NaN3) and within a few days the sexually reproduced tadpoles hatched, swam to the setonlics and metamorphoses into an oozoid, with a sessile body plan (Fig. 1a). The oozoid begins a cyclical budding process of asexual reproduction, forming a colony of genetically identical zooids and buds (Fig. 1b). The colonial individuals are united under a single gelatinous tunic by a network of blood vessels, which terminate in sausage-shaped protrusions (ampullae; Fig. 1a–c). Throughout adult life, B. schlosseri regenerates its entire body every two weeks. This stem-cell-mediated cycle of development includes the formation of all body organs including the heart, respiratory system, digestive system and neural complex. Sexual reproduction starts with fertilization and progresses through classic embryonic stages into a tadpole larva with chordate characters. Embryonic development commences when the ovary and testis are formed.
included the endostyle epithelial cells, cells circulating in the subendostylar sinuses and the main and lateral subendostylar nervous, which extend along the endostyle anterior–posterior axis. Dissected endostyle samples were flash-frozen in liquid nitrogen to minimize RNA degradation and stored at −80 °C. Using a mechanized Kone tissue grinder and pestle, samples were homogenized in the presence of lysis buffer (Qiagen RNAeasy Microkit #74004), and total RNA was extracted following the manufacturer’s protocol. The resultant RNA was cleaned and concentrated (Zymo Research RNA Clean and Concentrator-5, R1015) and analysed using an Agilent 2100 Bioanalyzer for quality analysis before library preparation. cdNA libraries were then prepared from high-quality samples (RNA integrity number (RIN) > 8) using Ovation RNA-seq v2 (Nugen). Size selection was performed before barcoding using 2ymo Research Select-a-Size DNA Clean and Concentrator Kit (D4080). Libraries were barcoded using NEBnext Ultra DNA Library Prep Kit Master for Illumina (New England Biolabs, E7370S) and NEBNext Multiplex Oligos for Illumina (New England Biolabs, E6690S). Barcoded library samples were then sequenced on an Illumina NextSeq 500 (2 x 150 bp, producing an average of 15 million reads per cell population).

**Cytopotoxicity and phagocytosis assays.** Three different ex vivo phagocytosis assays were used to identify the B. schlosseri myeloid lineage phagocytic cell populations: (i) phagocytosis of fluorescent beads; (ii) phagocytosis of a fluorescently labelled marine bacterium (Vibrio diazotrophicus); and (iii) allogeneic phagocytosis, during which we tested the capability of cells from different colonies to engulf allogeneic cells (Extended Data Fig. 7). FACS was used to identify the phagocytic cells in each of the ex vivo assays and to track their cell populations. Confocal microscopy and ImageStream analysis were used to confirm engulfment (Extended Data Fig. 7).

**Bacteria and beads phagocytosis assays.** Cells (10⁶–2 x 10⁶ [u]) of each of the four labelled populations were incubated overnight at 18 °C in a 2:1 ratio of bead:cells using Flouresbrite YG Carboxylate Microspheres 1.00 μm (Polysciences). To measure phagocytosis of bacteria, the marine bacterium V. diazotrophicus was heat-inactivated at 95 °C for 5 min, and labelled with Alexa Fluor 647 (Invitrogen #A20066). The analysis of cells positive for beads or bacteria was done by flow cytometry using two fluorescent channels—green channel was used to detect beads and the far-red channel was used to detect bacteria (Extended Data Fig. 7). FlowJo V10 (FlowJo) was used to analyse the flow cytometry data.

**Phagocytosis of allogeneic cells.** The labelling and incubation of cells were done as described for the cytotoxicity assay below. Phagocytosis was defined as double-positive cells in the FACS plots of the two labelling markers. Owing to the natural fluorescence of B. schlosseri cells, the level of double-positive cells in wells with separation of each one of the labels was reduced from the double-positive in the experimental wells (about 5% background). To validate phagocytosis, cells were sorted and observed by confocal microscopy. To compare cytotoxicity to phagocytosis, the optimal ratio (1:1) for allogeneic phagocytosis was prepared and compared to cytotoxicity at the same ratio, using 10⁶ cells of each labelled group per well.

