Transcriptome profiling of Trichoplax adhaerens highlights its digestive epithelium and a rich set of genes for fast electrogenic and slow neuromodulatory cellular signaling

Yuen Yan Wong
University of Toronto - Mississauga
Phuong Le
University of Toronto - Mississauga
Wassim Elkhatib
University of Toronto - Mississauga
Thomas Piekut
University of Toronto - Mississauga
Adriano Senatore (adriano.senatore@utoronto.ca)
University of Toronto - Mississauga https://orcid.org/0000-0001-5180-1180

Research article

Keywords: ,

Posted Date: September 17th, 2019

DOI: https://doi.org/10.21203/rs.2.14504/v1

Abstract

Background Trichoplax adhaerens is a fascinating early-diverging animal that lacks a nervous system and synapses, and yet is capable of directed motile feeding behavior culminating in the external digestion of microorganisms by secreted hydrolytic enzymes. The mechanisms by which Trichoplax cells communicate with each other to coordinate their activity and behavior is unclear, though recent studies have suggested that secreted regulatory peptides might be involved.

Results Here, we generated a high quality mRNA transcriptome of Trichoplax adhaerens , and predicted secreted proteins to identify gene homologues for digestion, development, immunity, cell adhesion, and peptide signaling. Detailed annotation of the expressed Trichoplax gene set also identified a nearly complete set of electrogenic genes involved in fast neural signalling, plus a set of 665 G-protein coupled receptors that in the nervous system integrate with fast signalling machinery to modulate cellular excitability. Furthermore, Trichoplax expresses an array of genes involved in intracellular signaling, including the key effector enzymes protein kinases A and C that functionally link fast and slow cellular signaling. Also identified were nearly complete sets of pre- and post-synaptic scaffolding genes, most encoding appropriate protein domain architectures. Notably, the Trichoplax proteome was found to bear slightly reduced counts of synaptic protein interaction domains such as PDZ, SH3 and C2 compared to other animals, but abundance of these domains did not appear to predict the presence of synapses in early-diverging groups.

Conclusions Despite its apparent cellular and morphological simplicity, Trichoplax expresses a rich set of genes involved in complex animal traits. The transcriptome presented here adds a valuable additional resource for molecular studies on Trichoplax genes, exemplified by our ability to clone cDNAs for nine full-length acid sensing ion channel proteins with almost perfect matches with their corresponding transcriptome sequences.

Background

Electrical signaling in the nervous system is achieved by an array of well characterized electrogenic genes, including those that establish ion and charge gradients across the cell membrane (i.e. ion pumps and exchangers), and those that exploit these gradients to produce transient, fast-propagating electrical signals (i.e. ion channels and ionotropic receptors) [1]. It is also evident that fast information flow in the nervous system is dynamically regulated by slower modalities of cellular communication, referred to as neuromodulation, which can alter neural circuit properties to produce different outputs for changes in behavior [2]. A major source for neuromodulation are ligand-activated G-protein coupled receptors (GPCRs), which activate intracellular G proteins and signaling pathways to alter the functional properties of ion channels and other electrogenic genes through effector kinases such as protein kinases A and C (PKA and PKC respectively) [3]. Studies of the crustacean stomatogastric ganglion have revealed that even simple neural circuits are heavily neuromodulated [4]. For instance, a single neuromodulatory ligand can cause distinct changes in the electrical properties of neurons within a neural circuit, by differentially

altering the functional properties of specific ion channel types in each cell. Neuromodulators can also differentially alter synaptic strength, and hence, the effective wiring of neural circuits. Furthermore, combinations of different neuromodulatory ligands can work together to impose complex effects on neural circuits. Given the extensive library of known neuromodulators and GPCRs for both vertebrates and invertebrates, the potential for complexity is staggering. It is therefore striking that the molecular machinery for fast ionotropic and slow neuromodulatory cell signaling produces a finite set of coherent neural output patterns associated with specific behaviors and transitions in behavior [4]. Indeed, understanding how the nervous system evolved must include a consideration of how fast and slow forms of cellular communication became functionally intertwined [5].

Trichoplax adhaerens (*Trichoplax* sp. H1; phylum Placozoa) is a simple, early-diverging marine invertebrate that remarkably, lacks a nervous system and yet is able to coordinate its different cell types to conduct motile behavior including feeding [6], chemotaxis and geotaxis [7]. Locomotion by this small, flat, disc-shaped animal is achieved through beating cilia located on its ventral epithelium, which interestingly, undergo coordinated cessation of beating upon detection of food [6], and in response to applied regulatory peptides encoded within its genome [8–11]. Furthermore, observation of spontaneous co-ordinated pauses among adjacent *Trichoplax* animals not in contact with each other points to regulated secretion of signaling molecules that elicit pausing, and that these signals can travel from animal to animal [6]. Based on ultrastructure [12] and single cell transcriptome [13] studies, *Trichoplax* only has 6–11 cell types, which lack direct connections with each other in the form of electrical or chemical synapses. Hence, the extent and manner by which *Trichoplax* cells communicate with each other is unknown. Interestingly, protein-coding sequences predicted from the genome of *Trichoplax adhaerens* identified a large set of "neural" gene homologues, including ion channels and genes involved in synaptic transmission [10]. Indeed, given *Trichoplax*'s cellular simplicity and strategic phylogenetic position near the base of Metazoa, it might prove useful for understanding how networks of genes underlying fast and slow neural communication co-evolved in the nervous system. A first step is therefore determining the presence of relevant genes in the *Trichoplax* expressed gene set.

Previously, an mRNA transcriptome for *Trichoplax adhaerens* was generated in an effort to better annotate its genome, and to compare its genome to the related placozoan *Hoilungia hongkongensis* [14]. Here, we build on this previous work by generating a deep, high quality mRNA transcriptome for *Trichoplax adhaerens,* sequenced in quadruplicate from whole animal total RNA. Through a combined *de novo* and *ab initio* assembly strategy, we produced a high quality converged dataset with the vast majority of transcripts bearing complete protein-coding sequences. Based on gene ontology annotation and inferred average mRNA expression levels, we report an enrichment of *Trichoplax* secreted proteins at the whole animal level. Furthermore, prediction and annotation of the *Trichoplax* secretome identified several novel neuropeptide precursors, and homologues for proteins associated with digestion, development, immunity/defense, and cell adhesion. Also identified in the transcriptome was a near-complete set of genes involved in fast electrical neural signaling, and over 665 GPCRs whose intracellular signaling pathways can alter fast signaling machinery to impose lasting changes in cellular excitability. *Trichoplax* also expresses an array of genes that link fast and slow signaling machinery, including

adenylyl cyclase, phospholipase C, PKA and PKC. Furthermore, the transcriptome was found to contain a complete set of genes involved in intracellular Ca^{2+} signaling, including voltage-gated calcium channels, Ca^{2+} pumps and exchangers, intracellular calcium channels IP_3 and ryanodine receptors, calmodulin kinase II, calcineurin, and very high level expression of the Ca^{2+} sensor protein calmodulin.

Also notable, the vast majority of genes involved in synapse function were identified as expressed, with most bearing appropriate domain architectures required for their synaptic interactions and functions. Comparing the content of synaptic protein interaction domains present within the proteomes of various animals and eukaryotes revealed that *Trichoplax* has a slightly reduced number of domains compared to other animals, including PDZ, SH3, and C2. However, counts of these domains do not predict the presence of synapses, since sponges and choanoflagellates, which lack synapses, contain more than ctenophores, which have synapses. The transcriptome dataset presented here represents a valuable addition to the toolsets available for conducting molecular and phylogenetic studies on *Trichoplax,* exemplified by our cloning of nine cDNAs for acid-sensing ion channels using primers designed against their respective transcriptome sequences.

Results

A high quality whole animal mRNA transcriptome for Trichoplax adhaerens

Total RNA was extracted and purified from four separate batches of ~20 whole *Trichoplax adhaerens* animals (Grell Strain, generously provided by Dr. Leo Buss from Yale University), using the RNeasy Plus Universal Midi Kit (Qiagen, 3 samples) and the PicoPure RNA Isolation Kit (Thermo Fisher Scientific, one sample). Four corresponding cDNA libraries were then prepared using the TruSeq Standed Total RNA Library preparation kit (Illumina) as per the manufacturer's instructions, with the exception of excluding the mRNA fragmentation step to ensure longer gaps between paired-end reads. The four cDNA libraries were then sequenced on an Illumina HiSeq 2500 and characterized using FastQC [15] (v0.11.8), revealing sets containing 24,863,374, 24,934,020, 29,064,422, and 27,171,952 paired-end reads, for a total of 212,067,536 125 base pair reads, with an average GC content of 41 % (submitted to NCBI under SRA SUB5265969).

