Molecular dynamics of the matrisome across sea anemone life history

B. Gideon Bergheim[1], Alison G. Cole[2], Mandy Rettel[3], Frank Stein[3], Stefan Redl[4], Michael W. Hess[4], Aissam Ikmi[3] and Suat Özbek[1*]

[1]University of Heidelberg, Centre for Organismal Studies, Department of Evolutionary Neurobiology, Im Neuenheimer Feld 230, 69120 Heidelberg, Germany

[2]Department of Neurosciences and Developmental Biology, Faculty of Life Sciences, University of Vienna, Djerassiplatz 1, 1030 Vienna, Austria

[3]European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany

[4]Innsbruck Medical University, Muellerstrasse 59, 6020 Innsbruck, Austria

*Correspondence: suat.oezbek@cos.uni-heidelberg.de

Abstract

The evolutionary expansion of extracellular matrix (ECM) molecules has been crucial for the establishment of cell adhesion and the transition from unicellular to multicellular life. Members of the pre-bilaterian phylum cnidaria offer an exceptionally rich perspective into the metazoan core adhesome and its original function in developmental and morphogenetic processes. Here, we present the ensemble of ECM proteins and associated factors for the starlet sea anemone *Nematostella vectensis* based on *in silico* prediction and quantitative proteomic analysis of decellularized mesoglea from different life stages. The integration of the matrisome with single cell transcriptome atlases reveals that *Nematostella's* complex ECM is predominantly produced by gastrodermal cells, confirming the homology of the cnidarian inner cell layer with bilaterian mesoderm. The transition from larva to polyp is characterized by an upregulation of metalloproteases and basement membrane components including all members of an unusually diversified SVEP1/Polydom family, suggesting massive epithelial remodeling. The enrichment of Wnt/PCP pathway factors during this process further indicates directed cell rearrangements as a key contributor to the polyp's morphogenesis. Mesoglea maturation in adult polyps involves wound response proteins suggesting similar molecular patterns in growth and regeneration. Our study identifies conserved matrisomal networks that coordinate transitions in *Nematostella's* life history.

Introduction

The evolution of extracellular matrix, a complex network of secreted, typically modular proteins, is closely linked to the emergence of metazoan life forms (Rokas 2008; Hynes 2009; Ozbek, et al. 2010; Hynes 2012; Naba 2024). While some ECM components, such as integrins and cadherin receptors can be traced back to unicellular organisms (Sebe-Pedros, et al. 2010; Nichols, et al. 2012), early diverging cnidarians (hydroids, jellyfish, corals, and sea anemones) are believed to possess one of the most complete adhesomes among pre-bilaterian clades (Ozbek, et al. 2010; Tucker and Adams 2014). Cnidarians, the sister group to bilaterians, are characterized by a simple gastrula-shaped body with a mouth opening surrounded by tentacles. They are diploblastic organisms, consisting of an outer ectoderm and an inner endoderm separated by a complex ECM called mesoglea (Sarras 2012; Bergheim and Ozbek 2019). The mesoglea, which is best studied in the freshwater polyp *Hydra*, forms a flexible, tri-laminar structure composed of a central, amorphous interstitial matrix (IM) interspersed with mainly collagenous filaments, sandwiched between two thin layers of basement membrane (BM) (Sarras, et al. 1991; Shimizu, et al. 2008; Aufschnaiter, et al. 2011; Bergheim and Ozbek 2019). Previous studies in *Hydra* have shown that the mesoglea can be separated intact from the epithelial cell sheets by a freeze-thaw technique (Day and Lenhoff 1981; Veschgini, et al. 2023). Previously, we analyzed the proteome of decellularized *Hydra* mesoglea and identified 37 unique protein sequences (Lommel, et al. 2018), including most of the described core matrisome components (Sarras 2012). Among medusozoans (jellyfish and hydroids), hydras stand out for having lost the free-swimming medusa form. They also lack a planula larva stage from which anthozoans (corals and sea anemones) typically produce sessile polyps. We therefore speculated that a cnidarian species with a complex life cycle could offer a more comprehensive picture of the non-bilaterian ECM repository. Here, we analyzed the matrisome of the anthozoan starlet sea anemone *Nematostella vectensis*, by employing *in silico* predictions of ECM proteins that were partially confirmed by a subsequent proteomic analysis of decellularized mesoglea. We detected a rich collection of matrisome proteins, comparable

in its core matrisome complexity to vertebrate species (Naba, et al. 2012). Furthermore, mapping of our matrisome data onto a previously established single-cell transcriptome dataset (Steger, et al. 2022; Cole, et al. 2024) revealed a prominent role of the gastroderm in ECM production. Cnidocytes, which produce the cnidarian stinging organelle, are characterized by a distinct set of ECM proteins that significantly contribute to the complexity of the cnidarian matrisome. Quantitative proteomics of mesoglea samples from different life stages (larva, primary and adult polyp) showed that while the larval mesoglea contained only few exclusively enriched factors, the transition from the larval stage to primary polyp was marked by an upregulation of a large fraction of the matrisome. This set of proteins included many metalloproteases and basement membrane factors, indicating significant epithelial reorganization. Remarkably, all members of an unusually diverse SVEP1/Polydom family were upregulated during this morphogenetic process, implicating a conserved role of this protein family for epithelial morphogenesis. Additionally, a significant enrichment of Wnt/planar cell polarity (PCP) signaling components, such as ROR2 and protocadherin Fat4, supports that directed cell movements underlie the axial elongation and morphogenesis of the polyp (Stokkermans, et al. 2022). The final transition to the adult animal involves an increased addition of elastic fiber components to the mesoglea and matricellular factors associated with wound healing, indicating common ECM-associated mechanisms in regeneration, growth and tissue differentiation.

Results

Molecular composition and structure of the *Nematostella* ECM

To investigate the components and dynamics of the ECM throughout various life stages of *Nematostella vectensis*, we employed a protocol for obtaining decellularized mesoglea, originally developed for *Hydra* (Day and Lenhoff 1981; Lommel, et al. 2018; Veschgini, et al. 2023). We isolated mesogleas from larvae at three days post-fertilization (3 dpf), primary polyps at 10 dpf, and from small adult polyps that were at least 1 year old. Protein extraction

was performed under strongly reducing conditions at high temperatures (90°C) to solubilize the cross-linked protein network of the ECM. The extracted proteins were then digested with trypsin and analyzed using quantitative mass spectrometry (Fig. 1A). Extensive studies in both vertebrates and invertebrates have identified specific characteristics of ECM proteins, typically based on conserved domains and domain arrangements (Engel 1996; Hohenester and Engel 2002; King, et al. 2008; Ozbek, et al. 2010; Naba, et al. 2012; Tucker and Adams 2014). Building on this knowledge, we conducted a *de novo* annotation of all predicted *Nematostella* protein models using InterProScan (Jones, et al. 2014), identifying 1,812 potential candidates for the *in silico* matrisome (Fig. 1B). These candidates were further classified into orthogroups based on similarity using OrthoFinder (Emms and Kelly 2019). To refine this analysis, we also predicted *in silico* matrisomes from protein models of several early-branching metazoan species, including two choanoflagellates, two sponges, three ctenophores, 9 additional representative cnidarians, and two placozoan species. SignalP (Teufel, et al. 2022) was used to confirm the presence of signal peptides, while DeepLoc (Thumuluri, et al. 2022) was employed to predict the cellular localization of the matrisome draft. Additionally, we explored the closest BLAST hits in the NCBI and SwissProt databases. After manually excluding duplicates and sequencing artifacts, we narrowed down the *in silico* matrisome to 829 candidate genes predicted to encode ECM proteins. Of these, 213 sequences (25%) were specific to cnidarians, and 20 were restricted to *Nematostella* (Fig. 1B, supplementary table S1). Each protein sequence was manually reviewed, with annotations assigned based on domain predictions, UniProt identifiers, and comparisons to known ECM proteins. Our subsequent mass spectrometry analysis initially identified a total of 5,286 proteins. However, we suspected that a significant portion might be cellular contaminants originating from residual amoebocytes within the mesoglea (Tucker, et al. 2011), which could not be completely removed during the isolation procedure. To address this, we used our curated *in silico* matrisome as a filter and identified 287 ECM components in the isolated *Nematostella* mesogleas, including 56 cnidarian-specific proteins (Fig. 1B, supplementary tables S1, S2). This number is significantly higher than previously reported for the *Hydra* matrisome (Lommel,

et al. 2018), suggesting a greater compositional diversity in *Nematostella*. A comparison of mesoglea samples of both species resolved by protein gel electrophoresis confirmed the higher complexity of the *Nematostella* mesoglea, particularly in the lower molecular weight fraction (supplementary figure S1). We further organized our curated dataset following the classification proposed by Hynes and Naba, who introduced the concept of a "core" matrisome, characterized by 55 signature InterPro domains, including EGF, LamG, TSP, vWFA, and collagen (Hynes and Naba 2012). Within this core matrisome, we identified 241 proteins, including collagens, proteoglycans, and ECM glycoproteins (e.g., laminins and thrombospondins) (Fig. 1C, supplementary table S1). Additionally, we identified a set of 310 "matrisome-associated" factors. This group includes molecules that (i) have structural or functional associations with the core matrisome, (ii) are involved in ECM remodeling (e.g., metalloproteases), or (iii) are secreted proteins, including growth factors. In summary, the predicted *Nematostella* matrisome comprises 551 proteins, constituting approximately 3% of its proteome (Artamonova and Mushegian 2013) and roughly half the size of the human matrisome with 1,056 proteins (Naba, et al. 2012). While the *Nematostella* core matrisome is comparable to that of bilaterians, vertebrate species exhibit a dramatic expansion in ECM-associated factors (Fig. 1D). The remaining 278 proteins, identified through negative selection based on exclusive domain lists for each ECM category (Naba, et al. 2012), primarily consist of transmembrane adhesion receptors (e.g., cadherins and IgCAM-like molecules), enzymes, and glycoproteins with specialized functions, such as venoms.

To validate the preservation of the isolated ECM, we generated polyclonal antibodies targeting unique peptide sequences in laminin gamma 1 (anti-Lam), type IV collagen NvCol4b (anti-Col4), a specific fibrillar collagen NvCol2c (anti-Col2), and a consensus motif for several *Nematostella* fibrillar collagens (anti-PanCol) (supplementary figure S2). We then performed immunofluorescence confocal microscopy on *Nematostella* whole mounts and decellularized mesoglea from all life stages (Fig. 1E-K). Consistent with recent findings in *Hydra* (Veschgini, et al. 2023), the *in vivo* and *ex vivo* images exhibited similar patterns along the polyp's body,

indicating the structural integrity of the mesoglea after decellularization (Fig. 1F-K). In cross-sections, the laminin antibody stained the thin double layer of the basement membrane (BM) lining the two epithelia, while the PanCol antibody detected the intervening fibrous layer of the interstitial matrix (IM) (Fig. 1E). Both antibodies also showed diffuse staining at the apical surface of ectodermal cells, likely due to the sticky nature of the glycocalyx. Ultrastructural examination of the mesoglea revealed a mean thickness of approximately 0.5 µm in larvae and 1.5 µm in primary polyps (supplementary figure S3A-B, G-H, supplementary table S3). In the triangular areas at the base of the gastrodermal mesenteric folds, the mesoglea was expanded (supplementary figure S3C-F). The BM lining the epithelia of larvae was very delicate, measuring only 70 nm, and the IM appeared as a loose array of thin fibrils (supplementary figure S3B). Primary polyps possessed a distinct BM, with a thickness of approximately 130nm, appearing as a dense meshwork of fibrils (supplementary figure S3H, supplementary table S3). Previous reports (Tucker, et al. 2011) indicated that the IM in older primary polyps was interspersed with thin fibrils of about 5 nm and extended thick fibrils of about 20-25 nm. Our samples from younger primary polyps (supplementary figure S3H) showed thin fibrils of approximately 6 nm and thick ones of about 13 nm, with occasional fibrils of up to 27 nm (supplementary table S3). Immunostainings with the laminin antibody revealed a thickened mesoglea at the aboral pole of the polyp, forming an unusual knot-like structure (supplementary figure S3J). Ultrastructural analysis showed that the fibrils in this region were densely packed and aligned along the oral-aboral axis (supplementary figure S3K-L), possibly providing a rigid attachment site for the mesentery retractor muscles. Immunoelectron microscopy confirmed the observations from immunofluorescence: anti-Lam immunogold labeling was primarily localized along the plasma membrane of ECM-lining cells, anti-Col4 labeling was found at the BM (supplementary figure S3M-N), and anti-PanCol labeling was predominantly distributed throughout the IM, with a similar pattern observed for the anti-Col2 antibody (supplementary figure S3O-P).

Cell type specificity of matrisome expression

Recently, single-cell RNA sequencing has been applied to identify the origin of neuroglandular cell lineages in *Nematostella* (Steger, et al. 2022), hypothesize on the origin of muscle cell types (Cole, et al. 2023), and catalog the distribution of cell states associated within all tissue types (Cole, et al. 2024). We made use of this developmental cell type atlas to determine the cell type specificity of matrisomal gene expression. Expression profiles for all genes of the matrisome across the entire life cycle are available in supplemental data (supplementary table S4). We calculated an average expression score for each of the matrisome gene sets (core, associated, and other) and found above average scores for the core matrisome associated with the mature gastrodermis and developing cnidocytes, and to a lesser extent also of the other two categories (Fig. 2). Core matrisome genes also showed additional high expression scores within an uncharacterized gland cell type (GD.1), matrisome-associated genes within the digestive gland set, and other genes within the maturing cnidocytes (Fig. 2A). We looked specifically at the distribution of core matrisome genes across all cell states and generated a list of differentially up-regulated genes (supplementary table S5, supplementary figure S4). Of note is the absence of any differentially expressed core matrisome factors within the ectodermal tissues, and contrastingly a large set that are specific to the mesoendodermal inner cell layer or cnidocytes (Fig. 2B). We also find sets of core matrisome genes that are specific to different secretory gland cell types, including mucin-producing, digestive-enzyme producing, and uncharacterized S2-class cell types. Interestingly, the gland cell-specific matrisome genes include several of the Polydom members upregulated during larva-to-primary polyp transition (see below). Altogether these expression profiles suggest that core components of the mesoglea (collagens, laminins) are produced from the inner cell layer, and that a large set of ECM glycoproteins and all of the minicollagens (David, et al. 2008; Zenkert, et al. 2011) have been recruited into the formation of the cnidarian synapomorphy, the cnidocytes.