We used a FACS-based cytotoxicity assay⁴¹ to measure killing of cells in vitro. Isolated B. schlosseri cells were labelled using CFSE (5 μM; Life Technologies) and Far Red dye (1 μM; Life Technologies) for 30 min at 18 °C to distinguish effectors and target cells and washed twice in staining medium: 3.3 X PBS, 2% FCS and 10 mM Heps. Cells were incubated overnight in 96-well U-shaped plates at 18 °C at different effector:target ratios in staining medium. After adding propidium iodide to wells to test cell viability, cells were analysed by FACS. The target or the effector cells were gated on 2D analysis using the labelling dyes (CFSE and Far Red). Spontaneous lysis was measured as the percentage of propidium iodide-positive cells in gated target cells in wells without effector cells, and sample lysis was quantified as the percentage of propidium iodide-positive cells in gated target cells. Specific lysis was calculated as follows: specific lysis % = 100 × ((sample lysis – spontaneous lysis)/(100 – spontaneous lysis)). The gating of FACS analysis was on 2D plots owing to the natural fluorescence of B. schlosseri cells. For BHF blocking assays, anti-BHF polyclonal mouse serum was used at 1:200, or mock serum as control. BHFs from known fusion or rejection outcomes from our experiments were used to validate autofluorescent cells. For the subendostylar sinus (the endo-niche) we detected no transplanted cells for CP3 (0/4) or the uninjected colonies (0/4), whereas 5/6 colonies were positive for cHSC cells, showing significant localization of cHSCs and control cells (CP3) to migrate to these sites. An additional channel was used to validate autofluorescent cells. For the subendostylar sinus (the endo-niche) we detected no transplanted cells for CP3 (0/4) or the uninjected colonies (0/4), whereas 5/6 colonies were positive for cHSC cells, showing significant localization of cHSCs and control cells (CP3) to migrate to these sites. An additional channel was used to validate autofluorescent cells.

**Electron microscopy.** For haematoxylin and eosin (H&E) staining, cells were incubated overnight on glass slides coated with poly-1-lysine (Sigma) and fixed with 4% PFA in 0.1 M MOPS in 0.5 M NaCl, pH 7.5 for 10 min and washed with PBS. The slides were stained with Harris haematoxylin for 5 min (Extended Data Fig. 1e). Immunohistochemistry was based on a published method⁴⁵. The labelling was done using rabbit anti phospho-histone H3 (pH3; Cell Signaling, 9701) at a concentration of 1:500 (Extended Data Fig. 5d).

**Electron microscopy.** Colonies were fixed overnight in 1.5% glutaraldehyde buffered with 0.2 M sodium cacodylate, 1.6% NaCl, pH 7.4. After washing in buffer and post-fixation in 1% OsO4 in 0.2 M cacodylate buffer plus 1.6% NaCl, specimens were dehydrated and embedded in Araldite. Sections were counterstained with toluidine blue, and for electron microscopy sections were stained for contrast with uranyl acetate and lead citrate. Micrographs were taken with a FEI Tecnai G2 electron microscope (operating at 100 kV) (Extended Data Fig. 5c).

**Differential expression analysis.** Determination of gene counts was performed using a Snakemake⁴⁶ pipeline. An outline of the steps is as follows: (i) low-quality bases and adaptor sequences were removed using Trimmomatic⁴⁷ (version 0.32); (ii) overlapping paired end reads were merged using FLASH⁴⁸ (version 1.2.11); (iii) reads were aligned to the UniVec Core database using Bowtie2⁴⁹ (version 2.2.4) to remove biological vector and control sequences; (iv) reads were aligned to the B. schlosseri transcriptome with BWA⁵⁰ (mem algorithm, version 0.7.12); and (v) aligned reads were sorted and indexed using SAMtools, PCR duplicates were removed using Picard Tools (MarkDuplicates, version 1.128) and then transcript level counts directly curated from the BAM file.

**Differential expression analysis.** Determination of gene counts was performed using a Snakemake⁴⁶ pipeline. An outline of the steps is as follows: (i) low-quality bases and adaptor sequences were removed using Trimmomatic⁴⁷ (version 0.32); (ii) overlapping paired end reads were merged using FLASH⁴⁸ (version 1.2.11); (iii) reads were aligned to the UniVec Core database using Bowtie2⁴⁹ (version 2.2.4) to remove biological vector and control sequences; (iv) reads were aligned to the B. schlosseri transcriptome with BWA⁵⁰ (mem algorithm, version 0.7.12); and (v) aligned reads were sorted and indexed using SAMtools, PCR duplicates were removed using Picard Tools (MarkDuplicates, version 1.128) and then transcript level counts directly curated from the BAM file.
least five counts per million in at least 80% of the smaller number of the compared samples. A simple model was used to compare the two sets of populations, with \( P \) values adjusted using the Benjamini–Hochberg method to produce a false discovery rate (FDR). FDRs below 0.05 were called as being differentially expressed. The comparisons between cell populations were performed in a one-versus-all approach, followed by selective aggregating of similar populations. Initially each population was compared to all others (except for CP2, which was an aggregate of all cells). Then all sets of two populations were compared to the remaining populations, with the best two being grouped together. The metric used was: if \( A \) is the set of genes found to be differentially expressed in population \( A \) versus all others, and \( n(A) \) is the number of genes in that set, then find the maximum of \( n(AB) - n(A \cup B) \). This attempts to find those populations that when grouped together are more distinct from all others compared to the populations individually. After this grouping was performed multiple times, the maximum was no longer above zero, and a new metric was used. This method was used to find the two groups of populations that maximized \( n(AB) - n(A \cup B) \). This approach was used to compare the different cell populations in a heat map (Extended Data Fig. 3a). Many of the populations have visually similar gene expression patterns (for example: CP3, CP5 and CP23; and CP25, CP33 and CP34), and in general, samples that were adjacent in FACS space are also similar in expression space (Extended Data Fig. 3b, 67% of neighbours in agreement), indicating a correlation between gene expression profile and morphology and marker expression.