For transcriptome assembly (Figure 1A), paired-end reads were quality trimmed using Trimmomatic–0.36 [16], then processed through a parallel *de novo* and *ab initio* transcriptome assembly pipeline of 27 transcriptomes converged into a final comprehensive assembly with Evidential Gene [17]. Briefly, depending on requirements/capabilities of different transcriptome assembly programs, paired-end read sets were either normalized to remove PCR duplicates via FastUniq [18], or inputted directly into assembly programs. Genome-guided transcriptome assemblies (i.e. *ab initio)* were generated using Trinity [19] v2.2.0 and Cufflinks [20] v2.2.1, with paired-end read sets aligned to the published *Trichoplax adhaerens* genome using GSNAP [21] v2016–05–01 for Trinity, and TopHat [22] (v2.1.1) and Bowtie [23] (v2.2.9) for

Cufflinks. For *de novo* assemblies, Trinity and three additional programs were used using the default k-mer length of 25 base pairs for Trinity, and 63, 71, 79, 87, 95, 103, and 111 base pairs for the others: Trans-ABySS [24] (v1.5.3), Velvet/Oases [25] (v1.2.10 and v0.2.09 respectively), and SOAP-denovo-Trans [26] (v1.03).

The resulting 27 independent transcriptomes were processed through the EvidentialGene pipeline [17], which selects transcripts bearing the longest open reading frame for each assembled gene sequence. The resulting output was then processed to remove redundant transcripts using CD-Hit [27] (v4.6), with a protein sequence identity threshold of 95%, producing a non-redundant transcriptome assembly of 17,411 unique protein-coding sequences. Sequences were then manually processed to 1) separate a small number of independent open reading frames contained within a fused transcript; 2) remove contaminating sequences assembled from RNA-Seq reads coming from other species/samples that were sequenced alongside the *Trichoplax* cDNA on the Illumina sequencer; and 3) remove transcripts bearing coding sequences shorter than 100 amino acids, also lacking sequence homology with proteins in the following biological databases (assessed via BLASTp [28] v2.3.0 with an e-value cut-off of $1e^{-5}$): Swiss-Prot [29] (accessed January 2017), RefSeq [30] (accessed January 2017), and TrEMBL [29] (accessed February 2017). Overall, this process removed 5,430 genes, producing a final high quality mRNA transcriptome for *Trichoplax adhaerens* consisting of 11,981 unique protein-coding genes (Figure 1A).

A histogram of the assembled transcripts, arranged according to sequence length in base pairs, revealed a somewhat normal distribution, with only minimal enrichment of shorter fragmented sequences (Figure 1B). The high quality of the assembly is also evident in a histogram of unique protein sequences binned by length in amino acids (Figure 1C), where the vast majority of transcripts contain compete open reading frames with both start and stop codons, with very few bearing truncated protein-coding sequences at the 5’ end (N-terminus), 3’ end (C-terminus), or both (Figure 1C). To further assess the quality of the transcriptome, the 11,981 genes were aligned against protein sequences in the Swiss-Prot, RefSeq and TremBL databases using BLASTp (Figure 2A). As expected, the highly curated protein database Swiss-Prot produced few alignment hits with high bit-scores, reflecting a general absence of experimentally validated *Trichoplax* gene in the scientific literature. Accordingly, the BLAST results were largely unchanged when the transcriptome was re-aligned against a filtered Swiss-Prot database lacking *Trichoplax* genes. In contrast, alignment with RefSeq and TrEMBL produced generally higher bit-scores, and removal of *Trichoplax* genes from these databases considerably reduced the overall bit-score quality. A plot of bit-scores of each aligned transcriptome gene, sorted by descending bit-score (Figure 2B) shows that bit-scores arising from alignment with RefSeq and TrEMBL undergo marked shifts towards lower values when *Trichoplax* genes were filtered out, while the Swiss-Prot results remained unchanged. We note that a considerable number of *Trichoplax* transcriptome genes either failed to find BLAST homology in these databases or produced poor alignment scores (i.e. bit-scores below 50), likely representing uncharacterized Placozoa-specific genes. Altogether, these analyses indicate that our transcriptome is of very high quality, contributing a valuable set of complete open reading frames for both molecular and phylogenetic/bioinformatics studies.

The transcriptome expands the known complement of expressed Trichoplax genes

To directly compare our transcriptome gene set to genes from the previously published genome [10], coding sequences of our 11,981 genes were matched against their single best counterparts in the genome-predicted gene-set via BLASTn (filtering criteria: best hit per query where alignment length ≥100 bp and sequence identity ≥95%). A total of 9,498 transcriptome genes aligned with a converged set of 9,260 genome-predicted genes, indicative of either fragmented assembly of some transcriptome sequences (i.e. two broken gene fragments mapping to a single genome-predicted gene), and/or of the presence of chimerically fused gene sequences in the genome-predicted set (i.e. two separate transcriptome genes mapping to a chimeric genome-predicted gene; Figure 3A). Furthermore, 2,483 transcriptome genes did not find a match in the genome-predicted set, while 2,260 genome-predicted genes were left unmatched. However, reciprocal BLASTn using genome-predicted genes as query against transcriptome sequences, using the same filtering criteria as above, resulted in only 1,418 genome-predicted genes not finding a match, while 10,102 found matches converging onto 9,250 transcriptome genes, leaving 2,731 transcriptome genes unmatched. Again, this convergence can reflect either fragmented assembly of some of the genome predicted genes, or chimeric assembly of transcriptome genes. Nevertheless, we note that for all *de novo* assembly programs used in our assembly pipeline, outside of Trinity, longer k-mer lengths were selected according to thresholds identified in a previous study that characterized the effect of k-mer length on chimeric transcript assembly [31].

We next sought to evaluate how many of the 11,981 transcriptome gene sequences mapped to the published *Trichoplax* genome scaffolds [10], using the program GMAP [21] (v2016–05–01) with a stringent threshold sequence identity of ≥ 90 base pairs, and a trimmed coverage of ≥ 90%. Of the 2,483 transcriptome genes that did not find BLASTn homology in the genome-predicted gene set, 1,768 found a match in the genome scaffolds, representing novel genes likely present in the genome scaffolds but that previously failed to be predicted (Figure 3B). Instead, 715 transcriptome genes did not map using our cut-offs, representing completely novel genes either located within un-sequenced gaps in the genome scaffolds, or alternatively, genes present in the genome but not meeting our cut-off. Indeed, of the 9,498 transcriptome genes that previously aligned to the genome predicted-gene set via BLASTn, 329 failed to meet our GMAP cut-offs. As a crude confirmation that the 715 novel genes are indeed expressed transcripts, we note that their average TPM expression level 166 ±88 standard error (SE), compared to 48 ±26 SE for the 1,768 novel gens that mapped to the genome.

Importantly, separate mRNA transcriptomes for *Trichoplax* sp. H1 (i.e. *Trichoplax adhaerens)* and sp. H2 were recently released [14], in an effort to better annotate the sequenced genome. The work presented here complements and improves upon this previous study by having longer paired-end read lengths (75 vs. 125 base pairs), total read counts (150.7 vs. 212.1 million paired-end reads) and a comprehensive combined *de novo* and *ab initio* assembly strategy as detailed above. Nevertheless, we used BLASTp to determine whether the newly identified genes from our study were also identified in this previous study. Of