Collagens constitute the primary structural components of the animal ECM and, being a highly diverse family of triple helical proteins (Fidler, et al. 2018), form a significant part of the core matrisome. Vertebrate collagens consist of 28 types (I-XXVIII) categorized as fibril-forming, network or beaded filament-forming, and transmembrane collagens (Kadler, et al. 2007). Type IV collagen, a major constituent of basement membranes, is considered to be a primordial component of the animal ECM based on studies in ctenophores that lack fibrillar collagens but display a remarkable diversity of collagen IV genes (Fidler, et al. 2017; Draper, et al. 2019). Our analysis revealed 12 bilaterian-type collagens (supplementary figure S5A) with conserved C-terminal non-collagenous trimerization domains (NC1) and extended triple helical stretches of ~1000 residues, including two sequences for type IV collagen (NvCol4a/b) together with a peroxidasin homolog (NV2.13306), indicating the presence of sulfilimine cross-links that stabilize the *Nematostella* BM (Fidler, et al. 2014). In *Hydra*, which lacks this specific post-translational modification, six collagen genes have been identified through cDNA cloning and proteomic analysis (Deutzmann, et al. 2000; Fowler, et al. 2000; Zhang, et al. 2007; Lommel, et al. 2018). We have classified the collagens according to the *Hydra* nomenclature (Zhang, et al. 2007) and identified three isoforms of NvCol1, each comprising an isolated minor triple helical domain at the N-terminus (supplementary figure S5A). In addition, we detected four NvCol2 paralogs, which contain an additional whey acidic protein 4 disulfide core (WAP) domain at the N-terminus. However, our dataset lacks a Hcol3-like collagen with alternating N-terminal WAP and vWFA domains, although it includes a protein with extensive WAP/vWFA repeats but lacking a collagen triple helix or NC1 domain (NV2.11346). NvCol5 consists of a continuous collagen domain with two minor interruptions following the signal peptide and is otherwise similar to NvCol1. We have not found sequences resembling Hcol6, which is characterized by triple helical sequences interrupted by multiple vWFA domains. Collagen XVIII-like, which is predicted to contain an unrelated Mucin-like insertion, includes an N-terminal Laminin-G/TSPN motif followed by a discontinuous central collagenous domain, similar to the BM-associated vertebrate α1(XVIII) collagen chains (Heljasvaara, et al. 2017) and *Drosophila* Multiplexin (Meyer and Moussian 2009). NvCol7 shares a similar domain

organization but differs by having only a single interruption of the triple helix near the N-terminus. According to Exposito et al., the *Nematostella* fibrillar collagen sequences are phylogenetically related to A clade collagens that in mammalians possess a vWFC module in their N-propeptide supposed to have evolved from the cnidarian WAP domain (Exposito, et al. 2008). NvCol7 is an exception and belongs to B clade collagens characterized by the possession of an N-terminal TSPN domain.

Most of the *Nematostella* collagens show a broad expression in diverse gastrodermal cell populations throughout all life stages (supplementary figure S6). NvCol2c is an exception, as it is predominantly expressed in neuronal cells of larvae indicating a specialized function in neurogenesis. NvCol7 is distinguished by showing an expression both in gastrodermal cells and two small cnidocyte cell populations described by Steger et al. to be exclusive for planula larvae (Steger, et al. 2022). Unlike the IM collagens, NvCol4a is additionally expressed during embryogenesis, emphasizing the pivotal role of the BM in organizing the epithelial tissue architecture during early development. In addition to these bilaterian-type collagens, we have identified four spongin-like proteins, which are short-chain collagens with derived NC1 domains (supplementary figure S5B). These molecules have been described as truncated variants of collagen IV with a wide distribution in several invertebrate phyla (Aouacheria, et al. 2006). Except for NvSpongin-like-2, their expression is not aligned with that of fibrillar collagens, but is mostly restricted to neuroglandular cells (supplementary figure S6). An additional set of “collagen-like” sequences that comprise short triple helical stretches without additional domain motifs (supplementary figure S5B) is broadly expressed across cell types and developmental stages (supplementary figure S6, supplementary table S4). In contrast, the large set of minicollagens (supplementary figure S5C) is strictly aligned with the cnidocyte lineage as detailed below.

A specialized cnidocyte matrisome

Cnidocytes, the stinging cells of cnidarians, are a key synapomorphy of this clade. Previous studies have demonstrated that, from a transcriptomic perspective, nematocyst capsule formation is distinct from the mature profile (Chari, et al. 2021; Steger, et al. 2022). We filtered the matrisome for genes expressed within cnidocytes and found 298 genes that are detectable above average in at least 5% of any cnidocyte transcriptomic state (supplementary figure S7, supplementary table S6). We further separated this list of genes into those that are absent from non-cnidocyte states (101 genes: ‘exclusive’) or are also detected within more than 50% of the non-cnidocyte transcriptomic states (41 genes: ‘ubiquitous’) (Fig. 3A). We consider the remaining genes shared across cnidocyte and non-cnidocyte profiles (156: ‘shared’). Of the cnidocyte-specific genes, we examined the proportion of genes specific to either the capsule-building specification profiles (79: ‘specification’) or the maturation phase of cnidogenesis (24: ‘maturation’). Most of these genes are restricted to the specification phase, with a smaller subset associated with the mature transcriptomic profile (Fig. 3A). The latter group includes several members of a vastly expanded family of Fibrinogen-related proteins (FREPs) (supplementary table S1), which have been implicated in innate immunity across various phyla (Zhou, et al. 2024) and may function as venom components. In the 'shared' gene set, most genes are associated with the mature cnidocyte profile and overlap with various neuroglandular subtypes (supplementary figure S7). This observation supports the hypothesis that the cnidocytes arose from an ancestral neuronal (Richards and Rentzsch 2014; Tournière, et al. 2020). We then calculated a gene module score for each gene set to estimate specificity across the dataset, summarizing the specificity of each gene set (Fig. 3B). Further examination of the gene lists reveals cnidocyte specificity of nematogalectins and minicollagens that serve as structural components for cnidocysts (Kurz, et al. 1991; Hwang, et al. 2010). Minicollagens, that have served as important phylum-specific genes (Holland, et al. 2011), comprise a short collagen domain (about 15 Gly-X-Y repeats) flanked by proline-rich regions and terminal cysteine-rich domains (CRDs) (David, et al. 2008). We identified all five previously described

Nematostella minicollagen sequences (Zenkert, et al. 2011; Steger, et al. 2022) (NvCol-1, -3, -4, -5, -6) along with four additional proteins with incomplete minicollagen sequence features, which we termed “minicollagen-like” (supplementary figure S5C). Intriguingly, we also identified a protein that combines features of both minicollagens and extended ECM-type collagens, suggesting a possible evolutionary origin of minicollagens from this gene family. This protein, NvNCol-7, contains a minicollagen pro-peptide sequence, proline-rich regions, and canonical N- and C-terminal CRDs (Tursch, 2016 #1245). Unlike previously described minicollagens, it includes an extended discontinuous collagen sequence of ~1000 residues, comprising 25 alternating blocks of mostly 12 or 15 Gly-X-Y repeats. These blocks are interrupted by either a single alanine or MPP/SPPSPP sequences, resembling degenerated collagen triplets. The presence of a minicollagen pro-peptide in this collagen suggests its expression in the cnidocyte lineage and secretion into the nematocyst vesicle as a structural component of cnidocyst walls or tubules (Adamczyk, et al. 2010; Garg, et al. 2023). This is confirmed by the single cell expression data, which show prominent and exclusive expression in cnidocyte lineages (supplementary figures S6, S7). Interestingly, *Nematostella* minicollagens exhibit differential expression across different cnidocyte subtypes, nematocytes and spirocytes (Fig. 3C). Spirocytes express NvCol-5 and NvNcol-like-3 and 4, while NvCol-1, 4 and 6 and NvNCol-7 are expressed within the nematocytes and NvCol-3 is expressed in both. Other cnidocyte-specific genes also show differential paralog expression between cnidocyte types, including the four NOWAs (Engel, et al. 2002; Garg, et al. 2023), and the vWFA domain proteins (Fig. 3C). NvTSR-2 vs NvTSR-3 distinguish between the two nematocyte lineages, nem.1 and nem.2. These are not confidently identified although postulated to be basitrichous haplonemas/isorhizas based on abundance and distribution across the adult tissue libraries. In summary, a significant fraction (23 %) of matrisomal genes is expressed within the cnidocyte lineage. The proportion of cnidarian-specific factors (30 %) in this subset is enriched as compared to the full matrisome (supplementary table S1), supporting the notion that the cnidocyte matrisome represents a specialized ECM adapted to

the unique assembly process and biophysical requirements of the cnidarian stinging organelle (Ozbek 2011).

Larva-to-polyp transition is marked by factors of basement membrane remodeling and Wnt/PCP signaling

Larva-polyp morphogenesis in *Nematostella* involves significant changes in body shape including the elongation of the body axis and the development of oral tentacles and internal mesenteries (Stokkermans, et al. 2022). We performed quantitative proteomics using tandem mass tag labeling (TMT) (see supplementary figure S8 for normalization steps) to examine whether this process is accompanied by stage-specific variations of matrisome components. As shown above, 35 % (287/829) of the predicted factors within the curated *in silico* matrisome were experimentally confirmed (Fig. 1B). We identified stage-specific matrisome components by assigning hits (>2-fold change and <0.05 false discovery rate (fdr)) using a modified t-test limma for each of the three life stage comparisons (supplementary table S7). As illustrated in the heat map in Fig. 4A, the transition from larvae to primary polyp is characterized by a general increase of ECM components. 94 proteins were differentially upregulated in primary polyps while only four, including vitellogenin and its receptor that are crucial for lipid transport from the ECM into the oocyte (Lebouvier, et al. 2022), were differentially abundant in larvae. The ECM components enriched in primary polyps as compared to larvae, besides various factors involved in general cell adhesion (e.g., cadherin-1, coadhesin-like proteins), contain two major functional groups (Fig. 4B, supplementary table S7): (1) factors involved in BM establishment and remodeling and (2) components of the Wnt/PCP signaling pathway. Both groups are indicative of a massive epithelial rearrangement, rather than cell proliferation, as a driver of larva-to-polyp transition (Stokkermans, et al. 2022). Notably, besides laminins, the BM proteoglycan perlecan and numerous (n = 10) astacin, ADAMTS and MMP family metalloproteases (see supplementary figure S11 for an overview of matrisomal proteases), the first group contained all members of an unusually expanded Polydom protein family

(supplementary figure S9). Polydom/SVEP1 is a secreted multidomain ECM protein initially discovered in a murine bone-marrow stromal cell line (Gilges, et al. 2000). In humans, it is composed of eight different domains including an N-terminal vWFA domain, followed by a cysteine-rich ephrin receptor motif, Sushi/CCP and hyalin repeat (HYR) units, EGF-like domains, a central pentraxin domain and a long tail of 28 Sushi domains terminating in three EGF repeats (supplementary figure S9). Polydom has recently been shown to be a ligand for the orphan receptor Tie1 and induce lymphatic vessel remodeling via PI3K/Akt signaling (Sato-Nishiuchi, et al. 2023). Earlier studies have shown basement membrane deposition of Polydom and a role in epithelial cell-cell adhesion via integrin binding (Sato-Nishiuchi, et al. 2012; Samuelov, et al. 2017). The *Hydractinia* homolog has been characterized as a factor specific to i-cells and to be potentially involved in innate immunity (Schwarz, et al. 2008). In comparison to vertebrate Polydoms, it contains additional Pan/Apple, FA58C, and CUB domains, but has a reduced number of terminal Sushi repeats (supplementary figure S9). In our study, we identified a *Nematostella* Polydom homolog (NvPolydom1) that shares high similarity with the *Hydractinia* and *Hydra* proteins, suggesting a conserved arrangement of domains in cnidarians (supplementary figure S9). This includes a pentraxin-PAN/Apple-FA58C core structure, a tail consisting of six Sushi repeats, and 1-2 terminal CUB domains. Two additional paralogs, NvPolydom2 and NvPolydom3, exhibit differences in their central region, with NvPolydom2 lacking the PAN/Apple domain and NvPolydom3 lacking both PAN/Apple and FA58C domains. Furthermore, we discovered four shorter Polydom-like sequences that share a common EGF-Sushi-HYR-TKE structure but lack vWFA and Pentraxin domains, as well as the terminal Sushi repeats. These Polydom-like proteins, resembling a truncated N-terminal part of canonical Polydoms, possess additional domains at the N- or C-termini, such as TSP1 repeats or Ig-like domains. The shortest member, referred to as Polydom-related, lacks EGF-like modules and consists of the central Sushi, HYR, and TKE domains found in all Polydom-like sequences, suggesting that this core motif may be essential for biological function. All Polydom and Polydom-like paralogs are expressed within the putative digestive cell state GD.1, with two of these showing additional expression within the

uncharacterized secretory cell type S2.tll.2&3 (supplementary figure S10). This, together with the exceptionally high isoform diversity, indicates a requirement for genetic robustness to account for perturbations of the developmental process regulated by the *Nematostella* Polydoms. Given that mouse Polydom is essential for endothelial cell migration in a Tie-dependent manner (Sato-Nishiuchi, et al. 2023), it is plausible that the *Nematostella* homologs could serve a similar function for rearrangement of epithelial cells along the BM during primary polyp morphogenesis. Interestingly, the top differentially abundant factor in the primary polyp mesoglea is a secreted integrin-alpha-related protein (sIntREP) containing three integrin-alpha N-terminal domains followed by a stretch of EGF repeats. It is an attractive hypothesis that sIntREP modulates Integrin-dependent cell adhesion by Polydoms and other factors to facilitate cell migration. The second group of proteins that contains several components of the Wnt/PCP pathway, including ROR2, protocadherin Fat4-like, and hedgling, further supports the assumption of epithelial cell migration as a main driver of primary polyp morphogenesis. Wnt/PCP signaling, originally described in *Drosophila melanogaster* (Gubb and Garcia-Bellido 1982; Adler 2002), has a well-established role in convergent extension movements of cells during gastrulation to facilitate the elongation of the embryo along its oral-aboral axis (Gao 2012). Its “core” factors include the transmembrane proteins Frizzled, Van Gogh/Strabismus, Flamingo, and the intracellular components Dishevelled, Prickle, and Diego (Simons and Mlodzik 2008). An additional level of PCP within tissues is regulated by the unusually large protocadherins Fat and Dachsous that interact in a heterophilic manner and display cellular asymmetries (Matakatsu and Blair 2004). Recently, the *Hydra* Fat-like homolog has been reported to be polarized along the oral-aboral axis and to organize epithelial cell alignment via organization of the actin cytoskeleton (Brooun, et al. 2020). Hedgling is an ancestral, non-bilaterian member of the cadherin superfamily with high similarity to FAT and Flamingo cadherins (Adamska, et al. 2007). It has a gastrodermal expression in the *Nematostella* primary polyp (Adamska, et al. 2007), which it shares with Wnt5a and ROR2 (supplementary table S4). ROR2, a highly conserved receptor tyrosine kinase, is the principal transducer of