**Gene domain finding.** The protein sequences of CP31-associated non-homologous genes (\( n = 52 \)) were saved to individual FASTA files. These were processed using the python-based RESTful client for InterPro5 (version 5.28–67.0) on ENSEMBLEs website using the default options to search for annotated domains within ascidian-specific genes that dominate the Botryllus cytotytic morula cells (Supplementary Table 5). Unannotated gene domains revealed: transmembrane domains (\( n = 7 \)), cell interaction and sugar binding domains (\( n = 4 \)) and signal peptides (\( n = 3 \)). This analysis also revealed genes with toxin and peptidases domains (g6777, g34113), hydrolase (g61144) and complement and coagulation domains (g6777, g23113). Moreover, g15971 contained a C-type lectin domain, a presence of immobile, bi-functional cells, PloS ONE 7, e35772 (2012).

**Data availability**

Sequencing data can be found on the NCBI Sequence Read Archive under accession PRJNA41486. RPKM values of gene expression and differential expression analysis results are shown in Supplementary Table 1. All other relevant data are available in the manuscript.

35. Boyd, H. C. et al. Growth and sexual maturation of laboratory-cultured Monterey Botryllus schlosseri. Biol. Bull. 170, 91–109 (1986).
36. Toolea, D. A. & De Narniasso, A. Botryllus schlosseri: tdna.109 (2015).
37. Belland, S. C. et al. Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. Science 332, 687–696 (2011).
38. Koh, P. W. et al. An atlas of transcriptional, chromatin accessibility, and surface marker changes in human mesoderm development. Sci. Data 3, 160109 (2016).
39. Burighel, P. & Brunetti, R. The circulatory system in the blastozooid of the colonial ascidian Botryllus schlosseri (Pallas). Ital. J. Zool. 38, 273–289 (1971).
40. Zaniolo, G., Lane, N. J., Burighel, P. & Manni, L. Development of the motor nervous system in ascidians. J. Comp. Neurol. 443, 124–135 (2002).
41. Edin-Brami, M. et al. Glycans in sera of amyotrophic lateral sclerosis patients and their role in neuronal cells. PloS ONE 7, e35772 (2012).
42. Sabbadin, A. & Graziani, G. Microgeographical and ecological distribution of colour morphs of Botryllus schlosseri (Asciidae). Nature 213, 815–816 (1967).
43. Cima, F. et al. Life history and ecological genetics of the colonial ascidian Botryllus schlosseri. Zool. Anz. 257, 54–70 (2015).
44. Laird, D. J. & Weissman, I. L. Continuous development precludes radioprotection in a colonial ascidian. Dev. Comp. Immunol. 28, 201–209 (2004).
45. Braden, B. P. et al. Vascular regeneration in a basal chordate is due to the presence of immobile, bi-functional cells. PloS ONE 9, e95460 (2014).
46. Köster, J. & Rahmann, S. Snakemake—a scalable bioinformatics workflow engine. Bioinformatics 28, 2520–2522 (2012).
47. Bolger, A. M., Lahme, M. & Usadel, B. Trimomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120 (2014).
48. Magot, T. & Salzberg, S. L. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27, 2957–2963 (2011).
49. Longmiread, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359 (2012).
50. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26, 589–595 (2010).
51. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140 (2010).
52. Wolff, L. & Hurnenich, R. Concise review: erythroid versus myeloid lineage commitment: regulating the master regulators. Stem Cells 31, 1237–1244 (2013).
53. Rosental, B. et al. The effect of chemotherapy/radiotherapy on cancerous pattern recognition by NK cells. Curr. Med. Chem. 19, 1780–1791 (2012).
54. Potts, K. S. & Bowman, T. V. Modeling myeloid malignancies using zebrafish. Front. Oncol. 7, 297 (2017).
55. Moss, L. et al. Identification of phagocytic cells, NK-like cytotoxic cell activity and the production of cellular exudates in the coelomic cavity of adult zebrafish. Dev. Comp. Immunol. 33, 1077–1087 (2009).
56. Wei, S. et al. The zebrafish activating immune receptor Nitr9 signals via Dap12. Immunogenetics 59, 813–821 (2007).
57. Sun, B. et al. Dissecting hematopoietic and renal cell heterogeneity in adult zebrafish at single-cell resolution using RNA sequencing. J. Exp. Med. 214, 2875–2887 (2017).
58. Han, Q. et al. Characterization of lamyepi-L17 family members and their receptors. J. Immunol. 195, 5440–5451 (2015).
59. Hirano, M., Das, S., Guo, P. & Cooper, M. D. The evolution of adaptive immunity in vertebrates. Adv. Immunol. 109, 125–157 (2011).
60. Han, Y., Huang, G., Zhang, Q., Yuan, S. & Liu, J. The primitive immune system of amphioxus provides insights into the ancestral structure of the vertebrate immune system. Dev. Comp. Immunol. 34, 791–796 (2010).
61. Rhodes, C. P., Ratcliffe, N. A. & Rowley, A. F. Presence of coelemocytes in the primitive chordate amphioxus (Branchiostoma lanceolatum). Science 217, 263–265 (1982).