the 715 genes that did not map to the genome, 450 (~63%) found a match in the published *Trichoplax* sp. H1 transcriptome with an alignment length cut-off of 100 amino acids and a percent identity cut-off of 60% (Table S1). Increasing percent identity to ≥ 95% reduced this number to 335 (47%), and to 153 (21%) at 100% identity. Instead, of the 1,768 novel genes that did map to the genome, 1,312 (74%) found a match with ≥ 60% identity, 1,123 (64%) with ≥ 95% identity, and 460 (26%) with 100% identity. Altogether, the two transcriptomes corroborate the existence and expression of these novel genes, and our dataset likely adds several hundred additional genes to the known *Trichoplax* gene set. Using TBLASTN (filtering criteria: best hit per query where alignment length ≥100 bp), we also compared the recently published transcriptome of *Trichoplax* sp. H2 [14] to our *Trichoplax* sp. H1 transcriptome. At a percent identity cut-off of 60%, 281 (39%) of the unmapped genes were also identified in the *Trichoplax* sp. H2 transcriptome, while these values decreased to 205 (29%) and 69 (10%) genes, at ≥ 95% and ≥ 100% identity cut-offs, respectively (Table S1). Of the novel genes that had a match in the genome scaffolds but not the genome-predicted gene set, 900 (51%), 762 (43%), and 218 (12%) genes were found in the *Trichoplax* sp. H2 transcriptome at ≥ 60%, ≥ 95%, and ≥ 100% identity cut-offs, respectively (Table S1). We similarly compared the novel *Trichoplax* genes to the 12,575 genes from the related placozoan *Hoilungia hongkongensis [*32] using BLASTp. Here, 178 (~25%) of the 715 un-mapped genes found a match with ≥ 60% identity, and only 3 with ≥ 95% (Table S1). Instead, 452 (26%) of the novel 1,768 mapped genes found a match with ≥ 60% identity, and only 2 with ≥ 95% (Table S1). Lastly, we sought to determine whether some of the 2,260 genome-predicted genes not present in our transcriptome were previously discarded during processing and filtering. The 2,260 genome-predicted genes were collapsed down to a 1,570 non-redundant gene set using CD-Hit (nucleotide sequence identity ≥ 90%) and aligned against a concatenated database of all 27 transcriptomes used for the EvidentialGene pipeline using BLASTp (Figure 1A). Imposing alignment length and percent identity cut-offs of ≥ 33 amino acids and ≥ 95%, respectively, we recovered 615 previously discarded transcriptome sequences, leaving only 952 genome-predicted genes unmatched in the transcriptome. In total, our transcriptome expands the repertoire of genome-predicted genes that are expressed at the mRNA level to 13,548 (Figure 3C). The resulting transcriptome was deposited to GenBank under accession SUB5274527.

Gene ontology highlights enriched expression of genes involved in ER trafficking and secreted/extracellular functions

BLAST2GO [33] was used to characterize the gene ontology [34] (GO) of the *Trichoplax* mRNA transcriptome, inferring level 2 GO terms from a BLASTp protein alignment with Swiss-Prot (e-value cut-off of 1E−5). Interestingly, despite the apparent anatomical and cellular simplicity of *Trichoplax,* a considerable number of genes received Biological Process annotations associated with complex animal traits including development, reproduction, immune system, locomotion, signaling and multicellular organismal processes (Figure 4A). In conjunction with this analysis, we sought to determine the mRNA expression levels of transcriptome genes by mapping the four *Trichoplax* RNA-Seq datasets to the

assembled transcripts using the program RSEM [35], providing transcripts per million (TPM) expression levels for all gene sequences. This permitted us to repeat the GO annotation on subsets of the transcriptome partitioned according to top TPM expression levels (i.e. top 100, 200, 500 or 1,000 transcripts/genes). This revealed a relative overrepresentation of genes involved in defense/immune function, protein localization and cellular compartment biogenesis, and metabolic processes. Furthermore, we conducted GO enrichment analysis to compare terms between the top 1,000 most highly expressed genes and the whole transcriptome (using Fisher's Exact Test in BLAST2GO; two-tailed analysis, false discovery rate = 0.05). Visualization of the top 90 enriched GO terms for all three gene ontology categories, selected according to lowest p-value, revealed enriched mRNA expression of genes with GO terms associated with protein localization to the endoplasmic reticulum, and hence the endomembrane/secretory pathway (Figure S1). Consistent with this is the marked overrepresentation of GO terms for high TPM genes that are associated with the endomembrane, cell membrane and extracellular compartments of the cell, including exosomes, which are secreted vesicles with roles in cellular communication [36] (Figures 4B and S1). Notably, although small in number, it is interesting to find GO terms in the Cellular Component category associated with synapses and cell junctions (Figure 4B), and mRNA enrichment of genes involved in vesicle function (Figure S1), despite *Trichoplax* lacking synapses. Altogether, these data reveal that *Trichoplax* expresses a diverse set of genes associated with complex metazoan traits, including those involved in intra- and inter-cellular communication.

The Trichoplax secretome identifies genes involved in digestion, cell signaling, development and immune-related functions

Using three separate prediction methods, Phobius [37], Spoctopus [38], and the SignalP [39] (v4.1)/tmHMM [40] (v2.0c)/TargetP [41] (v1.1b) pipeline, we identified a set of 593 putative secreted proteins with predicted N-terminal signal peptides, no transmembrane helices or mitochondrial targeting motifs, and independently identified by at least two of the three algorithms (Figure 5A; Supplementary File 1). Annotation of these genes via BLAST homology with Swiss-Prot sequences identified sets of genes with putative functions in digestion (Table 1), immunity and defense (Table 2), development/reproduction (Table 3) and cell matrix/cell adhesion (Table 4). Notably, mRNAs for digestive enzymes are highly expressed at the whole-animal level, consistent with *Trichoplax*'s external digestion of food algae [6], presumably through secretion of enzymes from lipophil cells located on its ventral epthelium [12, 13]. Included among these are homologues for peptidases such as trypsin, kallikrein, cathepsin, lysozyme and chymotrypsinogen A, lipases such as phospholipase A1 and A2, and enzymes that hydrolyze sugars or amino acids such as alpha-amylase, chitobase and phospholipase B (Table 1). Hydrolytic enzymes are also important for immunity and defense. *Trichoplax* exhibits high expression levels of the protective enzymes granzyme K and lysozyme, and others involved in immune signaling such as cathepsin (Table 2). A putative orthologue for complement factor C3 was also identified (Table 2), however, domain prediction analysis via InterPro demonstrated that this protein more closely resembles the alpha–2-

macroglobulin domain-containing CD109 antigen, previously identified in *Trichoplax* sp. H2 [42]. Thus, our work supports the previous report that *Trichoplax* lack a *bona fide* complement system and yet harbor a rich set of immunity-related genes that likely figure in determining self from non-self [41]. Also highly expressed is the cytokine granulin, and other immunity-related genes involved in cellular stress responses and apoptotic signaling. In contrast, secretome genes involved in development/reproduction generally exhibited lower average TPM expression levels (Table 3). It is however notable that *Trichoplax* homologues for ovochymase–2 and vitellogenin–2, two genes involved in sexual reproduction [43, 44], exhibit considerably high TPM values, given the uncertainty about whether placozoans are capable of sexual reproduction [10, 45–47]. Also identified in the secretome were numerous genes with putative extracellular matrix and cell adhesion functions, including stereocilin, involved in the organization of mechanosensitive stereocilia in vertebrate sensory hair cells [48], and adventurous-gliding motility protein Z, involved in cellular gliding along external surfaces driven by intracellular molecular transport motors [49] (Table 4).

Expected within the secretome are pre-pro-proteins for regulatory peptides and pro-hormones that are secreted after enzymatic cleavage into mature peptides in the ER. Previously, several such peptide precursors were identified from *Trichoplax* genome-predicted genes bearing repeated convertase cleavage sites predicted with the program NeuroPred [8, 50]. Furthermore, several studies have shown that *Trichoplax* motile behavior can be dynamically modulated by exogenously applied peptides [9, 11], indicating that *Trichoplax* likely uses secreted peptides for cellular communication and coordination. Given the completeness of protein-coding ORFs in our transcriptome, we reasoned that a similar analysis might identify additional regulatory peptides precursor genes. Indeed, we able to identify all 9 of the previously predicted short regulatory peptide precursor sequences, plus 3 novel ones bearing canonical convertase cleavage sites (WWamide, ITKL, and YPFFGN), and 4 bearing non-canonical cleavage sites (PERI, FALF, ERSA peptides, and DAYQamide; Supplementary File 2; Figure 5B). We also confirmed expression of *Trichoplax* insulin 1 and 2 pro-hormones, the recently discovered MFPF peptide, and granulin, the latter showing very high mRNA expression with an average TPM value of 965 ±118 SE (Figure 5C). One of these novel peptides, YPFFGN (i.e. QDYPFFGN/S), has a striking resemblance to the vertebrate peptide endomorphin (YPFFamide), however, a C-terminal asparagine/serine residue makes it uncertain whether the glycine residue is amidated after cleavage. Nevertheless, ectopic application of an amidated version of this peptide, as well as vertebrate endomorphin, causes *Trichoplax* to pause ciliary locomotion as occurs during feeding [9]. Coincidentally, the YPFFGN gene is one of the 715 genes present in the transcriptome but completely absent from the genome (Figure 3B), but its existence has been confirmed by cloning and sequencing [9] (NCBI accession number KY675296).