PCP signaling via Wnt5a in the *Xenopus* embryo (Hikasa, et al. 2002; Schambony and Wedlich 2007). It has also been reported to induce directional cell movements in mammals (He, et al. 2008) and to induce filopodia formation as a prerequisite for directed cell migration (Nishita, et al. 2006). Taken together, the matrisome dynamics revealed by quantitative mass spectrometry suggests a massive epithelial rearrangement and directed cell migration as the main morphogenetic process underlying axial elongation during primary polyp morphogenesis. Interestingly, IM components such as fibrillar collagens do not appear to contribute largely to this process. This changes in the adult mesoglea, which, compared to the primary polyp, is significantly enriched in elastic fiber components, such as fibrillins and fibulin, likely adding to the visco-elastic properties (Gosline 1971a, b) of the growing body column (Fig. 4A,C, supplementary table S7). In addition, the adult mesoglea contains several matricellular factors associated with different aspects of wound healing in vertebrate organisms (Cárdenas-León, et al. 2022). These include SPARC-related follistatin domain proteins, uromodulin, and periostin. The *Nematostella uromodulin* gene was previously reported to become highly upregulated in the wound ectoderm, probably as part of the innate immune response (DuBuc, et al. 2014). The same study showed a circular upregulation of the MMP inhibitor *NvTIMP* around the wound site. Indeed, while metalloproteases are similarly upregulated in adults and primary polyps, we observed a noticeable increase of diverse classes of protease inhibitors in adults (7 vs 2 in primary polyps), including TIMP, Kunitz and Kazal-type protease inhibitors, as well as thyroglobulin repeat proteins (Novinec, et al. 2006) (supplementary table S7). This is indicative of a high degree of protease activity regulation in tissue morphogenesis during growth as also observed during *Nematostella* whole-body regeneration (Schaffer, et al. 2016). Taken together, the transition from the primary polyp to the adult is characterized by an increase of elastic fibrillar IM components contributing to the long-range elasticity and resilience of the mesoglea, and a recruitment of wound response factors that together with a complex network of proteases and protease inhibitors likely regulate growth, organogenesis and tissue differentiation.

Discussion

The evolution of the ECM, a complex proteinaceous network that connects cells and organizes their spatial arrangement in tissues, has been a key innovation driving the emergence of multicellular life forms (Rokas 2008; Brunet and King 2017). Although ctenophores have recently been identified as the sister group to all other animals (Schultz, et al. 2023), the hitherto available genomic evidence suggests that they possess only a minimal repertoire of conserved ECM and ECM-affiliated proteins, limiting comparative studies with bilaterians (Draper, et al. 2019). In contrast, cnidarian genome data have offered broad evidence for conserved adhesomes anticipating the complexity of mammalian species (Tucker and Adams 2014). Here, we identified 829 ECM proteins that comprise the matrisome of the sea anemone *Nematostella vectensis* through *in silico* prediction using transcriptome databases and analyzed the dynamics of 287 ECM factors by TMT labeling and LC-MS/MS of decellularized mesoglea samples from different life stages. Utilizing cell-type specific atlases we showed that the inner gastroderm is the major source of the *Nematostella* ECM, including all 12 collagen-encoding genes. This finding supports the model of germ layer evolution proposed by Steinmetz et al. (Steinmetz, et al. 2017) where, based on both transcription factor profiles and structural gene sets, the cnidarian inner cell layer (endoderm) is homologous to the bilaterian mesoderm that gives rise to connective tissues. The observation that in *Hydra* both germ layers contribute to the synthesis of core matrisome proteins (Epp, et al. 1986; Zhang, et al. 2007) might be related to a secondary loss of the anthozoan-specific mesenteries, which represent extensions of the mesoglea into the body cavity sandwiched by two endodermal layers. According to fossil records, anthozoa are considered the basal branch in cnidarians with the ancestral form exhibiting an octoradial symmetry with mesenteries (Ou, et al. 2022). The primacy of the gastrodermis in ECM synthesis might therefore represent an ancestral feature, which in medusozoans was modified to adapt to the immensely enlarged mesoglea of the medusa life-stage. To evaluate the complexity of *Nematostella's* matrisome across cnidarians and other metazoan phyla, we plotted matrisome sizes from published databases

and newly generated *in silico* matrisomes of representative species against orthogroup counts (Fig. 5). For the *in silico* matrisomes, only orthogroups shared with at least one found in published matrisomes were counted. Surprisingly, anthozoan species generally exhibit a higher complexity than medusozoans and populate a transitory region between bilaterians and non-bilaterians in the evolutionary trajectory. This indicates that the acquisition of complex life cycles in medusozoa, that are distinguished by the pelagic medusa stage, led to a secondary reduction in the matrisome repertoire. This might have been a strategy to minimize the cost for ECM remodeling during metamorphosis and rely on a restricted set of highly expressed genes to form the expanded jellyfish mesoglea.

A significant fraction of *Nematostella's* exceptionally rich matrisome is devoted to the formation of the cnidocyst, a unique cellular novelty of the cnidarian clade (Jekely, et al. 2015; Babonis, et al. 2023). Cnidocyst-specific proteins follow an unusual secretion route into the lumen of the growing cnidocyst vesicle, which topologically represents extracellular space (Ozbek 2011). It has therefore been speculated that they originated from neurosecretory vesicles used for predation in early metazoans (Balasubramanian, et al. 2012). The recent finding that cnidocysts are instrumental for the predatory lifestyle of the *Aiptasia* larvae (Maegele, et al. 2023) supports the hypothesis of a deeply-rooted extrusive mechanism for prey capture in metazoan evolution. In this context, it is intriguing that the majority of *Nematostella* cnidocyte genes shared with other cell types is expressed within neuroglandular subtypes (supplementary figure S7). In addition, the cnidocyst-specific matrisome contains diverse proteins with repetitive ECM domains that likely have general bioadhesive or fibrous properties that might play a role in entangling and ingesting prey organisms. These include several fibropellin-like and other EGF repeat proteins (poly-EGF) as well as thrombospondin type-1 repeat (TSR) proteins, such as properdin-likes.

The changes in the matrisome profiles across *Nematostella's* major life stages suggest a highly dynamic epithelial rearrangement during primary polyp morphogenesis. The involvement of Wnt/PCP factors in this process indicates similar cell migration and re-

orientation events as during convergent extension processes in gastrulation. Kumburegama et al. have shown that primary archenteron invagination and apical constriction of bottle cells in *Nematostella* is dependent on the PCP components *strabismus* (Kumburegama, et al. 2011) and *Fzd10* (Wijesena, et al. 2022). Mesentery formation in primary polyps, which involves sequential folding events of the endodermal epithelium (Berking 2007), likely involves similar molecular pathways. As already observed by Appelöf (Appelöf 1900), mesoglea synthesis follows invagination during this process. The molecular network composed of Wnt/PCP and basal membrane factors that our data revealed might therefore indicate an actomyosin-controlled invagination of the endodermal layer followed by BM production to re-align the cells in their apico-basal architecture. The upregulation of wound response factors in the adult animal might indicate a transient loss of tissue integrity during collective cell migration which could entail an actin-based purse-string mechanism as during wound (Begnaud, et al. 2016; Bischoff, et al. 2021). Future work will further decipher the gene-regulatory network controlling polyp morphogenesis in this anthozoan model.

Materials and Methods

***Nematostella* culture.** Adult *Nematostellas* were kept in plastic boxes at 18 °C in 1/3 artificial sea water (~11 ppm) (*Nematostella* medium) in the dark. They were fed with freshly hatched *Artemia* nauplii and cleaned once per week. To induce spawning, animals were transferred to 27 °C *Nematostella* medium in light for 8 h and then washed with 18 °C *Nematostella* medium. The egg patches were collected and dejellied in 5 % cysteine solution for 15 min. Unless otherwise stated, the embryos were left to develop in *Nematostella* medium at room temperature (RT) in normal day/night cycles.

Immunocytochemistry. Larvae on 3 dpf, primary polyps on 10 dpf, and small adult polyps were collected and left to relax at 27 °C in direct light for 30 min. A solution of 7 % $MgCl_2$ in seawater was slowly added, and the animals were anesthetized for 20 min. Fixation was

performed with Lavdovsky's fixative (50 % ethanol, 36 % H_2O, 10 % formaldehyde, 4 % acetic acid) for 30 min at RT after which the samples were incubated in 150 mM Tris, pH 9.0, 0.05 % Tween-20, for 10 min. An incubation step at 70 °C for 10 min and a subsequent cooling to RT was followed by three 10 min washing steps in PBS, 0.1 % Tween-20, PBS, 0.1% Tween-20, 0.1% Triton-X100, and PBS, 0.1% Tween-20, respectively. Primary antibody incubation (rabbit anti-Laminin, rat anti-Pan-Collagen) was performed at 1:100 in 0.5 % milk powder overnight at 4 °C. The samples were washed 3 times for 10 min in PBS, 0.1% Tween-20 and incubated with secondary antibodies at 1:400 for 2.5 h at RT. Prior to mounting on object slides with Mowiol, DAPI was added at 1:1000 for 30 min. Decellularized mesogleas were transferred onto a microscopy slide lined with liquid blocker. The mesogleas of larvae and primary polyps were carefully stuck to the slide using an eyelash. The staining protocol followed the whole mount immunocytochemistry protocol with 10 min fixation and only one washing step per wash to avoid washing off of the mesogleas. The antibodies were incubated on the slides in a Petri dish with a wet paper towel to prevent evaporation. Images were acquired with an A1R microscope at the Nikon Imaging Facility Heidelberg Further image processing was performed with Fiji ImageJ v1.53t.

Mesoglea decellularization. About 500.000 larvae (3 dpf) and primary polyps (10 dpf), and four adult animals were collected. The adult polyps were cut open along the oral-aboral axis using a scalpel to ease the decellularization of the endoderm. All samples were incubated in 0.5 % N-lauryl-sarcosinate for 5 min and then frozen in liquid nitrogen. After thawing at RT, the samples were transferred to ddH_2O using a 70 µm sieve for larvae and primary polyps. The mesogleas were decellularized in ddH_2O by repeated pipetting with a flamed glass pipette and frequent water changes. The progress of decellularization was checked repeatedly by phase contrast microscopy at 60x magnification using a Nikon 80i microscope. As a final quality control, a few sample mesogleas were stained with DAPI for 10 min in PBS to visualize residual cells or nematocysts. Decellularized mesogleas were then picked individually for further analysis.

Sample preparation for SP3 and TMT labeling. Isolated mesogleas (larvae and primary polyps, N = 150, adults, N = 4) were dissolved in 1 M dithiothreitol (DTT) for 30 min at 90 °C and protein extraction was performed using trichloroacetic acid (TCA) precipitation. TCA pellets were resuspended in 50 µL 1% SDS, 50 mM HEPES pH 8.5. Reduction of disulfide bridges in cysteine containing proteins was performed with DTT (56 °C, 30 min, 10 mM in 50 mM HEPES, pH 8.5). Reduced cysteines were alkylated with 2-chloroacetamide (RT, in the dark, 30 min, 20 mM in 50 mM HEPES, pH 8.5). Samples were prepared using the SP3 protocol (Hughes, et al. 2014; Hughes, et al. 2019) and trypsin (sequencing grade, Promega) was added in an enzyme to protein ratio of 1:50 for overnight digestion at 37°C. Then, peptide recovery was performed in HEPES buffer by collecting the supernatant on a magnet and combining it with the second elution wash of beads with HEPES buffer. Peptides were labelled with TMT10plex (Werner, et al. 2014) Isobaric Label Reagent (ThermoFisher) according to the manufacturer's instructions. For further sample clean up an OASIS® HLB µElution Plate (Waters) was used for each sample separately. A control run was performed to be able to mix equal peptide amounts based on the MS signal in each run and samples were combined for TMT9plex accordingly. Offline high pH reverse phase fractionation was carried out on an Agilent 1200 Infinity high-performance liquid chromatography system, equipped with a Gemini C18 column (3 µm, 110 Å, 100 x 1.0 mm, Phenomenex) (Reichel, et al. 2016).

Mass spectrometry and data analysis. LC-MS/MS, Liquid Chromatography (LC) was performed as previously described for *Hydra* mesoglea (Veschgini, et al. 2023). IsobarQuant (Franken, et al. 2015) and Mascot (v2.2.07) were used to process the acquired data, which was searched against the *Nematostella vectensis* NV2 (wein_nvec200_tcsv2) protein models (https://simrbase.stowers.org/starletseaanemone) containing common contaminants and reversed sequences. The following modifications were included into the search parameters: Carbamidomethyl (C) and TMT10 (K) (fixed modification), Acetyl (Protein N-term), Oxidation (M) and TMT10 (N-term) (variable modifications). For the full scan (MS1) a mass error tolerance of 10 ppm and for MS/MS (MS2) spectra of 0.02 Da was set. Further parameters

were: trypsin as protease with an allowance of maximum two missed cleavages, a minimum peptide length of seven amino acids, and at least two unique peptides were required for a protein identification. The false discovery rate on peptide and protein level was set to 0.01. The raw output files of IsobarQuant (protein.txt – files) were processed using R. Contaminants were filtered out and only proteins that were quantified with at least two unique peptides were considered for the analysis. 5056 proteins passed the quality control filters. Log2-transformed raw TMT reporter ion intensities ('signal_sum' columns) were first cleaned for batch effects using limma (Ritchie, et al. 2015) and further normalized using variance stabilization normalization (Huber, et al. 2002) (see supplementary figure S8 for an overview of these steps). Proteins were tested for differential expression using the limma package. The replicate information was added as a factor in the design matrix given as an argument to the 'lmFit' function of limma. A protein was annotated as a hit with a false discovery rate (fdr) smaller 5% and a fold-change of at least 2. For the heatmap shown in Fig. 4A, the log2-transformed median abundance of all samples for each life stage was calculated.