62. Muñoz-Chápuli, R., Carmona, R., Guadix, J. A., Macías, D. & Pérez-Pomares, J. M. The origin of the endothelial cells: an evo-devo approach for the invertebrate/vertebrate transition of the circulatory system. Evol. Dev. 7, 351–358 (2005).
63. Lanot, R., Zachary, D., Holder, F. & Meister, M. Postembryonic hematopoiesis in Drosophila. Dev. Biol. 230, 243–257 (2001).
64. Meister, M. & Lagueux, M. Drosophila blood cells. Cell. Microbiol. 5, 573–580 (2003).
65. Nakayama, K., Nomoto, A. M. & Nishijima, M. Morphological and functional characterization of hemocytes in the giant clam Tridacna crocea. J. Invertebr. Pathol. 69, 105–111 (1997).
Extended Data Fig. 1 | B. schlosseri cell sorting workflow. a, Outline of cell purification process. Unsorted cells (light microscopy) are loaded into a FACS (i) and sorted, and this is followed by morphological observation (ii). Cells were labelled with diverse markers and screened using CyTOF (iii) for differential labelling. On the basis of SPADE cluster analysis (iv) markers were selected for FACS gating (v) before a final sort was performed (vi; c). b, Sorting based on FCS/SSC in the lower panel, and natural fluorescence in the upper panels. The analysis is after gating propidium iodide (PI)-negative cells (live cells). The specific excitation laser and optical filter for emission measurements were as follows (excitation laser in nm and filters stated as long pass (LP) and band pass (BP)): 488 nm for FSC and SSC, 488 nm (550LP 575/25BP) - PE, 405 nm (450/50BP) - Pacific Blue, 633 nm (660/20BP) - APC, 633 nm (755LP 780/60BP) - APC-Cy5.5. Nomenclature based on published work3,26. Experiment was performed three times. Scale bars, 20 μm. c, Sorting panels of 34 cell populations using FSC, SSC (central panel), CD49d, CD57, concanavalin-A (ConA), BHF and AP after gating PI-negative cells (live cells). Central panel is FCS/SSC, from which additional populations were differentiated. The specific excitation laser and optical filter for emission measurements were as follows: 488 nm (505LP 530/30BP) - AP, 488 nm (755LP 780/60BP) - CD49d, 405 nm (450/50BP) - CD49d, 405 nm (450/50BP) - CD57, 633 nm (660/20BP) - ConA, 633 nm (755LP 780/60BP) - BHF. d, Hierarchy of sorted cell populations by main parameters of differentiation. For example, the control populations used in Fig. 2b–e (CP19, 20, 21) are all included in CP18. CP18 is derived from CP9. e, H&E staining of the end point cell populations isolated in a rectangle of original population by FSC/SSC. Live imaging was done three times; H&E was one experiment with three replicates. Colour key (c, d) for different populations identified in this study and the figures that further describe them.
Extended Data Fig. 2 | Screen for differentiation markers of cell populations for FACS-based sorting. a, b, Examples of antibodies screened by CyTOF mass cytometry with B. schlosseri cells analysed by 2D mass spectrometry. The CyTOF screen was performed once. a, Examples of antibodies considered as nonspecific binders because different antibodies showed the same binding patterns. b, Examples of antibodies considered as specific binders because a cell population was bound by one antibody but not the other (red rectangle). c, d, Examples of validation screen by flow cytometry of antibodies with specific binding by CyTOF, done in 2D fluorescence excited by the same laser owing to autofluorescence. Negative FACS validation was done twice and positive three times. c, Negative validation of CCR2 labelled with PE. d, Positive validation of CD57 labelled with Pacific Blue. e, Flow cytometry of live B. schlosseri cells labelled with anti-BHF mouse serum. Anti-mouse Alexa Fluor-647 was used as a secondary antibody. f, Example of positive differential labelling by the lectin concanavalin-A in PE-Cy7 by FACS. The experiment was performed three times. g, Confocal imagery of membrane concanavalin-A labelling with Alexa Fluor-633 in red and Hoechst DNA labelling in blue. The experiment was performed once. Scale bar, 20 μm.
Extended Data Fig. 3 | *B. schlosseri* cell population clustering based on transcriptome analysis. **a**, We used 250 genes with the largest weights in the first 11 principal components (explaining 90% of the variance in mean-adjusted log-transformed gene counts) to cluster the different cell populations in a heat map. **b**, Transcriptome sequencing of *B. schlosseri* cell populations compared to FACS analysis. 2D projection of the distances between transcriptomes of cell populations based on all differential genes. Lines are drawn between the nearest two neighbours. Blue, FACS adjacency of the populations in the differentiation panel; red, genetic level proximity not predicted by FACS panel adjacency. The proximities of twenty (of thirty) genetic level cellular populations were predicted by FACS. Widths of lines are inversely proportional to distances.
Extended Data Fig. 4 | Gene expression in *B. schlosseri* cell populations.