Our analysis also identified several proteins bearing repeating stretches of arginine residues, which coincidentally, failed to find BLASTp homology in the Swiss-Prot and RefSeq protein databases and thus likely represent genes that are unique to *Trichoplax* (Figure 2A, B). This prompted us to look more carefully within the secretome, where we manually identified several additional arginine-enriched sequences (Supplementary File 2). Given the apparent non-canonical nature of these arginine stretches/repeats, it is uncertain whether they are cleaved by convertase, or instead, these novel proteins

remain complete after removal of the signal peptide. However, it is notable that some exhibit extremely high mRNA expression levels, with TPM values of 7,252 (evg1172538), 5,180 (evg1228357) and 3,837 (evg452961) for the three most highly expressed transcripts. In contrast, we note relatively lower expression levels for all regulatory peptides in general, with FALF and FFNPamide precursors showing moderate average TPM values of 387 ±15 SE and 325 ±10 SE (Figure 5C).

Trichoplax expresses a nearly complete set of genes involved in fast neural signaling

In vitro electrophysiological characterization of acid-sensing ion channel (ASIC) homologues cloned from gastropod molluscs revealed that select amidated regulatory peptides such as FMRFamide could activate them to produce depolarizing cation currents in cells [51–53]. Similarly, ASIC homologues cloned from the cnidarian *Hydra magnipapillata* activate in response to amidated regulatory peptides, where they might play a role in synaptic transmission across the neuromuscular junction [54]. Indeed, genomic expansion of ASIC channels and regulatory peptide genes in ctenophores and cnidarians has led to hypotheses that early-diverging animals employ secreted peptides and ASIC channel homologues for synaptic transmission [54, 55]. In the *Trichoplax* transcriptome, we were able to identify 10 full length ASIC channel coding sequences, also predicted from the genome, which we named *Tad*ASIC1 to *Tad*ASIC10. An additional ASIC sequence, *Tad*ASIC11, was predicted from the genome but was absent in our transcriptome. A maximum likelihood phylogenetic tree of the *Trichoplax* ASIC channels, inferred from a MUSCLE [56] protein alignment with homologues from other animals, revealed that most of the *Tad*ASICs are tightly clustered with poor bootstrap support linking them to other clades, while *Tad*ASIC10 forms a relatively strong clade with vertebrate epithelial sodium channels (ENaCs), mechanosensitive channels Mec–4 and Mec–10 from *C. elegans,* the uncharacterized *Hydra* ASIC subunit HyNaC12, and the FMRFamide peptide-gated channels from gastropod molluscs *Lymnaea stagnalis, Helix aspersa* and *Helisoma trivolvis [*51–53] (Figure 5D). Whole animal mRNA TPM expression levels of the *Tad*ASICs revealed relatively low expression for *Tad*ASIC2 to *Tad*ASIC8, very low expression for *Tad*ASIC1, and moderate expression for *Tad*ASIC9 and *Tad*ASIC10. As part of an ongoing study to functionally characterize the *Trichoplax* ASIC channels *in vitro,* we sought to clone their cDNAs into the mammalian expression vector pIRES2-EGFP from whole-animal total RNA (via RT-PCR using primers denoted in Table S2). *Tad*ASIC2 to *Tad*ASIC10 were successfully cloned in triplicate, producing consensus mRNA coding sequences for these genes (submitted to NCBI with respective accession numbers of MK547543, MK547544, MK547545, MK547546, MK547547, MK547548, MK547549, MK547550, and MK547551). Instead *Tad*ASIC1 and *Tad*ASIC11 failed to amplify by RT-PCR, likely due to nominal mRNA expression observed for *Tad*ASIC1 (Figure 5E), and possibly also for *Tad*ASIC11, which failed to show up in our transcriptome and hence is likely not expressed at the whole animal level (under our culturing conditions). Worth noting is that protein alignment of the validated ASIC channel cDNA sequences with their counterparts from the transcriptome vs. those predicted from the genomes of *Trichoplax adhaerens [*10]and the related placozoan *Trichoplax* sp. H2 [14] corroborates the high quality of our transcriptome,

where transcriptome sequences lack numerous deletions/insertions present in the other sequences (Supplementary File 3).

Using a database of different ion channel types as query, we identified numerous additional ion channel candidates in the *Trichoplax* transcriptome, which were subsequently carefully annotated using BLASTp protein alignment against the Swiss-Prot and NCBI non-redundant databases, combined with SMART-BLAST and InterPro [57] analyses, and in some cases, maximum-likelihood phylogenetic inference. Indeed, given the absence of bona fide neurons in *Trichoplax,* it is interesting that along with ASIC channels they express mRNAs for the vast majority of ion channels that are crucial for fast electrical signaling in the nervous system. For example, *Trichoplax* expresses a full complement of Pore-loop channels including fast voltage-gated sodium, calcium and potassium channels, a large conductance (BK) calcium-activated potassium channel, numerous AKDF (AMPA/Kainate-like) and epsilon subfamily ionotropic glutamate receptors, cation leak channels such as the sodium leak channel NALCN and a two-pore potassium channel, a cyclic nucleotide-gated channel, numerous inward rectifying potassium channels and numerous transient receptor potential (TRP) channels (Table S3).

Consistent with previous reports [55, 58], we failed to identify ion channels of the Cys-loop family which tend to play important roles in synaptic transmission, including ionotropic Glycine, GABA, acetylcholine, histidine and serotonin receptors, as well as innexins and connexins, which form gap junctions or hemi-channels [59]. Other identified channels included two ATP-gated P2X channels, one ryanodine receptor and three IP_3 receptors which operate in the ER, the mechanosensitive channel piezo, an orai calcium release-activated calcium channel and several CLC proton/chloride transporters that can act as voltage-gated chloride channels. Generally, the expression of these ionotropic genes at the whole animal mRNA level is low (Table S3). However, their presence in the genome is indicative of considerable potential for *Trichoplax* cells to undertake complex forms of electrical and calcium signaling. Unfortunately, to our knowledge, Placozoa is the only early-diverging phylum for which cellular electrical activity has not been successfully recorded using electrophysiology, and the role of fast electrical signaling in *Trichoplax* biology is completely unknown [60].

Trichoplax bears over 665 G protein-coupled receptor genes involved in slow neuromodulatory signalling

In the nervous system, the machinery underlying fast electrical signaling is dynamically regulated by neuromodulatory input, which operates along much slower and longer timescales and hence serves to alter neural circuit function to exact changes in behavior and physiology [4]. The most prominent form of neuromodulation occurs through ligand-activated G protein coupled receptors (GPCRs). These membrane receptors respond to various secreted neurotransmitters and neuropeptides by activating intracellular G proteins (i.e. Gα and Gβγ), which subsequently activate a series of intracellular signaling molecules including kinases that can phosphorylate ion channel proteins to alter their function (e.g. such as protein kinase C and protein kinase A) [3]. In some cases, activated Gβγ subunits physically bind ion channels to

alter their activity, as occurs during feedback auto-inhibition of pre-synaptic Ca_v2 calcium channels. In this context, ligand activation of peri-synaptic GPCRs by secreted neurotransmitters activates Gβγ proteins, which in turn bind to and inhibit pre-synaptic Ca_v2 channels to dampening their activity and hence neurotransmitter exocytosis [61]. Indeed, how these two forms of cell-cell signaling, fast ionotropic and slow neuromodulatory, became integrated during nervous system evolution to give rise to coherent neural output patterns underlying transitions in behavior is not known. *Trichoplax,* in its early-divergence and cellular simplicity, perhaps poses unique opportunities for exploring this question.

Here, we sought to identify *Trichoplax* GPCRs from the complete gene set using combined prediction with GPCRHMM [62] and InterProScan [63] (version 5.23–62.0). The two methods yielded 769 and 929 transcripts, respectively. The convergent dataset of 713 putative GPCRs was reduced down to 665 genes/proteins after removing fragmented gene sequences and ones that failed to align with GPCRs in the NCBI non-redundant database via BLASTp (Figure 6A; Supplementary File 4). A maximum likelihood phylogenetic tree constructed from a MUSCLE protein alignment trimmed with trimAl [64] (v1.2rev59), using IQTREE [65] (v1.6.5), revealed several clades of Rhodopsin-like (class A) GPCRs together comprising 533 genes or ~80% of the *Trichoplax* GPCRome (Figure 6B). By contrast, only 64 glutamate (class C) and 63 adhesion/secretin-like (class B) GPCRs could be identified, and even fewer frizzled, cAMP and ocular albinism-like GPCRs with 2, 2 and 1 identified GPCRs, respectively. Previously, Kamm et al. (2018) reported that the Rhodopsin-like class of GPCR represents the single most abundant gene family/domain in *Trichoplax* sp. H2, with over 700 genes identified using InterPro [14]. Our work here evidences a similar abundance in *Trichoplax adhaerens,* and delineates a robust GPCRome, validated not only with InterPro but also with a hidden Markov model whose sensitivity for detecting *bona fide* GPCRs is approximately 15% higher than that of transmembrane predictors [62]. Of the 655 *Trichoplax* GPCRs for which TPM mRNA expression data was available (i.e. not from the genome-predicted gene set), TPM values ranged from 0.07 to 125.9 with a mean and standard deviation of 4.6 and 9.4, respectively.