***In silico* matrisome prediction.** Domains for the protein models of the SIMRbase *Nematostella vectensis* NV2 (wein_nvec200_tcsv2) transcriptome (https://simrbase.stowers.org/starletseaanemone) were *de novo* annotated using InterProScan (Jones, et al. 2014). The proteins were then filtered positively using a list of known ECM protein domains and negatively by the presence of the respective exclusive domains according to Naba et al. (Naba, et al. 2012). Signal peptides were predicted using SignalP-6.0 (Teufel, et al. 2022) and DeepLoc-2 (Thumuluri, et al. 2022). The latter was also used to predict the cellular localization of proteins. In addition, OrthoFinder vers. 2.5.4 was used with default settings to predict orthogroups from all predicted matrisomes and published matrisomes from *M. musculus*, *H. sapiens (Naba, et al. 2012), B. taurus (Listrat, et al. 2023)*, *C. japonica (Huss, et al. 2019)*, D*. melanogaster (Davis, et al. 2019)*, *D. rerio (Nauroy, et al. 2018)*, *S. mediterranea (Cote, et al. 2019)* and *C. elegans (Teuscher, et al. 2019)*. Finally, all *Nematostella* sequences were manually annotated by comparing their domain architecture to

published protein groups. Domain and orthogroup analysis were performed using custom python scripts (See supplementary data). For the orthogroup analysis, the phylogenetically hierarchical orthogroups predicted by OrthoFinder were analyzed. To prevent domain redundancy, we restricted the analysis to SMART domains for the domain comparison, as domain comparisons for other domain databases showed similar results. To achieve a better orthogroup definition we predicted additional *in silico* matrisomes for a number of available protein model datasets in non-bilaterian species: Choanoflagellata: *Monosiga brevicollis*, *Salpingoeca rosetta*; Porifera: *Amphimedon queenslandica, Ephydatia muelleri*; Ctenophora: *Mnemiopsis leidyi*, *Pleurobrachia bache*i, *Beroe ovata* (http://ryanlab.whitney.ufl.edu/bovadb), *Pleurobrachia bachei* (Moroz, et al. 2014); Placozoa: *Tricoplax adhaerens, Tricoplax spec*; Cnidaria: *Aurelia aurita* (Gold, et al. 2019), *Clytia hemispherica* (http://marimba.obs-vlfr.fr), *Exaiptasia diaphana* (Oakley, et al. 2016), *Hydra vulgaris*, *Acropora digitifera* (Shinzato, et al. 2021), *Stylophora pistillata*, *Calvadosia cruxmelitensis* (Ohdera, et al. 2019), *Morbakka virulenta* (Khalturin, et al. 2019), *Thelohanellus kitauei*, *Acropora digitifera*, *Porites asteroides* (Kenkel, et al. 2013), *Xenia spec.* (Hu, et al. 2020), Chordata: *Branchiostoma belcheri*, *Branchiostoma floridae*, *Branchiostoma lanceolatum* (Uniprot Reference Proteomes). To identify orthogroups specific to cnidarians, we filtered the OrthoFinder-derived orthogroups and phylogenetic hierarchies, selecting only those that exhibited cnidarian exclusivity.

Single cell RNA expression. The expression matrix corresponding to the matrisome was extracted from the updated single cell atlas (https://doi.org/10.1186/s12983-024-00529-z). To generate expression data, the dataset was first separated into three life-cycle stages: samples from 18 h gastrula until 4 d planula were classified as ‘larva’, samples from 5 d through 16 d primary polyps were classified as ‘primary polyp’, and all samples derived from juvenile or adult tissues were classified as ‘adult’. The full gene matrix was filtered down to only the 829 models corresponding to the matrisome. For plotting expression values, the principal tissue-type annotations were further collapsed to cluster together early and late ectodermal clusters, specification and mature cnidocyte states, and to collapse the primary germ cells and putative

stem cells into a single data partition. Differentially expressed genes were calculated across all annotated cell-type states using the Seurat vs.4 function Seurat::FindAllMarkers, requiring a return-threshold of 0.001, and a minimum detection in 20% of any cluster. Module expression scores for different gene sets (core, associated, and other for the full dataset, and 'ubiquitous', 'shared', 'specification-specific' and 'mature-specific' for the cnidocyte subset) were calculated using the Seurat function Seurat:: AddModuleScore. The cnidocyte genes were binned as described above according to summarized expression data generated by the Seurat::DotPlot function, in the 'data' matrix of the resulting ggplot.

Electron microscopy. For morphology, *Nematostella* larvae and primary polyps were processed as previously described for *Hydra* (Bottger, et al. 2012; Garg, et al. 2023). Briefly, animals were subjected to cryofixation (high-pressure freezing, freeze-substitution and epoxy resin embedding: HPF/FS: supplementary figure S2A, B, G, H) or to standard chemical fixation (glutaraldehyde, followed by OsO4, resin embedding: CF: supplementary figure S2F, L). Ultrathin sections were optionally stained with uranyl acetate and lead for general contrast enhancement or with periodic acid, thiocarbohydrazide and silver proteinate to highlight periodic acid-Schiff-positive constituents (supplementary figure S2H). For Tokuyasu-immunoelectron microscopy (Tokuyasu 1973) samples were either fixed for >3 days at RT with 4 % w/v formaldehyde solution in PHEM (5 mM HEPES, 60 mM PIPES, 10 mM EGTA, 2 mM $MgCl_2$), pH 7.0 (TOK: supplementary figure S2M, O) or by using a new modification of established HPF/FS-sample rehydration methods (Ripper, et al. 2008; Schmiedinger, et al. 2013); this modification included freeze substitution with methanol containing 3.2 % w/v formaldehyde, 0.08 % w/v uranyl acetate and 8.8 % H_2O, removal of uranyl acetate at 4° C (on ice) and partial sample rehydration and postfixation through incubation in Lavdovski's fixative for 1 hour at RT (HPF/FS/RH-Lav: supplementary figure S2N, P). Fixed samples were rinsed with PHEM buffer and further processed for thawed cryosection immunogold labelling (Tokuyasu 1973) as previously described (Garg, et al. 2023). Anti-Lam and anti-PanCol label on standard TOK-sections were visualized with goat anti-rabbit or goat anti-rat secondary

antibodies coupled to 10 nm colloidal gold (#EM.GAR10/1, #EM.GAT10/1, British Biocell). Anti-Col4 and anti-Col2c labelling was performed on HPF/FS/RH-Lav-samples by using Nanogold®-IgG Goat anti-Guinea Pig IgG or Nanogold®-IgG Goat anti-Rat IgG (H+L) (#2054, 2007, Nanoprobes), respectively, followed by silver enhancement with HQ silver™ (#2012, Nanoprobes).

Acknowledgements

This work was supported by the German Science Foundation (DFG) (Collaborative Research Center 1324 (B07) and OE 416/8-1 to S.Ö.). G.B. acknowledges the Schmeil Foundation Heidelberg for its support. We thank Josephine C. Adams for fruitful discussions.

Data Availability Statement

The MS run results can be found in the PRIDE Database (PXD045345). An R script for generating all single cell RNA expression figures and the sequences and OrthoFinder results of the *in silico* matrisomes are available at https://github.com/suatoezbek/Oezbek-matrisome.

Declaration of Interests

The authors declare no competing interests.

Figure Legends

Fig. 1. Analysis of the *Nematostella* matrisome. (A) Mesoglea from larvae, primary polyps and adults was decellularized and analyzed by mass spectrometry. In parallel, an *in silico* matrisome was predicted using a computational approach, and curated manually. (B) An *in silico* matrisome of 1812 proteins was predicted bioinformatically. The manually curated matrisome consists of 829 proteins 287 of which were confirmed by mass spectrometry. About 25 % (213) of the ECM proteins are specific to cnidarians. (C) The curated *in silico* matrisome proteins were manually annotated and sorted into core matrisome, matrisome-associated and other (mostly transmembrane) proteins (see supplementary table S1 for detailed annotations). (D) Comparison of the *Nematostella* matrisome size with published matrisomes of other species. While the complexity of the *Nematostella* core matrisome is comparable to that of vertebrates, the number of ECM-associated proteins is disproportionally lower. The *Drosophila* core matrisome is characterized by significant secondary reduction. (E) Laminin antibody stains the bilaminar structure of the BL (magenta) at the base of the epithelial cell layers, while the pan-Collagen antibody (yellow) detects the central IM. Scale bar, 10 µm. The three life stages of *Nematostella* before (F-H) and after (I-K) decellularization. The mesoglea is stained with Laminin antibody to demonstrate its structural preservation and by DAPI (cyan) to visualize residual nuclei and nematocysts. The decellularized mesoglea retains morphological structures such as tentacles (t) and mesenteries (m). Scale bars: F, G, I, J, 100 µm; H, K, 1 mm.

Fig. 2. Single cell atlas of core matrisome genes. (A) Dimensional reduction cell plot (UMAP) highlighting cell clusters showing over-abundant expression of the core matrisome, matrisome-associated, and other ECM gene sets. Expression values correspond to gene module scores for each set of genes. (B) Dotplot expression profiles of upregulated genes of the core matrisome across cell type partitions, separated across phases of the life cycle. Illustrated are the top 5 genes with expression in at least 20% of any cell state cluster,

calculated to be upregulated with a p-value of <=0.001. See supplementary table S5 for a full list of differentially expressed core matrisome genes. Larva = 18hr:4day samples; Primary Polyp = 5:16 day samples; Adult = tissue catalog from juvenile and adult specimens.

Fig. 3. Cell-type specificity of cnidocyte-expressed ECM genes. A) The distribution of cnidocyte-expressed genes categorized as 'ubiquitous' (blue: 41), 'shared' (red: 27), 'mature-specific' (green: 38), or 'specification-specific' (purple: 88). (B) Expression of the module scores of each gene-subset across the main tissue-type data partitions, illustrated on UMAP dimensional reduction. (C) Sequential gene expression activation illustrated on a dotplot of top 5 differentially expressed genes (p-value <= 0.001) for each cnidocyte cell state. Nematocyte (nem) specification shares many genes, while spirocyte specification uses a distinct gene set.

Fig. 4. Mesoglea dynamics across life stages assessed by quantitative proteomics of isolated mesoglea. (A) 2-log transformed median abundances of proteins across different life stages. The curated matrisome was filtered for proteins with a 2-fold change in any of the life stages and a false discovery rate of 0.05 using a moderated t-test (limma). The heatmap shows the 2-log transformed median abundance of 4 samples per life stage. Most proteins are upregulated in only one of the life stages. Notably, BM factors including all polydoms are upregulated in the primary polyp. Most ECM protein categories can be clearly divided into adult specific and primary polyp specific proteins underscoring the differential composition of the mesoglea at different life stages. (B-C) Volcano plots showing the differential abundance of proteins in the mesoglea extracts of the three different life stages. (B) Proteins involved in BM organization including all polydoms and in Wnt/PCP signaling are upregulated during larva-to-primary polyp transition as highlighted. (C) The adult mesoglea compared to primary polyps is characterized by an enrichment of elastic fibril components and matricellular

glycoproteins involved in wound response and regeneration. gray = non-matrimonial background, orange = insignificant, magenta = differentially abundant matrisome proteins.

Fig.5. Matrisome complexity across metazoan phyla. Matrisome sizes of published and newly generated *in silico* matrisomes of representative cnidarians and other metazoan species were plotted against their respective orthogroup count. Only proteins from orthogroups shared with at least one published matrisome were counted. Anthozoans generally show a higher matrisome complexity than medusozoan species populating a transitory region between bilaterians and non-bilaterians in the evolutionary trajectory.

References

Adamczyk P, Zenkert C, Balasubramanian PG, Yamada S, Murakoshi S, Sugahara K, Hwang JS, Gojobori T, Holstein TW, Ozbek S. 2010. A non-sulfated chondroitin stabilizes membrane tubulation in cnidarian organelles. J Biol Chem 285:25613-25623.

Adamska M, Matus DQ, Adamski M, Green K, Rokhsar DS, Martindale MQ, Degnan BM. 2007. The evolutionary origin of hedgehog proteins. Curr Biol 17:R836-837.

Adler PN. 2002. Planar signaling and morphogenesis in Drosophila. Dev Cell 2:525-535.

Aouacheria A, Geourjon C, Aghajari N, Navratil V, Deleage G, Lethias C, Exposito JY. 2006. Insights into early extracellular matrix evolution: Spongin short chain collagen-related proteins are homologous to basement membrane type IV collagens and form a novel family widely distributed in invertebrates. Molecular biology and evolution 23:2288-2302.

Appelöf A. 1900. Studien über Actinien-Entwicklung: Bergens Museums Aarbog.

Artamonova II, Mushegian AR. 2013. Genome Sequence Analysis Indicates that the Model Eukaryote Nematostella vectensis Harbors Bacterial Consorts. Applied and Environmental Microbiology 79:6868-6873.

Aufschnaiter R, Zamir EA, Little CD, Ozbek S, Munder S, David CN, Li L, Sarras MP, Jr., Zhang X. 2011. In vivo imaging of basement membrane movement: ECM patterning shapes Hydra polyps. J Cell Sci 124:4027-4038.

Babonis LS, Enjolras C, Reft AJ, Foster BM, Hugosson F, Ryan JF, Daly M, Martindale MQ. 2023. Single-cell atavism reveals an ancient mechanism of cell type diversification in a sea anemone. Nat Commun 14:885.

Balasubramanian PG, Beckmann A, Warnken U, Schnolzer M, Schuler A, Bornberg-Bauer E, Holstein TW, Ozbek S. 2012. Proteome of Hydra nematocyst. J Biol Chem 287:9672-9681.

Begnaud S, Chen T, Delacour D, Mege RM, Ladoux B. 2016. Mechanics of epithelial tissues during gap closure. Curr Opin Cell Biol 42:52-62.

Bergheim BG, Ozbek S. 2019. Extracellular matrix and morphogenesis in cnidarians: a tightly knit relationship. Extracellular Matrix 63:407-416.