**a.** Enrichment scores of the top nine systems (left), nine tissues (centre) and ten cell types (right) of annotated genes upregulated in CP33, CP34 and CP25 using the GeneAnalytics tool (compared to human). In systems the haematopoietic system has the highest score, in tissues the blood has the highest score, and within the cells, the granulocytes and HSC have the highest scores.

**b–d.** Geneset Activity Analyses using the Gene Expression Commons tool on a mouse haematopoiesis model of different *B. schlosseri* cell populations. **b.** Analysis of CP8 (pigment cells) based on 12 significantly upregulated genes, showing that CP8 is part of the haematopoietic system with gene activity resembling known cell types. **c.** Analysis of CP3 (small cells) based on 235 significantly upregulated genes, showing that this population is not part of the haematopoietic system. On the bottom: enrichment scores of the top ten tissues using CP3 genes by GeneAnalytics tool compared to human; the highest score is for the testes, suggesting that this population is a gonadal population. **d.** Analysis of CP19 (enriched with morula cells) based on 96 upregulated genes (*P* < 0.25), showing that CP19 has gene activity resembling cells in the lymphoid lineage using Geneset analysis.
Extended Data Fig. 5 | Subendostylar sinuses as an HSC niche.

a, Reduction in DiO fluorescence suggesting cell proliferation, three weeks after transplantation. The experiment was performed once with two pools for each population from five animals each. b, Candidate HSC (cHSC) population and a control population (CP3) were isolated, labelled with DiD and transplanted into CFSE-labelled compatible colonies; in vivo tracing of transplanted cell migration was used to identify niches. There were no cells detected for CP3 (0/4) or the uninjected colonies (0/4) in the subendostylar sinus, whereas 5/6 colonies injected with cHSC cells showed significant localization of the cHSCs to the subendostylar sinus. $P = 0.048$, Fisher’s exact test, two-tailed. Although in the cell islands 4/4 were positive with CP3, 5/6 positive with cHSC, and 2/4 positive in the uninjected colonies, there are high levels of autofluorescent cells in the cell islands. Full image panels of Fig. 2g. c, Transverse sections of an adult zooid counterstained with toluidine blue (top two left) where the endostyle (green arrowhead) and endo-niche (blue arrowhead) are enlarged (scale bar, 30 μm). Electron microscopy section of the same animal’s endostyle and endo-niche (right and bottom, enlarged). Yellow arrowheads indicate cells with haemoblast (HSC) morphology that are enriched within the endo-niche (scale bar, 5 μm). The experiment was performed three times. Full image panels of Fig. 2b. d, Immunohistochemistry with antibodies against phospho-histone H3 (PHH3), suggesting that there are mitotic cells in the endostyle region in the developing primary bud and also in the adult zooid endostyle.
Extended Data Fig. 6 | Gene expression in an HSC niche: the endostyle.