Trichoplax expresses genes for glutamatergic and GABAergic biosynthesis and degradation

Interestingly, the soft-bodied sponge *Tethya wilhelma* contracts upon application of the neurotransmitters glutamate and gamma-aminobutyric acid (GABA), presumably through activation of metabotropic GPCRs [66]. Given the basal position of sponges within the Metazoa, it thus seems likely that these two ligand systems were adopted very early on during animal evolution for cellular communication, if not before animals emerged. A quick survey of *Trichoplax* GPCRs identified numerous putative glutamate and GABA metabotropic GPCRs in the transcriptome, as well as other transmitter-activated GPCRs such norepinephrine and opioid receptors (data not shown). We thus sought to update the annotation of *Trichoplax* genes for neurotransmitter biosynthesis and degradation, using BLASTp alignment against Swiss-Prot and NCBI non-redundant databases, combined with SMART-BLAST and InterPro analysis, and when required, phylogenetic analysis [10, 58]. Not surprisingly, pathway components for glutamate and

GABA appear to be complete in *Trichoplax* (Table S4). However, most other transmitter pathways were found to be incomplete, including those for the synthesis and/or degradation of catecholamines dopamine, epinephrine and norepinephrine, as well as those for serotonin, octopamine and histamine (Table S4). Interestingly, we did find two *Trichoplax* homologues for choline acetyltransferase, previously reported as absent [58], but not the acetylcholine degradation enzyme acetylcholinesterase. Finally, we identified 2 *Trichoplax* enzyme homologues for synthesis of the transmembrane diffusible ligand, nitric oxide, suggesting that one of the three inducible nitric oxide synthase (iNOS)-like genes previously reported in *Trichoplax adhaerens* [67] is not expressed, or, accordant with its namesake, is expressed only under select conditions yet to be identified.

Trichoplax points to an early establishment of the synaptic proteome and synaptic signaling machinery

With our expanded *Trichoplax* gene set, we carried out an extensive annotation of synaptic gene homologues, using the same thorough annotation strategy as described for ion channels and neurotransmitter biosynthesis pathways above. *Trichoplax* was found to express all members of the exocytotic SNARE complex (i.e. synaptobrevin, SNAP25, and syntaxin–1), and Ca^{2+}-sensitive elements of the exocytotic machinery synaptotagmin and complexin (Figure 7A, B; Table S5). In addition to the pre-synaptic Ca_V2 calcium channel noted previously, *Trichoplax* also expresses an array of pre-synaptic scaffolding proteins that interact with Ca_V2 calcium channels to regulate their pre-synaptic localization and function, including RIM, RIM-BP, Mint, CASK and CAST/ELK [68–72] (Figure 7A, B; Table S6). Interestingly, *Trichoplax* possess two RIM (Rab3 Interacting Molecule) homologues, both bearing canonical Zn^{2+}-finger motifs that mediate interactions with Munc–13, and two C2 domains for interactions with α-liprin and ELKs [70]. However, one homologue lacks a PDZ domain that in fruit fly and mouse is thought to bind the distal C-terminus of Ca_V2 channels promoting their localization at the synapse active zone [73, 74]. RIM is particularly interesting being recently marked as one of only 25 genes that are both unique to animals and have resisted genetic loss across all major metazoan clades [75]. Some notable genes that we were unable to identify in the transcriptome were the related active zone scaffolding proteins Bassoon and Piccolo, and the invertebrate ELK-related protein Bruchpilot [76, 77].

Overall, it is striking that the vast majority of *Trichoplax* homologues for the pre-synaptic assembly bear appropriate domain architectures required for specific interactions and functions at the synapse. We note that the expression of these genes is quite variable at the whole-animal level (Figure 7B; Tables S5, S6), and a conclusive determination about whether these genes are functionally co-expressed in *Trichoplax* cells will require deep single cell transcriptome profiling. Recently, a single cell transcriptome study on *Trichoplax,* identifying about 100 co-expressed genes per cell type, revealed that synaptobrevin and synaptotagmin are co-expressed in cells also expressing putative regulatory peptides [13]. Perhaps with deeper sequencing, it will become clearer whether cellular excitation and Ca^{2+}-regulated exocytosis is taking place in select *Trichoplax* cell types. Of course, this would require co-expression with an array of

ion channels that regulate membrane excitability and exocytosis. Interestingly, sponges appear to have lost core genes required for neural excitability including voltage-gated Na_v and K_v channels [60, 78–80]. In the Great Barrier Reef sponge *Amphimedon queenslandica,* which lacks synapses, the absence of temporal co-expression of synaptic genes suggests that synapse and neural evolution involved emergent co-expression of pre-existing genes/proteins in the primordial neuron [81], or alternatively, the loss of a nervous system in sponges [82] involved genetic losses and desynchronization of neuronal gene expression networks.

Analysis of the *Trichoplax* post-synaptic proteome similarly revealed that most relevant genes are present and expressed at the whole-animal level, and bear appropriate domain architectures required for their synaptic function (Table S6). For example, present in the transcriptome are disc-large homologues DLG1/SAP97 and DLG5, members of the membrane associated guanylate kinases (MAGUK) protein family that are core organizers of the post-synaptic density [83], and key proteins required for synapse formation and stabilization including neurexin and neuroligin that span the synaptic cleft [84]. Interestingly, neuroligins appear to be absent in sponges, ctenophores and choanoflagellates [55, 80, 85], suggesting that *Trichoplax* and the Placozoa are the most early-diverging lineage to possess this particular gene. Also present is Gephryn, a key organizer of inhibitory synapses that clusters Cys-loop GABA and glycine receptors at the post-synaptic membrane [86]. Notably, we failed to identify GRIP and Shank, both involved in post-synaptic localization of ionotropic glutamate receptors (i.e. AMPA and NMDA) as well as GPCRs [83, 87]. Shank also interacts with various other post-synaptic proteins that were found in the transcriptome, including Homer and Cortactin, and a post-synaptic Ca_v1 channel homologue that in neurons (i.e. $Ca_v1.3$) controls Ca^{2+}-dependent gene expression changes associated with synaptic plasticity and learning and memory [88]. Interestingly, like RIM and Ca_v2, Shank interacts with the distal C-terminus of the $Ca_v1.3$ channel via a PDZ protein interaction domain, however whether this interaction is conserved in invertebrates is not known. Also identified were homologues for a rich ensemble of genes involved in synaptic intracellular signaling, including the kinase PKA, challenging a previous suggestion that the *PRKACA* (Protein Kinase A Catalytic Subunit) gene is only found in the genome of *Trichoplax* sp. H2, and not *Trichoplax adhaerens* [14]. Also identified were PKC, calmodulin kinase II (CaMKII), the PKA anchoring protein AKAP, and the phosphatases PP2A, PP2B and calcineurin. Both CaMKII and calcineurin are regulated by changes in intracellular [Ca^{2+}] via calmodulin, which is very highly expressed in *Trichoplax* with an average TPM of 2,091.6. Clearly, elaborate intracellular Ca^{2+} signaling is likely for *Trichoplax,* with detectable mRNA expression of ER Ca^{2+} channels ryanodine and IP_3 receptors, plasma membrane and sarcoplasmic reticulum Ca^{2+} ATPase pumps, several $Na^+/(K^+)/Ca^{2+}$ exchangers, and GPCR effector enzymes adenylate cyclase and phospholipase C. Furthermore, homologues for the cell growth and proliferation signaling pathways mTOR [89] (e.g. mTOR, raptor and rictor) and MAPK/ERK [90] (e.g. extracellular signal-regulated kinases/ERKs, mitogen-activated protein kinases/MAPKs, Ras, Raf, cAMP response element binding protein/CREB) were also found to be expressed at the mRNA level (Table S6).