Berking S. 2007. Generation of bilateral symmetry in Anthozoa: a model. J Theor Biol 246:477-490.

Bischoff MC, Lieb S, Renkawitz-Pohl R, Bogdan S. 2021. Filopodia-based contact stimulation of cell migration drives tissue morphogenesis. Nat Commun 12:791.

Bottger A, Doxey AC, Hess MW, Pfaller K, Salvenmoser W, Deutzmann R, Geissner A, Pauly B, Altstatter J, Munder S, et al. 2012. Horizontal gene transfer contributed to the evolution of extracellular surface structures: the freshwater polyp Hydra is covered by a complex fibrous cuticle containing glycosaminoglycans and proteins of the PPOD and SWT (sweet tooth) families. PLoS One 7:e52278.

Brooun M, Klimovich A, Bashkurov M, Pearson BJ, Steele RE, McNeill H. 2020. Ancestral roles of atypical cadherins in planar cell polarity. Proc Natl Acad Sci U S A 117:19310-19320.

Brunet T, King N. 2017. The Origin of Animal Multicellularity and Cell Differentiation. Dev Cell 43:124-140.

Cárdenas-León CG, Mäemets-Allas K, Klaas M, Lagus H, Kankuri E, Jaks V. 2022. Matricellular proteins in cutaneous wound healing. Frontiers in Cell and Developmental Biology 10.

Chari T, Weissbourd B, Gehring J, Ferraioli A, Leclere L, Herl M, Gao F, Chevalier S, Copley RR, Houliston E, et al. 2021. Whole-animal multiplexed single-cell RNA-seq reveals transcriptional shifts across Clytia medusa cell types. Sci Adv 7:eabh1683.

Cole AG, Jahnel SM, Kaul S, Steger J, Hagauer J, Denner A, Murguia PF, Taudes E, Zimmermann B, Reischl R, et al. 2023. Muscle cell-type diversification is driven by bHLH transcription factor expansion and extensive effector gene duplications. Nature communications 14.

Cole AG, Steger J, Hagauer J, Denner A, Ferrer Murguia P, Knabl P, Narayanaswamy S, Wick B, Montenegro JD, Technau U. 2024. Updated single cell reference atlas for the starlet anemone Nematostella vectensis. Front Zool 21:8.

Cote LE, Simental E, Reddien PW. 2019. Muscle functions as a connective tissue and source of extracellular matrix in planarians. Nature communications 10.

David CN, Ozbek S, Adamczyk P, Meier S, Pauly B, Chapman J, Hwang JS, Gojobori T, Holstein TW. 2008. Evolution of complex structures: minicollagens shape the cnidarian nematocyst. Trends Genet 24:431-438.

Davis MN, Horne-Badovinac S, Naba A. 2019. In-silico definition of the Drosophila melanogaster matrisome. Matrix Biol Plus 4:100015.

Day RM, Lenhoff HM. 1981. Hydra mesoglea: a model for investigating epithelial cell-basement membrane interactions. Science 211:291-294.

Deutzmann R, Fowler S, Zhang X, Boone K, Dexter S, Boot-Handford RP, Rachel R, Sarras MP, Jr. 2000. Molecular, biochemical and functional analysis of a novel and developmentally important fibrillar collagen (Hcol-I) in hydra. Development 127:4669-4680.

Draper GW, Shoemark DK, Adams JC. 2019. Modelling the early evolution of extracellular matrix from modern Ctenophores and Sponges. Extracellular Matrix 63:389-405.

DuBuc TQ, Traylor-Knowles N, Martindale MQ. 2014. Initiating a regenerative response; cellular and molecular features of wound healing in the cnidarian Nematostella vectensis. BMC Biol 12:24.

Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol 20:238.

Engel J. 1996. Domain organizations of modular extracellular matrix proteins and their evolution. Matrix Biology 15:295-299.

Engel U, Ozbek S, Engel R, Petri B, Lottspeich F, Holstein TW. 2002. NOWA, a novel protein with minicollagen Cys-rich domains involved in nematocyst formation in *Hydra*. J. Cell Sci. 115.

Epp L, Smid I, Tardent P. 1986. Synthesis of the mesoglea by ectoderm and endoderm in reassembled hydra. J Morphol 189:271-279.

Exposito JY, Larroux C, Cluzel C, Valcourt U, Lethias C, Degnan BM. 2008. Demosponge and sea anemone fibrillar collagen diversity reveals the early emergence of A/C clades and the maintenance of the modular structure of type V/XI collagens from sponge to human. J Biol Chem 283:28226-28235.

Fidler AL, Boudko SP, Rokas A, Hudson BG. 2018. The triple helix of collagens - an ancient protein structure that enabled animal multicellularity and tissue evolution. Journal of Cell Science 131.

Fidler AL, Darris CE, Chetyrkin SV, Pedchenko VK, Boudko SP, Brown KL, Jerome WG, Hudson JK, Rokas A, Hudson BG. 2017. Collagen IV and basement membrane at the evolutionary dawn of metazoan tissues. Elife 6.

Fidler AL, Vanacore RM, Chetyrkin SV, Pedchenko VK, Bhave G, Yin VP, Stothers CL, Rose KL, McDonald WH, Clark TA, et al. 2014. A unique covalent bond in basement membrane is a primordial innovation for tissue evolution. Proceedings of the National Academy of Sciences of the United States of America 111:331-336.

Fowler SJ, Jose S, Zhang X, Deutzmann R, Sarras MP, Jr., Boot-Handford RP. 2000. Characterization of hydra type IV collagen. Type IV collagen is essential for head regeneration and its expression is up-regulated upon exposure to glucose. J Biol Chem 275:39589-39599.

Franken H, Mathieson T, Childs D, Sweetman GM, Werner T, Togel I, Doce C, Gade S, Bantscheff M, Drewes G, et al. 2015. Thermal proteome profiling for unbiased identification of direct and indirect drug targets using multiplexed quantitative mass spectrometry. Nat Protoc 10:1567-1593.

Gao B. 2012. Wnt regulation of planar cell polarity (PCP). Curr Top Dev Biol 101:263-295.

Garg N, Stibler UK, Eismann B, Mercker M, Bergheim BG, Linn A, Tuchscherer P, Engel U, Redl S, Marciniak-Czochra A, et al. 2023. Non-muscle myosin II drives critical steps of nematocyst morphogenesis. iScience 26:106291.

Gilges D, Vinit MA, Callebaut I, Coulombel L, Cacheux V, Romeo PH, Vigon I. 2000. Polydom: a secreted protein with pentraxin, complement control protein, epidermal growth factor and von Willebrand factor A domains. Biochem J 352 Pt 1:49-59.
Gold DA, Katsuki T, Li Y, Yan X, Regulski M, Ibberson D, Holstein T, Steele RE, Jacobs DK, Greenspan RJ. 2019. The genome of the jellyfish Aurelia and the evolution of animal complexity. Nat Ecol Evol 3:96-104.
Gosline JM. 1971a. Connective Tissue Mechanics of Metridium-Senile .1. Structural and Compositional Aspects. Journal of Experimental Biology 55:763-774.
Gosline JM. 1971b. Connective Tissue Mechanics of Metridium-Senile .2. Visco-Elastic Properties and Macromolecular Model. Journal of Experimental Biology 55:775-795.
Gubb D, Garcia-Bellido A. 1982. A genetic analysis of the determination of cuticular polarity during development in Drosophila melanogaster. J Embryol Exp Morphol 68:37-57.
He F, Xiong W, Yu X, Espinoza-Lewis R, Liu C, Gu S, Nishita M, Suzuki K, Yamada G, Minami Y, et al. 2008. Wnt5a regulates directional cell migration and cell proliferation via Ror2-mediated noncanonical pathway in mammalian palate development. Development 135:3871-3879.
Heljasvaara R, Aikio M, Ruotsalainen H, Pihlajaniemi T. 2017. Collagen XVIII in tissue homeostasis and dysregulation - Lessons learned from model organisms and human patients. Matrix Biology 57-58:55-75.
Hikasa H, Shibata M, Hiratani I, Taira M. 2002. The Xenopus receptor tyrosine kinase Xror2 modulates morphogenetic movements of the axial mesoderm and neuroectoderm via Wnt signaling. Development 129:5227-5239.
Hohenester E, Engel J. 2002. Domain structure and organisation in extracellular matrix proteins. Matrix Biology 21:115-128.
Holland JW, Okamura B, Hartikainen H, Secombes CJ. 2011. A novel minicollagen gene links cnidarians and myxozoans. Proc Biol Sci 278:546-553.
Hu M, Zheng X, Fan CM, Zheng Y. 2020. Lineage dynamics of the endosymbiotic cell type in the soft coral Xenia. Nature 582:534-538.
Huber W, von Heydebreck A, Sultmann H, Poustka A, Vingron M. 2002. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. Bioinformatics 18 Suppl 1:S96-104.
Hughes CS, Foehr S, Garfield DA, Furlong EE, Steinmetz LM, Krijgsveld J. 2014. Ultrasensitive proteome analysis using paramagnetic bead technology. Molecular Systems Biology 10:757.
Hughes CS, Moggridge S, Muller T, Sorensen PH, Morin GB, Krijgsveld J. 2019. Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. Nat Protoc 14:68-85.
Huss DJ, Saias S, Hamamah S, Singh JM, Wang JH, Dave M, Kim J, Eberwine J, Lansford R. 2019. Avian Primordial Germ Cells Contribute to and Interact With the Extracellular Matrix During Early Migration. Frontiers in Cell and Developmental Biology 7.
Hwang JS, Takaku Y, Momose T, Adamczyk P, Ozbek S, Ikeo K, Khalturin K, Hemmrich G, Bosch TC, Holstein TW, et al. 2010. Nematogalectin, a nematocyst protein with GlyXY and galectin domains, demonstrates nematocyte-specific alternative splicing in Hydra. Proc Natl Acad Sci U S A 107:18539-18544.
Hynes RO. 2012. The evolution of metazoan extracellular matrix. Journal of Cell Biology 196:671-679.
Hynes RO. 2009. The extracellular matrix: not just pretty fibrils. Science 326:1216-1219.
Hynes RO, Naba A. 2012. Overview of the matrisome--an inventory of extracellular matrix constituents and functions. Cold Spring Harb Perspect Biol 4:a004903.
Jekely G, Keijzer F, Godfrey-Smith P. 2015. An option space for early neural evolution. Philos Trans R Soc Lond B Biol Sci 370.
Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30:1236-1240.

Kadler KE, Baldock C, Bella J, Boot-Handford RP. 2007. Collagens at a glance. Journal of Cell Science 120:1955-1958.
Kenkel CD, Meyer E, Matz MV. 2013. Gene expression under chronic heat stress in populations of the mustard hill coral (Porites astreoides) from different thermal environments. Mol Ecol 22:4322-4334.
Khalturin K, Shinzato C, Khalturina M, Hamada M, Fujie M, Koyanagi R, Kanda M, Goto H, Anton-Erxleben F, Toyokawa M, et al. 2019. Medusozoan genomes inform the evolution of the jellyfish body plan. Nat Ecol Evol 3:811-822.
King N, Westbrook MJ, Young SL, Kuo A, Abedin M, Chapman J, Fairclough S, Hellsten U, Isogai Y, Letunic I, et al. 2008. The genome of the choanoflagellate Monosiga brevicollis and the origin of metazoans. Nature 451:783-788.
Kumburegama S, Wijesena N, Xu R, Wikramanayake AH. 2011. Strabismus-mediated primary archenteron invagination is uncoupled from Wnt/beta-catenin-dependent endoderm cell fate specification in Nematostella vectensis (Anthozoa, Cnidaria): Implications for the evolution of gastrulation. Evodevo 2:2.
Kurz EM, Holstein TW, Petri BM, Engel J, David CN. 1991. Mini-collagens in hydra nematocytes. J Cell Biol 115:1159-1169.
Lebouvier M, Miramon-Puertolas P, Steinmetz PRH. 2022. Evolutionarily conserved aspects of animal nutrient uptake and transport in sea anemone vitellogenesis. Curr Biol 32:4620-4630 e4625.
Listrat A, Boby C, Tournayre J, Jousse C. 2023. Bovine extracellular matrix proteins and potential role in meat quality: First in silico Bos taurus compendium. J Proteomics 279:104891.
Lommel M, Strompen J, Hellewell AL, Balasubramanian GP, Christofidou ED, Thomson AR, Boyle AL, Woolfson DN, Puglisi K, Hartl M, et al. 2018. Hydra Mesoglea Proteome Identifies Thrombospondin as a Conserved Component Active in Head Organizer Restriction. Sci Rep 8:11753.
Maegele I, Rupp S, Ozbek S, Guse A, Hambleton EA, Holstein TW. 2023. A predatory gastrula leads to symbiosis-independent settlement in Aiptasia. Proc Natl Acad Sci U S A 120:e2311872120.
Matakatsu H, Blair SS. 2004. Interactions between Fat and Dachsous and the regulation of planar cell polarity in the Drosophila wing. Development 131:3785-3794.
Meyer F, Moussian B. 2009. Drosophila multiplexin (Dmp) modulates motor axon pathfinding accuracy. Dev Growth Differ 51:483-498.
Moroz LL, Kocot KM, Citarella MR, Dosung S, Norekian TP, Povolotskaya IS, Grigorenko AP, Dailey C, Berezikov E, Buckley KM, et al. 2014. The ctenophore genome and the evolutionary origins of neural systems. Nature 510:109-114.
Naba A. 2024. Mechanisms of assembly and remodelling of the extracellular matrix. Nat Rev Mol Cell Biol.
Naba A, Clauser KR, Hoersch S, Liu H, Carr SA, Hynes RO. 2012. The matrisome: in silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices. Mol Cell Proteomics 11:M111 014647.
Nauroy P, Hughes S, Naba A, Ruggiero F. 2018. The in-silico zebrafish matrisome: A new tool to study extracellular matrix gene and protein functions. Matrix Biology 65:5-13.
Nichols SA, Roberts BW, Richter DJ, Fairclough SR, King N. 2012. Origin of metazoan cadherin diversity and the antiquity of the classical cadherin/beta-catenin complex. Proc Natl Acad Sci U S A 109:13046-13051.
Nishita M, Yoo SK, Nomachi A, Kani S, Sougawa N, Ohta Y, Takada S, Kikuchi A, Minami Y. 2006. Filopodia formation mediated by receptor tyrosine kinase Ror2 is required for Wnt5a-induced cell migration. J Cell Biol 175:555-562.
Novinec M, Kordis D, Turk V, Lenarcic B. 2006. Diversity and evolution of the thyroglobulin type-1 domain superfamily. Molecular biology and evolution 23:744-755.
Oakley CA, Ameismeier MF, Peng L, Weis VM, Grossman AR, Davy SK. 2016. Symbiosis induces widespread changes in the proteome of the model cnidarian Aiptasia. Cell Microbiol 18:1009-1023.