a, Comparison between the transcriptome sequence data from 10 samples of dissected endostyles and the transcriptome data for 34 whole colonies revealed a list of 327 genes that were significantly upregulated in the endostyle and showed homology to genes expressed in human haematopoietic bone marrow. Heat map includes the top 100 (by log(FC)) of the bone-marrow-associated endostyle genes. b, Geneset Activity Analysis of top 200 genes upregulated in the *B. schlosseri* endostyle associated with the blood system, found by RNA-seq (this study) using the Gene Expression Commons tool on a mouse haematopoiesis model. The enriched populations are bone marrow stromal cells and HSCs. c, Similar analysis, but for *C. robusta* based on previous in situ work, revealed enriched mouse bone marrow stromal cells as well, based on 188 genes that are expressed in the *C. robusta* endostyle and are associated with the blood system.
Extended Data Fig. 7 | Discovery of a myeloid lineage phagocytic population. 
a, FACS analysis of *B. schlosseri* cells that are fluorescently positive in one of three phagocytosis assays performed: (first and second from left) phagocytosis of green fluorescent beads, (third) phagocytosis of *V. diazotrophicus* labelled with AF647, and (fourth) allogeneic phagocytosis. Three phagocytic populations were identified: amoebocytes (A), myeloid cells (M), and large phagocytes (LP). The experiment was repeated twice. The myeloid cells were the main contributors to phagocytosis, contributing more than 40% to each of the phagocytosis assays. The large phagocytes contribute mainly to allogeneic phagocytosis compared to the other assays. 
b, Live images of the three isolated phagocytic populations. The experiment was performed three times. Scale bar, 20 μm. 
c, We isolated the three main phagocytic populations and a small cell population (CP3) as a control, and incubated each one with fluorescent beads to validate the engulfment capacity of each population. The experiment was repeated twice. Plots show FACS analysis of green fluorescent bead phagocytosis by sorted populations. 
d, Amoebocytes, myeloid cells, and large phagocytes all had significantly higher engulfment rates than the small cell population. Moreover, amoebocytes and myeloid cells had significantly higher cell percentages than the large phagocyte population. Percentage analysis was carried out on two samples for each sorted population. Unpaired *t*-test, two-tailed; *P* < 0.05, data shown as mean. 
e, Representative confocal images of the three phagocytic populations after engulfment of beads. Scale bar, 20 μm. 
f, ImageStream analysis confirmed that the three phagocytic populations assayed engulfed the beads. The positive cells had mainly the morphology of amoebocytes, myeloid cells, or large phagocytes. The experiment was performed once on ImageStream. Representative images of the three phagocytic populations after engulfment of beads. Scale bar, 7 μm.
Extended Data Fig. 8 | Cytotoxicity and the two morphs of morula cells at PORs. **a**, Cytotoxicity assays of isolated LGL cells compared to small cells, and to isolated morula cells (MCs). In both cases the LGL cells had significantly higher cytotoxicity compared to the other cell types. The experiment with isolated cells was performed twice with triplicates. Unpaired t-test, two-tailed; *P = 0.003, **P = 0.0013; data shown as mean. **b**, LGL cells were isolated (upper left) and incubated overnight either in syngeneic (upper right) or in allogeneic challenge (lower left). FSC/SSC analysis of LGL cell (lower population) and morula cells (upper population). Insets, sample light microscopy images of the cells after incubation for each treatment. Lower right, analysis of LGL and morula cells in syngeneic compared to allogeneic challenge. The experiment was performed once with duplicates and validated by light microscopy. Bars show mean. **c**, H&E-stained section of *B. schlosseri* colonies undergoing rejection. In the ampule (AMP) the inactivated form of cytotoxic morula cells/large granular lymphocyte-like cells (LGL) can be observed (top). On the other hand, the activated form with the brown pigmentation of morula cells can be observed at POR (bottom). **d**, Confocal imagery of phagocytosis assays to validate the allogeneic engulfment. Colonies are labelled with CFSE in green and with DiD in red after allogeneic phagocytosis assay. Large phagocytic cells can be seen after engulfment of allogeneic cells or vesicles. Validation of allogeneic phagocytosis by confocal imaging was performed twice. Scale bar, 20 μm. **e**, Example of cytotoxicity assay with different effector to target (E:T) ratios, where the targets are compatible or rejecting colony cells to the effector colony. In the rejecting colony, specific lysis is significantly higher. The experiment was performed three times with triplicates. ANOVA two-factor with replication; *P = 0.0015; data shown as mean ± s.d.
Extended Data Table 1 | Antibodies screened by CyTOF for binding of *B. schlosseri* cells

| Symbol | Mass | Panel1 antigen | Clone | Supplier | CyTOF | FACS | Panel2 antigen | Clone | Supplier | CyTOF | FACS |
|--------|------|---------------|-------|----------|-------|------|---------------|-------|----------|-------|------|
| Pd/Cd  | 110-114 | CD3 | S4.1 | Invitrogen | Yes | No |
| In     | 113 | CD7 | M-T701 | BD | Yes | No |
| In     | 115 | CD45 | H30 | Biolegend | Yes | No |
| La     | 139 | CD57 | HCD57 | Biolegend | Yes (low) | Yes (~14%) | CD2 | RPA-2.10 | Biolegend | Yes (low) | No |
| Pr     | 141 | NKP46 | 195314 | R&D system | No | CD61 | VI-PL2 | Biolegend | No |
| Nd     | 142 | CD48 | TU145 | BD | No |
| Nd     | 144 | CC55 | HEK/1/B5a | Biolegend | Yes | N/A | CD94 | DX22 | Biolegend | Yes (low) | N/A |
| Nd     | 145 | ULRB1 | 292319 | R&D | No |
| Nd     | 146 | CD309 | 89106 | BD | No |
| Sm     | 147 | CD8 | RPA-T8 | Biolegend | No |
| Nd     | 148 | KIR3DL1 | DX9 | BD | No | CRTAM | Cr24.1 | Biolegend | No |
| Sm     | 149 | DNA1 | DX11 | BD | No |
| Eu     | 151 | NKG2D | 1011 | Biolegend | No | CD84 | CD84.1.21 | Biolegend | No |
| Sm     | 152 | TNFR2 | 22235 | R&D | No |
| Eu     | 153 | NK2G2C | 134522 | R&D | Yes (low) | N/A |
| Sm     | 154 | NKP44 | P44-8 | Biolegend | No | Notch1 | MHN1-519 | Biolegend | No |
| Ed     | 155 | CRACC | 162.1 | Biolegend | No |
| Gd     | 156 | KIR2DL3 | 180701 | R&D | No |
| Gd     | 158 | CD161 | HP-3G10 | Biolegend | No | dec12A | SOC1 | Biolegend | No |
| Tb     | 159 | CD11c | Bu15 | Biolegend | No |
| Gd     | 160 | NKP3O | 210845 | R&D | No | 2B4 | C1.7 | Beckman | No |
| Dy     | 161 | CD15 | W6D3 | Biolegend | No | CD4 | RPA-T4 | Biolegend | Yes | No |
| Dy     | 162 | CD49d | 9F10 | Biolegend | Yes | Yes (~28%) |
| Dy     | 163 | CD16 | 3G8 | Biolegend | No | KIR2DS1 | EB6.B | Beckman | No |
| Er     | 166 | TIGIT | MB5A43 | eBioscience | No | KIR3DL2 | DX31 | Gift from Dr. Lanier | No |
| Er     | 167 | CD27 | O323 | Biolegend | No | CCR7 | G043H7 | Biolegend | No |
| Er     | 168 | KIR2DL1 | 143211 | R&D | No | CCR2 | 48607 | BD | Yes | No |
| Er     | 170 | CD11a | HI111 | Biolegend | No | CD11b_act | CRBM1/5 | Biolegend | No |
| Yb     | 172 | KIR2DL4 | 181703 | R&D | No | CD34 | 8G12 | BD | Yes (low) | N/A |
| Yb     | 173 | CD33 | WM53 | Biolegend | No |
| Yb     | 174 | CD144 | TEA/31 | Beckman | No | NKG2A | Z199 | Beckman | No |
| Lu     | 175 | KLRG1 | 13F12F2 | ThermoFisher | No |
| Yb     | 176 | CD56 | NCAM16.2 | BD | Yes (low) | N/A |
| Er     | 191/193 | DNA | All cells | DNA | Sigma | All cells | Sigma | All cells | No |
| Pt     | 195 | cisplatin | Sigma | No | Sigma | No | Sigma | No |