Although not exclusive to the synapse, several types of protein-protein interaction domains are essential for the assembly of both pre- and post-synaptic protein complexes, such as PDZ, Src Homology 3 (SH3), C2, FYVE zinc-finger (Zn-ginger), guanylate kinase (GK), Rho-GTPase-activating protein (Rho-GAP), Unc13 homology and CaMKII domains (e.g. Figure 7A). Previously, comparison of the PDZ domain content and diversity in proteomes of several metazoans and related single-celled eukaryotes revealed both expansion and specialization of PDZ domains in animals associated with increased molecular complexity at the synapse [91]. To extend that analysis, we compared the total numbers of different synaptic domains within the *Trichoplax* proteome to those in other animals. Domain prediction was carried out on the *Trichoplax* predicted proteome, plus 23 additional proteomes, using InterProScan. PDZ, SH3 and C2 domains were predicted using the Superfamily application (expect value ≤1E−6), while Zn-finger, GK, Rho-GAP Unc13 homology and CaMKII domains were identified with Pfam (expect value ≤1E−6), and counted to produce the plot shown in Figure 8. Not surprisingly, vertebrates possess expanded counts for all domain types consistent with an enriched genetic repertoire brought about by several rounds of genome duplication [92]. Interestingly, chelicerate horseshoe crabs are also thought to have undergone whole genome duplication relative to other arthropods, and we note a corresponding enrichment in domains for *Limulus polyphemus* compared to arthropods *Drosophila melanogaster* and *Apis mellifera.* Compared to other animals, the proteomes of *Trichoplax* and the two ctenophores *Mnemiopsis leidyi* and *Hormiphora californiensis* contain fewer predicted PDZ, SH3 and C2 domains, although *Trichoplax* has considerably fewer genes in total. Indeed, it is interesting that sponges *Amphimedon queenslandica* and *Oscarella carmela* possess more PDZ and SH3 domains than ctenophores and *Trichoplax,* as do the single-celled choanoflagellates *Monosiga brevicollis* and *Salpingoeca rosetta.* Another animal with a notable enrichment in domains is for the bivalve mollusc *Crassostrea gigas,* which has considerably more PDZ, SH3 and C2 domains than the *Octopus bimaculoides* or *Aplysia californica.* Researchers have previously reported massive selective duplication of innate immunity-related genes [93] and protein kinase genes [94] in *Crassostrea,* however genome duplications have not been reported to the best of our knowledge. What is evident from this analysis is that counts of “synaptic” protein interaction domains does not predict the presence of synapses since choanoflagellates, sponges and *Trichoplax,* which lack synapses, possess domain counts that are comparable to those of animals with synapses. Indeed, the sponge proteomes contain considerably more of these domains than ctenophores, which have synaptically connected neurons as well as neuromuscular synapses [95].

Discussion

Here, we provide the first detailed annotation of the *Trichoplax adhaerens* whole animal mRNA transcriptome. We note that although a transcriptome shotgun assembly was recently released for *Trichoplax* sp. H1 [14], the focus of that study was on placozoan genomics, and not transcriptome annotation. Our deep sequencing and comprehensive assembly strategy produced a transcriptome data set that significantly expands upon the former, by identifying several hundred new genes, and with the majority of assembled transcripts bearing complete open reading frames which is useful for functional studies that require gene cloning. This was exemplified by our ability to successfully clone 9 *Trichoplax*

acid-sensitive ion channel homologues using primers designed against their corresponding transcriptome sequences, without the need for classic and laborious gene cloning methods such as degenerate PCR and 5'/3' RACE (rapid amplification of cDNA ends). In separate ongoing studies, we have similarly used the transcriptome to clone other ion channel types, some of which bear protein-coding sequences longer than 5,000 base pairs. Clearly, the gene sequences that were generated provide a valuable resource for functional studies of *Trichoplax* genes.

Sequencing the *Trichoplax* whole animal mRNA transcriptome from four separate cDNA libraries permitted quantification of average gene expression levels (i.e. transcripts per million or TPM). Among the most highly expressed genes were those that encode putative secreted proteins, including hydrolytic enzymes such as proteases, lipases and amylases. Interestingly, a single cell transcriptomic study identified a specific *Trichoplax* cell type that co-expresses numerous hydrolytic enzymes, including those identified in this study (e.g. phospholipase A2, trypsin) [13]. However, whether these cells correspond with lipophil cells, proposed digestive cells located along the ventral epithelium that bear large acidophilic/lipophilic vesicles, is not clear [6, 12]. The ability of *Trichoplax* to quickly break down living unicellular microorganisms along its ventral epithelium for feeding has been documented, and it is likely that regulated secretion of hydrolytic enzymes occurs in response to localized sensory feedback indicating the presence of food, since secretion/digestion seems to occur at discrete sites along the ventral epithelium, not throughout [6]. Perhaps instead, some of the identified enzymes are contained within lysosomal compartments in fiber cells, which are proposed to endocytose entire microalgae as a secondary form of food uptake [96].

Also identified within the *Trichoplax* secretome were genes important for multicellular processes such as development, reproduction and immunity. Of course, the mere presence of these genes is not enough to expect that their functions are necessarily similar, where at least a subset are likely to have pleiotropic and/or divergent functions. For example, the egg yolk protein vitellogenin, which was identified in the *Trichoplax* secretome (Table 3), has taken up non-reproductive functions in select arthropods, such as control of longevity [97] and social interactions [98]. Also, the protein Homer (Table S6), a key post-synaptic scaffolding protein, was found to localize to the nucleus in the choanoflagellate *Salpingoeca rosetta,* and to interact with the lipid raft protein flotillin [99], suggestive of roles in nuclear scaffolding and signaling. In this case, these observations do not represent functional divergence, but rather, an ancient pleiotropic function since both the interaction between Homer and flottilin, and the nuclear localization of both, were subsequently identified in vertebrates astrocytes [99].

The *Trichoplax* secretome was also found to contain several novel short regulatory peptide gene precursors, which is interesting because regulatory peptides appear to be important for *Trichoplax* motile behavior. For example, ectopic application of the YPFFGN endomorphin-like peptide identified in this study causes *Trichoplax* to suddenly pause ciliary locomotion [9], as occurs during feeding [6]. Also, the peptides FFNP, ELPE, MFPF and WPFF trigger animal flattening and internal movements; PWN and SIFGamide trigger detachment and folding, and LF and LFNE trigger rotation and flattening [11]. To our knowledge, the 7 novel peptides identified in this study have not yet been tested for their effects on

Trichoplax motile behavior, including FALF and WWamide that exhibit fairly high mRNA expression at the whole animal level (Figure 5C). Also interesting are the arginine rich secretome proteins (Supplementary File 2), in that they appear to be unique to *Trichoplax* and most exhibit very high expression levels. The function of these genes is completely unknown, but it is interesting to note that *Trichoplax* fiber cells possess rickettsial endosymbionts within the endoplasmic reticulum [12, 100, 101], a rare compartment for intracellular bacteria in animals, where perhaps, a subset of secretome proteins remain the ER to mediate symbiotic interactions. Alternatively, arginine-rich proteins are endowed with membrane permeating capabilities, where perhaps, these genes are involved in cellular signaling, or defense/immunity against pathogenic microorganisms [102].

It is striking that despite lacking a *bona fide* nervous system, *Trichoplax* expresses a nearly full complement ion channels, pumps and exchangers that permit fast electrical signaling in the nervous system. Sponges, which similarly lack nervous systems, are different from placozoans in that they appear to have lost key electrogenic genes including voltage-gated sodium and potassium channels [78]. Nevertheless, glass sponges exhibit electrical signals that propagate slowly between cells via cytoplasmic junctions, signaling arrest of choanocyte ciliary beating and the feeding current [103]. Furthermore, sponge contractile cells require transient influx of divalent cations such as Ca^{2+}, Mg^{2+} and/or Sr^{2+} [103–105], and for some but not all species, contract in response to transient membrane depolarization caused by increasing extracellular [K^{+}] [104]. Not known is whether *Trichoplax* cells are also capable of electrical signaling in the form of graded or action potentials, or excitation-induced intracellular calcium signaling, such as exocytosis and contraction, since neither electrophysiological recording nor voltage/calcium imaging have been reported for any placozoan [60]. In a recent study, *Trichoplax* dorsal epithelial cells were shown to undergo ultrafast contractions, which interestingly, spread from cell to cell in waves suggestive of mechanical coupling. Contractions were proposed to be mediated by non-muscle actinomyosin filaments. Interestingly, application of ionomycin, which increases cytoplasmic [Ca^{2+}] via release from internal stores [106], caused all dorsal epithelial cells to contract, suggesting the contractile filaments are regulated by Ca^{2+}. In this study, the authors did not test whether the contractions were also dependent on external Ca^{2+}, which could flow into the cytoplasm through select ion channels. Should this be the case, the spread of contraction between cells might conceivably occur through the following mechanism: mechanical stress from adjacent cells activates stretch-sensitive ion channels at the cell membrane, which in turn either depolarize the cell to recruit voltage-gated calcium channels, or directly conduct Ca^{2+} into the cell. The presence of ryanodine receptors in the transcriptome suggests that cells are capable of calcium-induced calcium release, where transient increases in cytoplasmic Ca^{2+} activate ryanodine receptors in the ER, which in turn conduct ER Ca^{2+} to further increase its cytoplasmic concentration. Interestingly, the soluble cytoplasmic Ca^{2+} sensor protein calmodulin was found to be highly expressed at the whole animal level. Localization of this protein at the cellular level might highlight cell types that rely on transient Ca^{2+} signalling. Another cell type that might be contractile is the *Trichoplax* fiber cell, located between the dorsal and ventral epithelium and proposed to mediate localized churning movements observed during feeding [6]. Although direct evidence that fiber

cells contract has not been reported, they do exhibit muscle-like filaments under electron microscopy [107]. Clearly, functional studies of distinct cell types, coupled with electrophysiology and/or optical recording, are a necessary next step in understanding the roles of electrical and calcium signaling in *Trichoplax* biology.