Ohdera A, Ames CL, Dikow RB, Kayal E, Chiodin M, Busby B, La S, Pirro S, Collins AG, Medina M, et al. 2019. Box, stalked, and upside-down? Draft genomes from diverse jellyfish (Cnidaria, Acraspeda) lineages: Alatina alata (Cubozoa), Calvadosia cruxmelitensis (Staurozoa), and Cassiopea xamachana (Scyphozoa). Gigascience 8.
Ou Q, Shu D, Zhang Z, Han J, Van Iten H, Cheng M, Sun J, Yao X, Wang R, Mayer G. 2022. Dawn of complex animal food webs: A new predatory anthozoan (Cnidaria) from Cambrian. Innovation (Camb) 3:100195.
Ozbek S. 2011. The cnidarian nematocyst: a miniature extracellular matrix within a secretory vesicle. Protoplasma 248:635-640.
Ozbek S, Balasubramanian PG, Chiquet-Ehrismann R, Tucker RP, Adams JC. 2010. The evolution of extracellular matrix. Mol Biol Cell 21:4300-4305.
Reichel M, Liao Y, Rettel M, Ragan C, Evers M, Alleaume AM, Horos R, Hentze MW, Preiss T, Millar AA. 2016. In Planta Determination of the mRNA-Binding Proteome of Arabidopsis Etiolated Seedlings. Plant Cell 28:2435-2452.
Richards GS, Rentzsch F. 2014. Transgenic analysis of a SoxB gene reveals neural progenitor cells in the cnidarian Nematostella vectensis. Development 141:4681-4689.
Ripper D, Schwarz H, Stierhof YD. 2008. Cryo-section immunolabelling of difficult to preserve specimens: advantages of cryofixation, freeze-substitution and rehydration. Biology of the Cell 100:109-123.
Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 43:e47.
Rokas A. 2008. The origins of multicellularity and the early history of the genetic toolkit for animal development. Annu Rev Genet 42:235-251.
Samuelov L, Li Q, Bochner R, Najor NA, Albrecht L, Malchin N, Goldsmith T, Grafi-Cohen M, Vodo D, Fainberg G, et al. 2017. SVEP1 plays a crucial role in epidermal differentiation. Exp Dermatol 26:423-430.
Sarras MP, Jr. 2012. Components, structure, biogenesis and function of the Hydra extracellular matrix in regeneration, pattern formation and cell differentiation. Int J Dev Biol 56:567-576.
Sarras MP, Jr., Madden ME, Zhang XM, Gunwar S, Huff JK, Hudson BG. 1991. Extracellular matrix (mesoglea) of Hydra vulgaris. I. Isolation and characterization. Dev Biol 148:481-494.
Sato-Nishiuchi R, Doiguchi M, Morooka N, Sekiguchi K. 2023. Polydom/SVEP1 binds to Tie1 and promotes migration of lymphatic endothelial cells. J Cell Biol 222.
Sato-Nishiuchi R, Nakano I, Ozawa A, Sato Y, Takeichi M, Kiyozumi D, Yamazaki K, Yasunaga T, Futaki S, Sekiguchi K. 2012. Polydom/SVEP1 Is a Ligand for Integrin α9β1. Journal of Biological Chemistry 287:25615-25630.
Schaffer AA, Bazarsky M, Levy K, Chalifa-Caspi V, Gat U. 2016. A transcriptional time-course analysis of oral vs. aboral whole-body regeneration in the Sea anemone. Bmc Genomics 17.
Schambony A, Wedlich D. 2007. Wnt-5A/Ror2 regulate expression of XPAPC through an alternative noncanonical signaling pathway. Dev Cell 12:779-792.
Schmiedinger T, Vogel GF, Eiter O, Pfaller K, Kaufmann WA, Flörl A, Gutleben K, Schönherr S, Witting B, Lechleitner TW, et al. 2013. Cryo-Immunoelectron Microscopy of Adherent Cells Improved by the Use of Electrospun Cell Culture Substrates. Traffic 14:886-894.
Schultz DT, Haddock SHD, Bredeson JV, Green RE, Simakov O, Rokhsar DS. 2023. Ancient gene linkages support ctenophores as sister to other animals. Nature 618:110-117.
Schwarz RS, Bosch TC, Cadavid LF. 2008. Evolution of polydom-like molecules: identification and characterization of cnidarian polydom (Cnpolydom) in the basal metazoan Hydractinia. Dev Comp Immunol 32:1192-1210.
Sebe-Pedros A, Roger AJ, Lang FB, King N, Ruiz-Trillo I. 2010. Ancient origin of the integrin-mediated adhesion and signaling machinery. Proc Natl Acad Sci U S A 107:10142-10147.

Shimizu H, Aufschnaiter R, Li L, Sarras MP, Jr., Borza DB, Abrahamson DR, Sado Y, Zhang X. 2008. The extracellular matrix of hydra is a porous sheet and contains type IV collagen. Zoology (Jena) 111:410-418.

Shinzato C, Khalturin K, Inoue J, Zayasu Y, Kanda M, Kawamitsu M, Yoshioka Y, Yamashita H, Suzuki G, Satoh N. 2021. Eighteen Coral Genomes Reveal the Evolutionary Origin of Acropora Strategies to Accommodate Environmental Changes. Mol Biol Evol 38:16-30.

Simons M, Mlodzik M. 2008. Planar cell polarity signaling: from fly development to human disease. Annu Rev Genet 42:517-540.

Steger J, Cole AG, Denner A, Lebedeva T, Genikhovich G, Ries A, Reischl R, Taudes E, Lassnig M, Technau U. 2022. Single-cell transcriptomics identifies conserved regulators of neuroglandular lineages. Cell Rep 40:111370.

Steinmetz PRH, Aman A, Kraus JEM, Technau U. 2017. Gut-like ectodermal tissue in a sea anemone challenges germ layer homology. Nat Ecol Evol 1:1535-1542.

Stokkermans A, Chakrabarti A, Subramanian K, Wang L, Yin S, Moghe P, Steenbergen P, Monke G, Hiiragi T, Prevedel R, et al. 2022. Muscular hydraulics drive larva-polyp morphogenesis. Curr Biol 32:4707-4718 e4708.

Teufel F, Armenteros JJA, Johansen AR, Gislason MH, Pihl SI, Tsirigos KD, Winther O, Brunak S, von Heijne G, Nielsen H. 2022. SignalP 6.0 predicts all five types of signal peptides using protein language models. Nature Biotechnology 40:1023-1025.

Teuscher AC, Jongsma E, Davis MN, Statzer C, Gebauer JM, Naba A, Ewald CY. 2019. The in-silico characterization of the Caenorhabditis elegans matrisome and proposal of a novel collagen classification. Matrix Biol Plus 1:100001.

Thumuluri V, Almagro Armenteros JJ, Johansen AR, Nielsen H, Winther O. 2022. DeepLoc 2.0: multi-label subcellular localization prediction using protein language models. Nucleic Acids Res 50:W228-W234.

Tokuyasu KT. 1973. A technique for ultracryotomy of cell suspensions and tissues. J Cell Biol 57:551-565.

Tournière O, Dolan D, Richards GS, Sunagar K, Columbus-Shenkar YY, Moran Y, Rentzsch F. 2020. NvPOU4/Brain3 Functions as a Terminal Selector Gene in the Nervous System of the Cnidarian. Cell Reports 30:4473-4489.

Tucker RP, Adams JC. 2014. Adhesion Networks of Cnidarians: A Postgenomic View. International Review of Cell and Molecular Biology, Vol 308 308:323-377.

Tucker RP, Shibata B, Blankenship TN. 2011. Ultrastructure of the mesoglea of the sea anemone Nematostella vectensis (Edwardsiidae). Invertebrate Biology 130:11-24.

Veschgini M, Suzuki R, Kling S, Petersen HO, Bergheim BG, Abuillan W, Linke P, Kaufmann S, Burghammer M, Engel U, et al. 2023. Wnt/beta-catenin signaling induces axial elasticity patterns of Hydra extracellular matrix. iScience 26:106416.

Werner T, Sweetman G, Savitski MF, Mathieson T, Bantscheff M, Savitski MM. 2014. Ion coalescence of neutron encoded TMT 10-plex reporter ions. Anal Chem 86:3594-3601.

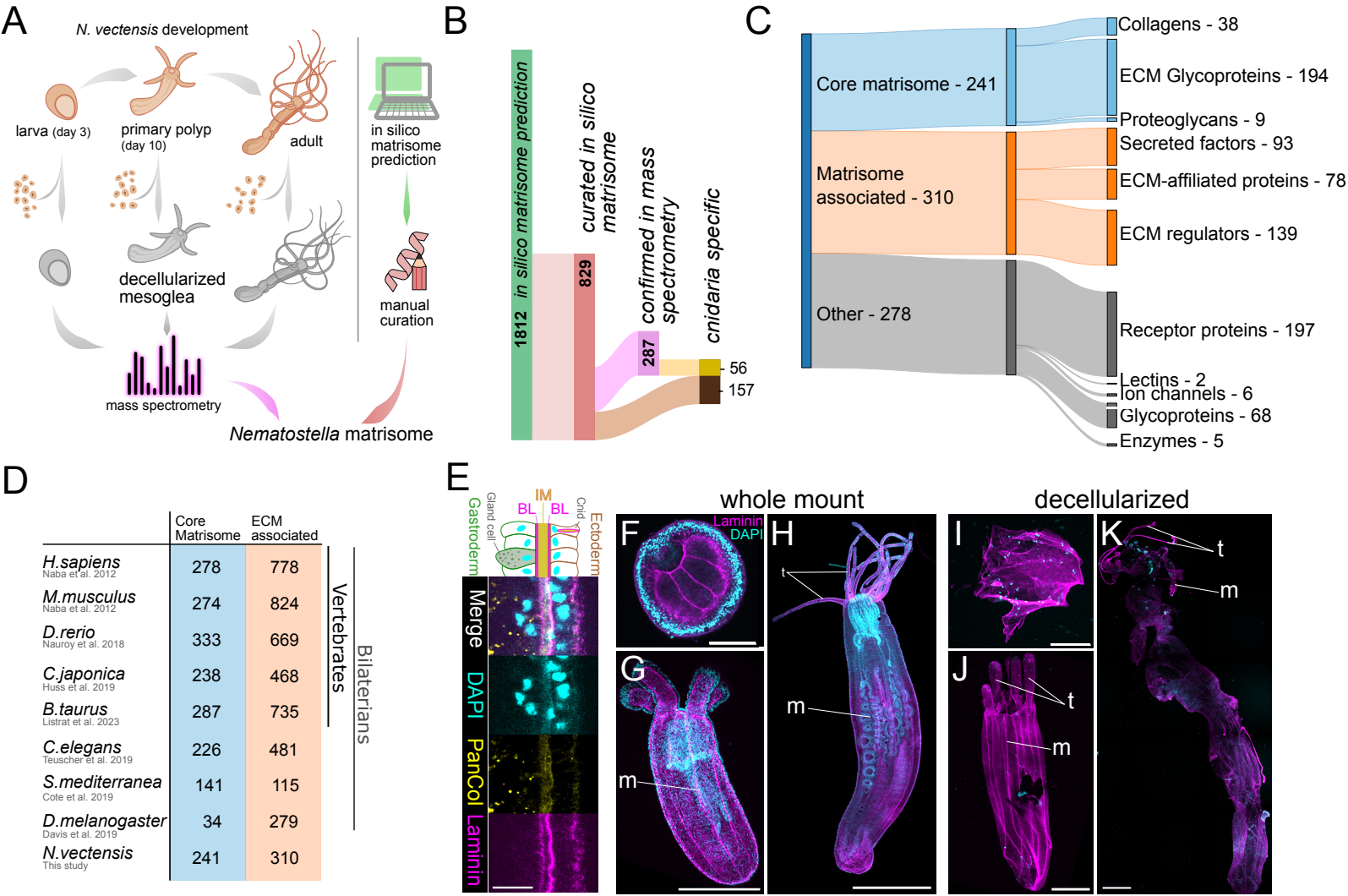
Wijesena N, Sun HY, Kumburegama S, Wikramanayake AH. 2022. Distinct Frizzled receptors independently mediate endomesoderm specification and primary archenteron invagination during gastrulation in Nematostella. Developmental biology 481:215-225.

Zenkert C, Takahashi T, Diesner MO, Ozbek S. 2011. Morphological and molecular analysis of the Nematostella vectensis cnidom. PLoS One 6:e22725.

Zhang X, Boot-Handford RP, Huxley-Jones J, Forse LN, Mould AP, Robertson DL, Lili, Athiyal M, Sarras MP, Jr. 2007. The collagens of hydra provide insight into the evolution of metazoan extracellular matrices. J Biol Chem 282:6792-6802.

Zhou Y, Chen L, Hao ST, Cao XK, Ni SS. 2024. Zebrafish ANGPT4, member of fibrinogen-related proteins, is an LTA-, LPSand PGN-binding protein with a bacteriolytic activity. Fish & Shellfish Immunology 147.