A screen of 49 antibodies that would cross-react with *B. schlosseri* cells and could potentially differentiate cellular populations. Symbol represents element; mass shows elemental isotope mass; antigen is the human antigen against which the antibody was produced. CyTOF column represents whether the *B. schlosseri* cell population was positive or not; low means less than 1% of cells were positive. FACS column represents whether there was binding by the same antibody clone by flow cytometry.
Extended Data Table 2  | Flow cytometry binding of *B. schlosseri* cells, and references for phagocytic and cytotoxic cells in different organisms

| a | Lectin | Fluorophore | % positive cells |
|---|---|---|---|
| ConA | AF-633 | 45% |
| PNA | PE-Cy7 | 16% |
| UEA | PE-Cy7 | 30% |
| Additional markers | | | |
| Alkaline Phosphatase | Green | 9% high 25% mid |
| Anti-BHF polyclonal | AF-647 | 35% |

| b | Animal: | Phagocytic cells | Cytotoxic cells |
|---|---|---|---|
| Homo (human) | Myeloid lineage review\textsuperscript{52} | Review\textsuperscript{53} |
| Danio (zebrafish) | Myeloid lineage review\textsuperscript{54}, phagocytosis\textsuperscript{55} | Recognition molecules\textsuperscript{56}, cytotoxicity\textsuperscript{55}, cell types\textsuperscript{57} |
| Petromyzon (lamprey) | Myeloid lineage\textsuperscript{58} | Lymphoid lineage review\textsuperscript{59} (cytotoxicity was not shown) |
| Branchiostoma (lancelet) | Large phagocytes\textsuperscript{60}, some suggestion of amoebocytes\textsuperscript{61} | |
| Strongylocentrotus (sea urchin) | Phagocytic cells review\textsuperscript{32} | Suggestion that colorless spherule cells are cytotoxic morula cells in review\textsuperscript{32} |
| Drosophila (fruit fly) | Amoebocytes \textsuperscript{62}, plasmic cells or pupal macrophage-like have resemblance to the large phagocytes\textsuperscript{63,64} | Crystal cells that contain the enzymes for melanization for cytotoxicity could resemble the morula cells at the enzymatic level\textsuperscript{63,64} |
| Limulus (horseshoe crab) | Amoebocytes and granular phagocytes in review\textsuperscript{33} | Cells with prophenoloxidase and melanization process in review\textsuperscript{33,34} |
| Tridacna (clam) | Amoebocytes and eosinophilic granular hemocytes that resemble large phagocytes\textsuperscript{65} | Morula cells\textsuperscript{65} |