At a genetic level, *Trichoplax* cells seem poised for complex electrical and calcium signaling, expressing numerous ion channel genes including: voltage-gated potassium and sodium channels that drive fast action potentials; ion channels that modulate action potential shape and frequency such as Ca^{2+}-activated potassium channels and cyclic-nucleotide gated channels; and channels that are activated by extracellular ligands such as ionotropic glutamate receptors, ASIC channels, P2X receptors and TRP channels (Table S3). Notably, *Trichoplax* is the most early-diverging animal to possess all three types of voltage-gated calcium channels found in animals, including a Ca_V1 channel homologue that in muscles drive excitation-contraction coupling, and a Ca_V2 channel homologue that in neurons drive excitation-secretion coupling (i.e. neurotransmitter exocytosis). Previously, we reported the electrophysiological properties of the cloned *Trichoplax* Ca_V3 calcium channel, expressed in a human cell line. In neurons and muscle, Ca_V3 channels help control action potential threshold, and play a major role in regulating cellular excitability. We found that despite its significant divergence at the amino acid sequence level, the *Trichoplax* Ca_V3 channel conducts low-voltage activated Ca^{2+} currents with hallmark features for this channel type that are required for regulating excitability [108]. By extension, it is plausible that other and perhaps numerous *Trichoplax* ion channels share core functional attributes with their counterparts from other animals, which could be co-expressed in varying functional assemblages as occurs in neurons to produce alternate modalities of cellular excitation.

Bioinformatic prediction of 665 *Trichoplax* GPCRs from our expanded gene dataset significantly expands the repertoire previously identified from the genome [109, 110]. Altogether, GPCRs encompass roughly ~5% of all *Trichoplax* protein-coding nuclear genes, which is remarkable given the apparent cellular and morphological simplicity of the animal. In contrast poriferans, which have more anatomically distinct cell types [111], are estimated to only have about 220–328 GPCRs [110, 112], while *Trichoplax* more resembles cnidarians and ctenophores in its GPCR content [110]. In general, GPCRs are for integrating an array of extracellular signals, both intrinsic and sensory, into intracellular responses through various intracellular signaling programs. In a previous study, *Trichoplax* was shown to possess all of the core intracellular regulators of GPCRs [113], including various classes of G proteins, G protein-coupled receptor kinases (GRKs), arrestins, phosducins, Ric8, GoLoco and regulators of G protein signaling (RGS) [109]. Here, we identified key downstream GPCR signalling elements that in the nervous system form a critical link between neuromodulatory GPCR signaling and electrogenic/synaptic signaling [114]. For example, *Trichoplax* expresses homologues for adenylyl cyclase (increases cytoplasmic cAMP levels when bound to activated G proteins), and the cAMP-dependent protein kinase A (PKA). Activated PKA has many electrogenic downstream targets such as Ca_V1 channels, causing long-lasting changes in their biophysical properties and hence the state of cellular excitability [115, 116]. Similarly, G-protein activation of phospholipase C leads to hydrolysis of membrane-bound phosphatidylinositol 4,5-bisphosphate

(PIP_2), producing diacylglycerol (DAG) and inositol triphosphate (IP_3), the former of which activates PKC and the latter ER-localized IP_3 receptors [117, 118]. Like PKA, PKC has various targets including ion channels and synaptic proteins [117, 119], and hence similarly serves to alter neural excitability and synaptic signaling for extended periods. Additionally, like ryanodine receptors, IP_3 receptors conduct ER Ca^{2+} into the cytoplasm upon activation, providing a functional link between extracellular GPCR ligands and intracellular Ca^{2+} signaling [120]. And lastly, activated G proteins themselves can directly bind and modulate ion channels, including G protein activated inward rectifying K^+ (GIRK) channels, pre-synaptic Ca_V2 calcium channels, and $Ca_V3.2$ calcium channels [121]. In the vertebrate nervous system, the complexity of integrated fast and slow neural signaling is expected to be staggering, given the diversity of cell types and complexity of tissues, and the broad lexicon of neuromodulatory ligands and GPCRs [4]. Indeed, the cellular simplicity of *Trichoplax* provides a tantalizing prospect for the comprehensive determination of how fast and slow neural signaling genes co-segregate in their cellular expression across an entire animal.

Trichoplax expresses a rich complement of genes involved in synapse formation and function, most bearing appropriate domain architectures that are required for their integration into specific proteomic complexes at pre- and post-synaptic sites (e.g. Figure 7). Many studies have now shown that early-diverging animals, and single-celled choanoflagellates, possess a large complement of synaptic gene homologues [122]. However, largely unexplored is whether these proteins assemble into homologous complexes as those found in synapses. A major next step forward is therefore a determination of the complexing of these proteins, using classic methods such as yeast II hybrid, co-immunoprecipitation and mass spectrometry. We sought to evaluate whether the absence of synapses in *Trichoplax* and sponges is reflected by a reduced number of synapse-associated protein-protein interaction domains (Figure 8). Sponges and choanoflagellates, which lack synapses, were found to generally possess more domains than *Trichoplax* and ctenophores, indicating that domain counts do reflect the presence/absence of synapses. What is clear from our analyses is that some lineages have considerably expanded domain counts, which for vertebrates and select arthropods, likely arose through independent genome duplication events. We note that our analyses did not explore the heterogeneity of the selected domains. For example, PDZ domains separate into distinct classes that bind different C-terminal sequence motifs on target proteins [123]. Previously, poriferans and choanoflagellates were found to have a reduced diversity in PDZ domain types [91], which might also be true for *Trichoplax* and ctenophores should they be analyzed in the same way. Interestingly, a recent study determined that a large proportion of metazoan proteins which interact with PDZ domains predate animals, and that their incorporation into PDZ-dependent protein complexes emerged later via *de novo* mutations that produced appropriate PDZ ligand motifs [124]. An interesting question is therefore: what are the ligands for different synaptic scaffolding proteins in early-diverging animals such as *Trichoplax,* ctenophores and poriferans, and how do their interactomes compare to those humans and other animals?