Fig.1



	Core Matrisome	ECM associated
H.sapiens Naba et al. 2012	278	778
M.musculus Naba et al. 2012	274	824
D.rerio Nauroy et al. 2018	333	669
C.japonica Huss et al. 2019	238	468
B.taurus Listrat et al. 2023	287	735
C.elegans Teuscher et al. 2019	226	481
S.mediterranea Cote et al. 2019	141	115
D.melanogaster Davis et al. 2019	34	279
N.vectensis This study	241	310

Vertebrates

Bilaterians

Fig. 2

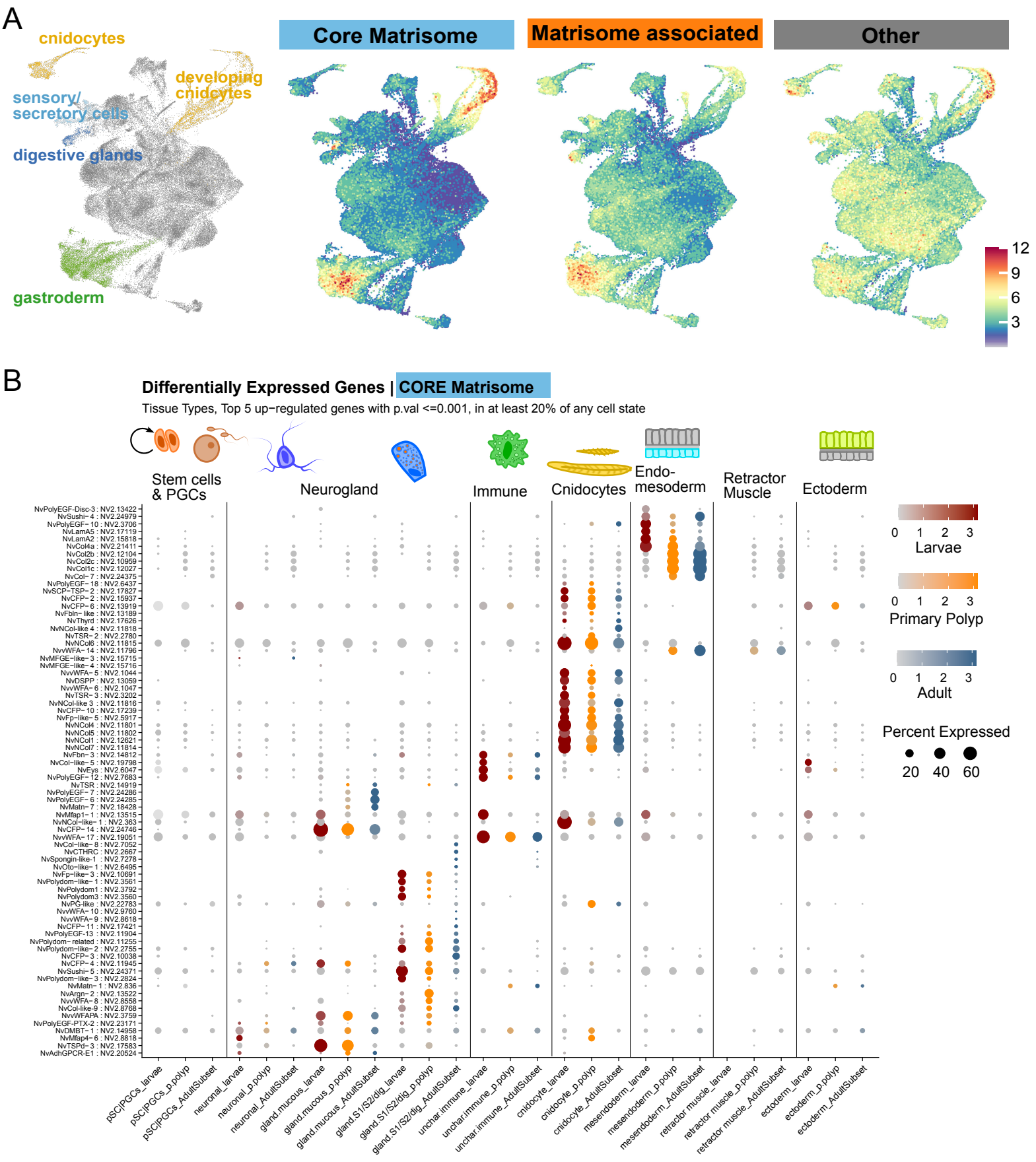

A
cnidocytes
developing cnidocytes
sensory/ secretory cells
digestive glands
gastroderm
Core Matrisome
Matrisome associated
Other
12
9
6
3
B
Differentially Expressed Genes | CORE Matrisome
Tissue Types, Top 5 up−regulated genes with p.val <=0.001, in at least 20% of any cell state
Stem cells & PGCs
Neurogland
Immune
Cnidocytes
Endo-mesoderm
Retractor Muscle
Ectoderm
Larvae
Primary Polyp
Adult
Percent Expressed
20 40 60
pSC|PGCs_larvae
pSC|PGCs_p.polyp
pSC|PGCs_AdultSubset
neuronal_larvae
neuronal_p.polyp
neuronal_AdultSubset
gland.mucous_larvae
gland.mucous_p.polyp
gland.mucous_AdultSubset
gland.S1/S2/dig_larvae
gland.S1/S2/dig_p.polyp
gland.S1/S2/dig_AdultSubset
unchar.immune_larvae
unchar.immune_p.polyp
unchar.immune_AdultSubset
cnidocyte_larvae
cnidocyte_p.polyp
cnidocyte_AdultSubset
mesendoderm_larvae
mesendoderm_p.polyp
mesendoderm_AdultSubset
retractor muscle_larvae
retractor muscle_p.polyp
retractor muscle_AdultSubset
ectoderm_larvae
ectoderm_p.polyp
ectoderm_AdultSubset

Fig. 3

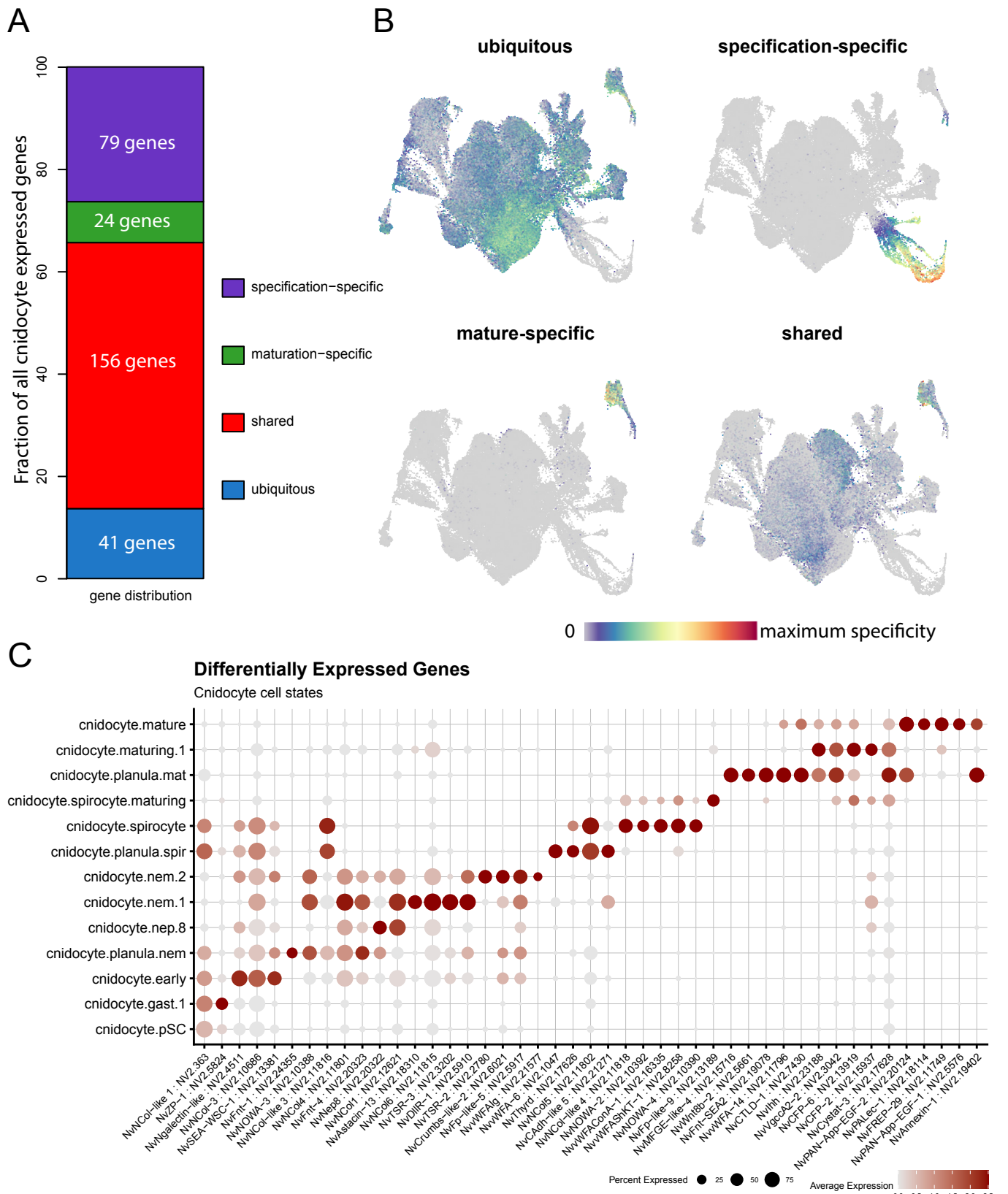

A
Fraction of all cnidocyte expressed genes
100
80
60
40
20
0
79 genes
24 genes
156 genes
41 genes
gene distribution
specification−specific
maturation−specific
shared
ubiquitous
B
ubiquitous
specification-specific
mature-specific
shared
0
maximum specificity
C
Differentially Expressed Genes
Cnidocyte cell states
cnidocyte.mature
cnidocyte.maturing.1
cnidocyte.planula.mat
cnidocyte.spirocyte.maturing
cnidocyte.spirocyte
cnidocyte.planula.spir
cnidocyte.nem.2
cnidocyte.nem.1
cnidocyte.nep.8
cnidocyte.planula.nem
cnidocyte.early
cnidocyte.gast.1
cnidocyte.pSC
NvNCol−like 1 : NV2.363
NvZP−1 : NV2.5824
NvNgalectin−like : NV2.4511
NvNCol−3 : NV2.10686
NvSEA−WSC−1 : NV2.13381
NvFnt−1 : NV2.24355
NvNOWA−3 : NV2.10388
NvNCol−like 3 : NV2.11816
NvNCol4 : NV2.11801
NvFnt−4 : NV2.20323
NvNep8 : NV2.20322
NvNCol1 : NV2.12621
NvAstacin−13 : NV2.18310
NvNCol6 : NV2.11815
NvTSR−3 : NV2.3202
NvDlIR−1 : NV2.5910
NvTSR−2 : NV2.2780
NvCrumbs−like−2 : NV2.6021
NvFp−like−5 : NV2.5917
NvWFAlg : NV2.21577
NvWFA−6 : NV2.1047
NvThyrd : NV2.17626
NvNCol5 : NV2.11802
NvCAdh−like 5 : NV2.21271
NvNCol−like 4 : NV2.11818
NvNOWA−2 : NV2.10392
NvWFAConA−1 : NV2.16335
NvWFAShKT−1 : NV2.8258
NvNOWA−4 : NV2.10390
NvFp−like−9 : NV2.13189
NvMFGE−like−4 : NV2.15716
NvWnt8b−2 : NV2.5661
NvFnt−SEA2 : NV2.19078
NvWFA−14 : NV2.11796
NvCTLD−1 : NV2.7430
NvIlh : NV2.23188
NvVgcA2−2 : NV2.3042
NvCFP−6 : NV2.13919
NvCFP−2 : NV2.15937
NvCystat−3 : NV2.17628
NvPAN−App−EGF−2 : NV2.20124
NvPALec−1 : NV2.18114
NvFREP−29 : NV2.11749
NvPAN−App−EGF−1 : NV2.5576
NvAnnexin−1 : NV2.19402
Percent Expressed
25
50
75
Average Expression
0.0
0.5
1.0
1.5
2.0
2.5

Fig. 4

A

CORE MATRISOME | MATRISOME ASSOCIATED | OTHER

Collagens

NvCol-like-9 (NV2t008768001.1)
NvLpxACol (NV2t011419002.1)
NvCol6-like-2 (NV2t023710001.1)
NvSpongin-like-2 (NV2t007899001.1)
NvCol6-like-4 (NV2t014089001.1)
NvSpongin-like-1 (NV2t007278001.1)
NvCol6-like-1 (NV2t023164001.1)
NvCol18 (NV2t011420001.1)
NvCol4a (NV2t021411001.1)

1.0 0.0 −1.0 −2.0 −3.0

ECM Glycoproteins

NvvWFA-WSC (NV2t021576001.1)
NvLamA-like (NV2t015819001.1)
NvLamB1-1 (NV2t001719001.1)
NvLamA2 (NV2t015818001.1)
NvvWFA-9 (NV2t008618001.1)
NvvWFA-4 (NV2t014087001.1)
NvvWFA-15 (NV2t014073001.1)
NvvWFA-3 (NV2t014076001.1)
NvDSLvWFA (NV2t009243001.1)
NvHmcn1-like 1 (NV2t023125001.1)
NvPolyEGF-9 (NV2t019497002.1)
NvTSP-1 (NV2t023884001.1)
NvSAMEGF (NV2t002046001.1)
NvvWFA-WAP (NV2t011346003.1)
NvThyr-2 (NV2t014040001.1)
NvFbn-2 (NV2t022932001.1)
NvFbn-1 (NV2t022930001.1)
NvFbn (NV2t022931001.1)
NvCFP-13 (NV2t023447001.1)
NvArgn-1 (NV2t014093001.1)
NvEmilin1 (NV2t010668001.1)
NvThyr-1 (NV2t014419001.1)
NvCFP-3 (NV2t010038001.1)
NvPostn (NV2t020449001.1)
NvFbln (NV2t003768001.1)
NvvWFA-13 (NV2t010983001.1)
EGF-trlLpxA (NV2t012181001.1)
NvMfap1 (NV2t013515001.1)
NvDMBT-1 (NV2t014958001.1)
NvTSR (NV2t014919002.1)
NvCFP-11 (NV2t017421001.1)
NvvWFA-12 (NV2t010017001.1)
NvTSR-1 (NV2t000737001.1)
NvTSR-3 (NV2t003202001.1)
NvMatn-7 (NV2t018428001.1)
NvDisc-lg (NV2t001116001.1)
NvThyr-3 (NV2t019319002.1)
NvFp-like-5 (NV2t005917001.1)
NvvWFA-11 (NV2t009856002.1)
NvPolyvWFA (NV2t008635001.1)
NvPolydom-like-3 (NV2t002824002.1)
NvvWFA-TSR (NV2t014005001.1)
NvPolydom-like-2 (NV2t002755001.1)
NvvWFA-21 (NV2t021578001.1)
NvTSR-2 (NV2t002780001.1)
NvPolydom-related (NV2t012555001.1)
NvPolydom3 (NV2t003560001.1)
NvPolydom-like-4 (NV2t013340001.1)
NvDisc-SEA (NV2t012171001.1)
NvPolyEGF-Sushi-lg (NV2t017270001.1)
NvPolydom-like-1 (NV2t003561001.1)
NvPolydom2 (NV2t003300001.1)
NvLamA1-like (NV2t022840001.1)
NvPolydom1 (NV2t003792001.1)
NvFp-like-6 (NV2t005918002.1)