\textsuperscript{a} The column ‘% positive cells’ shows the percentage of the cells that were positively labelled by the marker. For alkaline phosphatase (AP), ‘high’ represents cells that labelled strongly and ‘mid’ represents cells that were positively but not strongly labelled. \textsuperscript{b} Rows contain references\textsuperscript{52-65} and any notes for each of the organisms reviewed in the production of Fig. 4.
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

| n/a | Confirmed |
|-----|-----------|
| ✔   | The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| ✔   | An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ✔   | The statistical test(s) used AND whether they are one- or two-sided |
| ✔   | Only common tests should be described solely by name; describe more complex techniques in the Methods section. |
| ✔   | A description of all covariates tested |
| ✔   | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ✔   | A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ✔   | For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted |
| ✔   | Give P values as exact values whenever suitable. |
| ✔   | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ✔   | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ✔   | Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated |
| ✔   | Clearly defined error bars |
| ✔   | State explicitly what error bars represent (e.g. SD, SE, CI) |

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection

Explained in the methods section. The pipeline was written in Snakemake (version 3). Reads trimmed with Trimmmomatic (version 0.32), FLASH used to merge reads (version 1.2.11), Bowtie2 (version 2.2.4) and BWA ("mem" algorithm, version 0.7.12) for alignments. SAMtools used for sorting and indexing reads, Picard ("MarkDuplicates" tool, version 1.128) to remove duplicates.

Data analysis

Explained in the methods section. Differential expression performed using edgeR in R (version 3.x). Other scripts written in python (version 3.6) and jupyter notebooks (version 5.x). Python libraries include matplotlib (version 2.x), pandas (version 0.22), scikit-learn (version 0.18) and HTSeq (version 0.8). Also used were the python-based RESTful client for InterPro5 (version 5.28-67.0) on ENSEMBLE’s website using the default options.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.
Data

Policy information about availability of data
All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Data availability: Sequencing data can be found on the NCBI Sequence Read Archive under accession PRJNA414486. RPKM values of gene expression and differential expression analysis results are in Supplementary Table 1. All other relevant data are available in the manuscript.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☑ Life sciences  ☐ Behavioural & social sciences  ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size
- No sample size calculation was performed, some of the experiments were limited by number of animals or the amount of cells that could be collected or sorted.

Data exclusions
- Cell population 17 was excluded since it was mainly debris. Explained in the text.

Replication
- Experiment replication explained in the methods and figure legends. For the main part of cell isolation and population by FACs in order to validate the reproducibility, we have used several machines including BD FACs Aria II- as the main sorter, BD FACs Accuri C6 for analysis, BD LSRFortessa for analysis, and validated the populations (FSC/SSC) by Sony SH800S cell sorter.

Randomization
- All the experiments were done together for the treatments and controls, there was no need for randomization since there was no difference between the samples in the experiments. All the animals that were compared to each other (mainly in transplantation assays) were from the same batches or the same offspring pool from the same mother colony, also for the treatments and the controls.

Blinding
- The analysis of the pigment cell based differentiation experiment was done in a single blind manner 20 days following transplantation. No additional blinding was performed since most of the data collected was numerical analysis.

Reporting for specific materials, systems and methods

Materials & experimental systems

| n/a | Involved in the study |
|-----|-----------------------|
| ☑ Unique biological materials |
| ☑ Antibodies |
| ☑ Eukaryotic cell lines |
| ☑ Palaeontology |
| ☑ Animals and other organisms |
| ☑ Human research participants |

Methods

| n/a | Involved in the study |
|-----|-----------------------|
| ☑ ChiP-seq |
| ☑ Flow cytometry |
| ☑ MRI-based neuroimaging |

Antibodies

Antibodies used
All antibodies described in the methods including the anti-BHF mouse serums. The antibodies used are CD49d PE-Cy7 (BioLegend 304313: Clone 9F10) 1μg, CD57 Pacific Blue (BioLegend 322316; clone HCD57) 0.25μg.

Validation
All antibodies described in the methods. Part of the research was the screen for antibodies that would make cluster of differentiation of Botryllus cell populations. The initial screen was done by Cytof of 49 antibodies. The antibodies that were positive by Cytof were validated by FACS. The positive antibodies were checked three times. Finally we have used the following antibodies: CD49d PE-Cy7 (BioLegend 304313: Clone 9F10), CD57 Pacific Blue (BioLegend 322316; clone HCD57).
### Animals and other organisms

Policy information about [studies involving animals; ARRIVE guidelines](#) recommended for reporting animal research

| Category               | Description                                                                 |
|------------------------|-----------------------------------------------------------------------------|
| Laboratory animals     | Explained in the methods. All work was done with invertebrate model Botryllus schlosseri. |
| Wild animals           | Explained in the methods. We have used animals from the mariculture facility. |
| Field-collected samples| There was no field collection animals. We have used animals from the mariculture facility. |

### Flow Cytometry

#### Plots

- Confirm that:
  - The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
  - The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
  - All plots are contour plots with outliers or pseudocolor plots.
  - A numerical value for number of cells or percentage (with statistics) is provided.

#### Methodology

| Component              | Explanation                                      |
|------------------------|--------------------------------------------------|
| Sample preparation     | Explained in the methods.                        |
| Instrument             | Explained in the methods.                        |
| Software               | Explained in the methods.                        |
| Cell population abundance | Explained in the relevant sections.             |
| Gating strategy        | Explained in the methods.                        |

- Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.