1 0 −1 −2 −3

Proteoglycans

NvPlc-like (NV2t014904001.1)
NvPlc (NV2t014907001.1)

0 −1

Larvae | Primary Polyp | Adult

ECM regulators

NvMep-13 (NV2t014862004.1)
NvADAMTS-9 (NV2t017536001.1)
NvAstacin-23 (NV2t013964001.1)
NvAstacin-1 (NV2t016729001.1)
NvMep-16 (NV2t016035001.1)
NvvWFAlg (NV2t021577002.1)
NvLox (NV2t019165003.1)
NvMep-8 (NV2t017021001.1)
NvHspe-2 (NV2t018651001.1)
NvMMP-3 (NV2t000807001.1)
NvMep-10 (NV2t005698001.1)
NvADAM-3 (NV2t013893002.1)
NvADAM-4 (NV2t018385002.1)
NvADAMTS-1 (NV2t022459002.1)
NvAstacin-12 (NV2t021687001.1)
NvM28MMP (NV2t013120001.1)
NvADAMTS-12 (NV2t021636001.1)
NvAstacin-9 (NV2t002433001.1)
NvCoag-like 5 (NV2t003648001.1)
NvMep-4 (NV2t016574001.1)
NvMMP-6 (NV2t001647001.1)
NvNaglu (NV2t004033001.1)
NvMMP-1 (NV2t008375002.1)
NvAstacin-24 (NV2t014830002.1)
NvAstacin-26 (NV2t017332002.1)
NvTIMP-3 (NV2t017964001.1)
NvCoag-related 2 (NV2t000581002.1)
NvAstacin-28 (NV2t018104001.1)
NvKSI (NV2t014070001.1)
NvKusi-2 (NV2t000320001.1)
NvIGFKz (NV2t006476001.1)
NvAstacin-3 (NV2t010250001.1)
NvSerpin-like 2 (NV2t021103001.1)
NvAstacin-like-2 (NV2t008027002.1)
NvADAMTS-18 (NV2t020821001.1)
NvAstacin-11 (NV2t004041001.1)
NvMMP-12 (NV2t020475001.1)
NvnWFAGonKz (NV2t014092001.1)
NvTIMP-1 (NV2t003671001.1)
NvADAMTS-20 (NV2t022999001.1)
NvAstacin-like-1 (NV2t014828001.1)
NvAstacin-17 (NV2t005152001.1)
NvvWFAKz-1 (NV2t014043001.1)
NvTryl-like (NV2t003632001.1)
NvMMP-2 (NV2t020474001.1)

1 0 −1 −2 −3

ECM-affiliated proteins

NvNOWA-1 (NV2t015872001.1)
NvSnaclec-like-4 (NV2t014543001.1)
NvNOWA-3 (NV2t010388002.1)
NvCTLD-8 (NV2t018585001.1)
NvSEMA5B (NV2t011758001.1)
NvAnnexin-3 (NV2t010289001.1)
NvCUB-ZP-1 (NV2t021283001.1)
NvNADPH1-2 (NV2t000021001.1)
NvNgalectin-1 (NV2t005200001.1)
NvCUB-misc-5 (NV2t003067003.1)
NvMuc-like-8 (NV2t012902001.1)
NvNOWA-2 (NV2t010392003.1)
NvCUB-ZP-2 (NV2t021285001.1)
NvNgalectin-3 (NV2t005196003.1)
NvCollectin (NV2t023165001.1)
NvSnaclec-like-3 (NV2t013847001.1)
NvNADPH1-1 (NV2t000027001.1)
NvSEMA-like-1 (NV2t007220001.1)
NvMuc-like-6 (NV2t012174001.1)

1.0 0.0 −1.0 −2.0

Secreted factors

NvVgt-2 (NV2t001922001.1)
NvKcp-like-2 (NV2t012484002.1)
NvTNF (NV2t022057001.1)
NvFst-like 1 (NV2t013423001.1)
NvFst-like 2 (NV2t023329001.1)
NvDlIR-1 (NV2t005910001.1)
NvVgt-1 (NV2t000914001.1)
NvEGF-lg (NV2t016522001.1)
NvSOM-TSP (NV2t016600001.1)

1 0 −1 −2 −3 −4 −5

Larvae | Primary Polyp | Adult

Glycoproteins

NvHml-like 2 (NV2t005728003.1)
NvHetC-like (NV2t002146001.1)
NvvThy (NV2t014046001.1)
NvVef1-2 (NV2t003719001.1)
NvUru-like-1 (NV2t012183001.1)
NvBr3-like-1 (NV2t021872001.1)
NvTLDcP-2 (NV2t006705001.1)
NvRyn1 (NV2t017096001.1)
NvCAdh-like 3 (NV2t016479001.1)
NvMAMFnt (NV2t014399001.1)
NvCAdh-like 2 (NV2t001610003.1)
NvCrumbs-like-2 (NV2t006021001.1)
NvSIntN (NV2t000650001.1)
NvCrumbs-like-5 (NV2t011903001.1)
NvTLDcP-4 (NV2t014921001.1)
NvCub-6 (NV2t013183001.1)
NvTDA6-2 (NV2t002896001.1)
NvFnt-3B-like (NV2t022291001.1)
NvSPP-like-2 (NV2t014010001.1)
NvCub-5 (NV2t005224001.1)
NvNep8 (NV2t020322001.1)
NvMAMCUB (NV2t023484001.1)

1 0 −1 −2 −3 −4

Ion channels

NvTRP (NV2t010225001.1)
NvVgccA2-1 (NV2t016117001.1)

0 −1

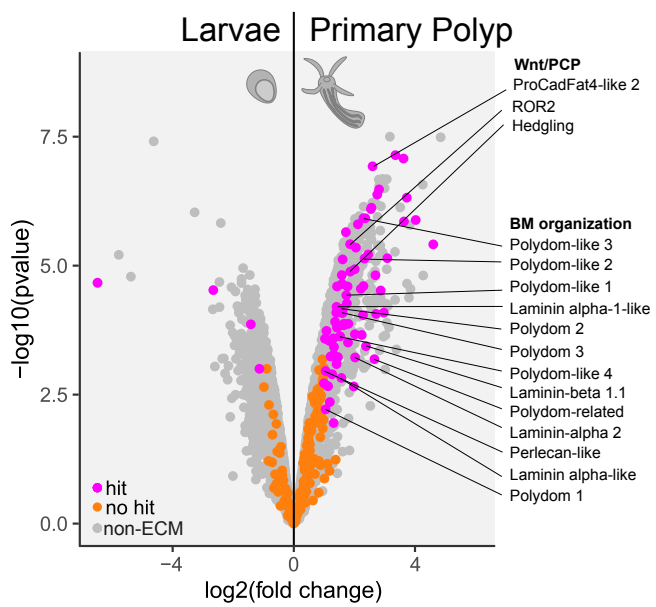
Receptor proteins

NvZP-2 (NV2t011069001.1)
NvvWFAc1-3 (NV2t001994001.1)
NvZP-1 (NV2t005824001.1)
NvWAPlgZP-2 (NV2t011632001.1)
NvCDH-1 (NV2t004715001.1)
NvWAPlgZP-1 (NV2t011345001.1)
NvIgSushi (NV2t017065001.1)
NvAntxR-3 (NV2t008406003.1)
NvAntxR-4 (NV2t008980001.1)
NvHedgeling (NV2t002229001.1)
NvSushiEGFZP-1 (NV2t012863002.1)
NvFATPCad4-like-3 (NV2t023594001.1)
NvNCAM-6 (NV2t021924001.1)
NvSushiEGFZP-2 (NV2t012864001.1)
NvFATPCad4-like-2 (NV2t023593001.1)
NvTGFbR (NV2t005822001.1)
NvAntxR-2 (NV2t008769001.1)
NvL1Cam-1 (NV2t001171001.1)
NvNrxn-like-4 (NV2t009601001.1)
NvIgCam-1 (NV2t000361004.1)
NvThyr-TSP-EGF (NV2t004066002.1)
NvVgtR-like-1 (NV2t012147001.1)
NvCDH-3 (NV2t004921001.1)
NvFat-like (NV2t003144001.1)
NvNCAM1-1 (NV2t021902003.1)
NvNCAM-related-1 (NV2t022495007.1)
NvSOL-like-3 (NV2t001288001.1)
NvSOL-like-6 (NV2t003169001.1)
NvCpne-2 (NV2t004625001.1)
NvROR2-like (NV2t006768001.1)
NvMmanR-1 (NV2t012165001.1)
NvNrxn (NV2t011326001.1)
NvMmanR-3 (NV2t012172001.1)

1 0 −1 −2 −3

Larvae | Primary Polyp | Adult

B



C

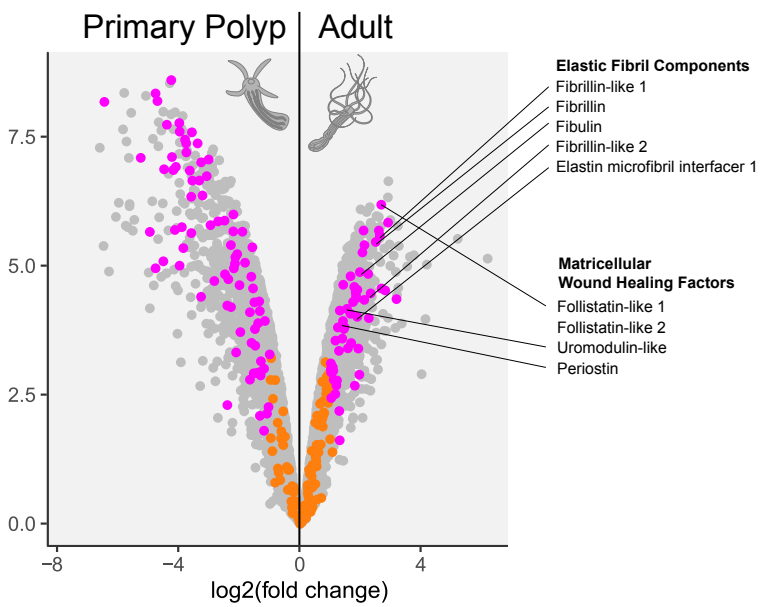
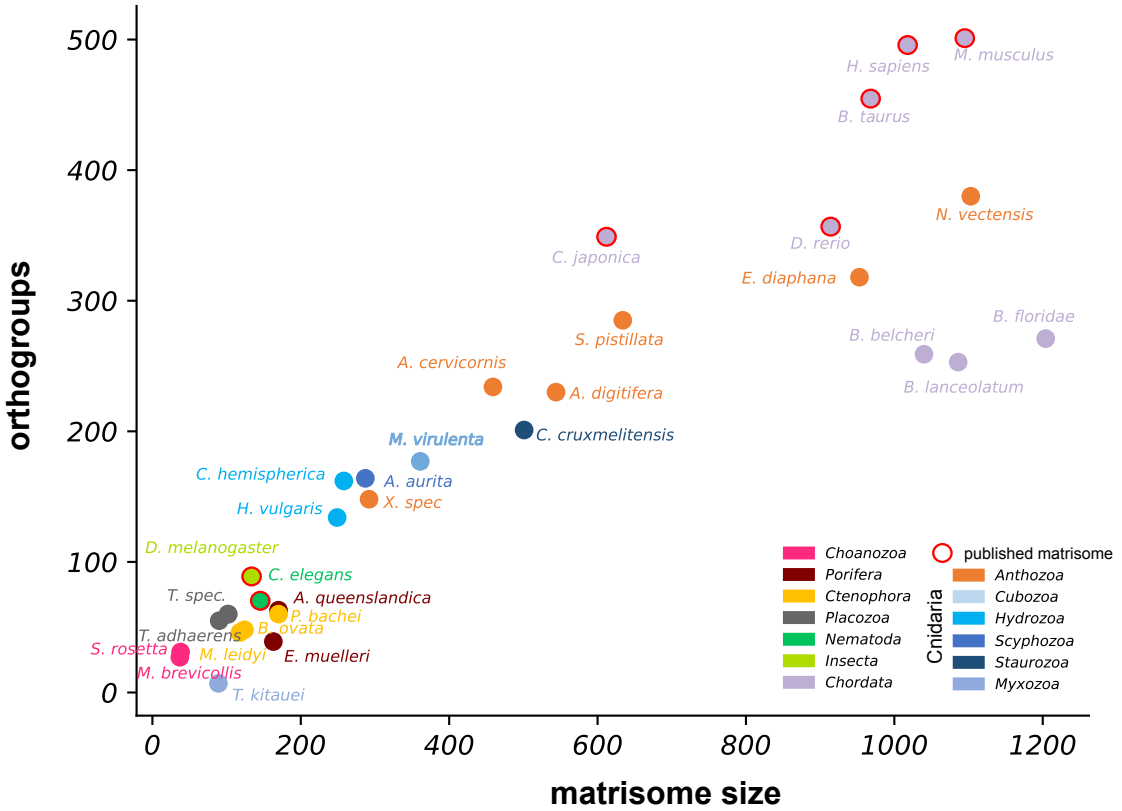


Fig. 5


orthogroups
matrisome size
0
100
200
300
400
500
0
200
400
600
800
1000
1200
H. sapiens
M. musculus
B. taurus
N. vectensis
D. rerio
C. japonica
E. diaphana
B. floridae
B. belcheri
S. pistillata
B. lanceolatum
A. cervicornis
A. digitifera
C. cruxmelitensis
M. virulenta
C. hemispherica
A. aurita
X. spec
H. vulgaris
D. melanogaster
C. elegans
T. spec.
A. queenslandica
P. bachei
B. ovata
T. adhaerens
S. rosetta
M. leidyi
E. muelleri
M. brevicollis
T. kitauei
Choanozoa
Porifera
Ctenophora
Placozoa
Nematoda
Insecta
Chordata
published matrisome
Cnidaria
Anthozoa
Cubozoa
Hydrozoa
Scyphozoa
Staurozoa
Myxozoa