Methods

Due to the *in silico* nature of this study, the majority of methods used to generate the data are contained within the results section and corresponding figure legends. Transcriptome annotation for, among other research objectives, the purpose of inferring orthology between *Trichoplax adhaerens* genes and those found in other species, consisted of a combination of BLAST against UniProt/Swiss-Prot databases, reciprocal BLAST against NCBI, and InterPro protein domain analyses. In cases where orthology was uncertain, maximum likelihood phylogenetic inference was also used. Secretome transcripts were not subject to reciprocal BLAST analysis, however, BitScores and E-Values from our BLASTp annotation are provided in Tables 1–4, allowing the reader to critically assess whether grounds for inferring homology exist. Below, we only detail background methods used for obtaining and generating the proteomes used for Figure 8, where we quantified the content of synaptic protein interaction domains in the proteomes of various species. Briefly, full proteomes for *Saccharomyces cerevisiae, Exaiptasia pallida, Acropora digitifera, Crassostrea gigas, Aplysia californica, Octopus bimaculoides, Helobdella robusta, Caenorhabditis elegans, Limulus polyphemus, Apis mellifera, Drosophila melanogaster, Strongylocentrotus purpuratus, Danio rerio, Mus musculus, Rattus norvegicus,* and *Homo sapiens* were downloaded from the Ref-Seq protein database (accessed February 2018). For the sponges *Amphimedon queenslandica* and *Oscarella carmela,* transcriptome-derived protein models were obtained from the *Amphimedon queenslandica* Transcriptome Resource [125] and compagne.org [126], respectively. Genome-derived protein models for the choanoflagellates *Monosiga brevicollis* and *Salpingoeca rosetta* were obtained from JGI Genome Portal (Project ID: 16178) [127] and Ensemble Genome Project accession: PRJNA37927, respectively. For the ctenophore *Hormiphora californiensis,* paired-end RNA-Seq reads derived from mRNA were obtained from the NCBI Sequence Read Archive (accession SRR1992642), and two *de novo* transcriptomes were generated using Trinity [16], one with *in silico* read normalization and one without. These were then converged into a single assembly using EvidentialGene. For *Mnemiopsis leidyi,* paired-end mRNA RNA-Seq reads were obtained from the NCBI sequence read archive (accession numbers SRR1971491, SRR4353882, SRR4353883, SRR4353884, SRR4353885, SRR4353886, SRR4353887, SRR4353888, SRR4353889, SRR4353890, SRR4353891, SRR4353892, SRR4353893, SRR4353894, SRR4374091, SRR4374265, SRR4374273, SRR4374274, SRR4374324, SRR4374325, SRR4374356, SRR4374357, SRR4374583, SRR4374709, SRR4374710, SRR4374711, SRR4374712, SRR4374713, SRR4374714, SRR4374715, SRR4374742 and SRR4374769). All reads were pooled into four sets of paired-end reads for normalization using FastUniq [18]. A condensed EvidentialGene pipeline, similar to the one used to generate the *Trichoplax* mRNA transcriptome, was used to generate a transcriptome for *Mnemiopsis,* but only using a single *ab initio* assembly (Cufflinks) and a single *de novo* assembly (Trinity). Predicted protein sequences with lengths shorter than 100 amino acids that did not find BLASTp homology in Swiss-Prot (accessed: April 2017), Ref-Seq (accessed: January 2017), or TrEMBL (accessed: February 2017), were removed (e-value > 1e−5).

Similarly, for *Nematostella vectensis,* single-end RNA-Seq reads were obtained from NCBI (accession numbers SRR5183917, SRR5183918, SRR5183919, SRR5183920, SRR5183921, SRR5183922, SRR5183923, SRR5183924, SRR5183925, SRR5183926, SRR5183927, SRR5183928, SRR5183929 and SRR5183930). These were assembled using a combined strategy as depicted in Figure 1A for the

Trichoplax transcriptome, using trimmed RNA-Seq reads that were either normalized or un-normalized for *de novo* assembly with Trinity (i.e. 2 assemblies), or normalized for *ab initio* assembly with Cufflinks (1 assembly). A fourth assembly was obtained from FigShare [128], and the four assemblies were merged using the EvidentialGene pipeline. Prior to domain analysis, all of the obtained proteomes were processed for removal of redundant sequences at the protein level using CD-Hit (sequence identity threshold of 95%).

Abbreviations

AKDF: AMPA Kainate Delta Phi; AMPA: α-amino–3-hydroxy–5-methyl–4-isoxazolepropionic acid; ASIC: acid-sensing ion channel; ATP: adenosine triphosphate; BK channel: big potassium; BLAST: basic local alignment search tool; CaMKII: calmodulin kinase II, cAMP: cyclic adenosine monophosphate; CASK: calcium/calmodulin-dependent serine protein kinase; CAST: cytomatrix at the active zone-associated structural protein; cDNA: complementary deoxyribonucleic acid; CLC: chloride-conducting ion channel; CREB: cAMP response element; DAG: diacylglycerol; EGFP: enhanced green fluorescent protein, ENaC: epithelial sodium channel; ER: endoplasmic reticulum; ERK: extracellular signal–regulated kinase; FYVE domain: Fab 1, YOTB, Vac 1, and EEA1 domain; GABA: gamma-aminobutyric acid; GIRK: G protein-activated inward rectifying K^+ channel; GK: guanylate kinase; GO: gene ontology; GPCR: g-protein coupled receptor; GRK: GPCR kinase; IP_3: inositol 1,4,5-triphosphate; MAGUK: membrane-associated guanylate kinase; MAPK: mitogen-activated protein kinase; Mint: munc–18-interacting protein; mRNA: messenger ribonucleic acid; mTOR: mammalian target of rapamycin; NALCN: sodium leak channel; NCBI: national center for biotechnology information; NMDA: N-methyl-D-aspartate; ORF: open reading frame; P2X: purinergic type 2 receptor X; PDZ domain: post synaptic density protein 95, Drosophila disc large tumor suppressor 1, and zonula occludens–1 protein domain; PKA: protein kinase A; PKC: protein kinase C; RGS: regulator of G protein signaling; Rho-GAP: RACE: rapid amplification of cDNA ends; Rho-GTPase-activating protein; RIM: rab3-interacting molecule; RIM-BP: RIM binding protein; RT-PCR: reverse transcription polymerase chain reaction; SH3 domain: SRC homology 3 domain; SNAP–25: synaptosomal nerve-associated protein 25; SNARE: soluble NSF attachment protein receptor; TPM: transcripts per million; TRP: transient receptor potential.

Declarations

Acknowledgements

We would like to thank Dr. Paul Katz and the Grass Fellowship program (Grass Foundation) for supporting AS in the initial stages of this project, Dr. Carolyn Smith for her comments on the manuscript, and Brian Novogradac for his support establishing software pipelines on the University of Toronto Mississauga Research Cluster.

Funding

This project was supported by an NSERC Discovery Grant (RGPIN–2016–06023), a Canadian Foundation for Innovation Grant (35297), University of Toronto Start-Up Funds and a Grass Fellowship to A. Senatore. These funding bodies did not play a role in the design of the study, the data acquisition and analysis, and the writing of the manuscript.

Availability of data and materials

The four cDNA libraries were deposited to NCBI under SRA SUB5265969, and the transcriptome assembly to GenBank under accession SUB5274527.

Author's contributions

AS designed the project and funded the research. AS, YYW and PL designed the experiments, and YYW and PL conducted the transcriptome assembly and bioinformatics analyses. AS and WE conducted the cloning and analysis of the ASIC receptor subunits. AS and YYW wrote the manuscript. TP performed necessary revisions of the manuscript and figures.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1.Hille B: *Ion channels of excitable membranes/Bertil Hille:* Sinauer Associates, Inc; 2001.

2.Kaczmarek LK, Levitan IB: *Neuromodulation: the biochemical control of neuronal excitability:* Oxford University Press, USA; 1987.

3.Pavlos NJ, Friedman PA: *GPCR signaling and trafficking: the long and short of it. Trends in Endocrinology & Metabolism* 2017, *28(*3):213–226.

4.Marder E: *Neuromodulation of neuronal circuits: back to the future. Neuron* 2012, *76(*1):1–11.

5.Katz PS, Lillvis JL: *Reconciling the deep homology of neuromodulation with the evolution of behavior. Current Opinion in Neurobiology* 2014, *29:*39–47.

6.Smith CL, Pivovarova N, Reese TS: *Coordinated feeding behavior in Trichoplax, an animal without synapses. PLoS One* 2015, *10(*9):e0136098.

7.Mayorova TD, Smith CL, Hammar K, Winters CA, Pivovarova NB, Aronova MA, Leapman RD, Reese TS: *Cells containing aragonite crystals mediate responses to gravity in Trichoplax adhaerens (Placozoa), an animal lacking neurons and synapses. PloS one* 2018, *13(*1):e0190905.

8.Nikitin M: *Bioinformatic prediction of Trichoplax adhaerens regulatory peptides. General and comparative endocrinology* 2015, *212:*145–155.

9.Senatore A, Reese TS, Smith CL: *Neuropeptidergic integration of behavior in Trichoplax adhaerens, an animal without synapses. Journal of Experimental Biology* 2017, *220(*18):3381–3390.

10.Srivastava M, Begovic E, Chapman J, Putnam NH, Hellsten U, Kawashima T, Kuo A, Mitros T, Salamov A, Carpenter ML: *The Trichoplax genome and the nature of placozoans. Nature* 2008, *454(*7207):955.

11.Varoqueaux F, Williams EA, Grandemange S, Truscello L, Kamm K, Schierwater B, Jekely G, Fasshauer D: *High cell diversity and complex peptidergic signaling underlie placozoan behavior. Current Biology* 2018, *28(*21):3495–3501. e3492.

12.Smith CL, Varoqueaux F, Kittelmann M, Azzam RN, Cooper B, Winters CA, Eitel M, Fasshauer D, Reese TS: *Novel cell types, neurosecretory cells, and body plan of the early-diverging metazoan Trichoplax adhaerens. Current Biology* 2014, *24(*14):1565–1572.

13.Sebé-Pedrós A, Chomsky E, Pang K, Lara-Astiaso D, Gaiti F, Mukamel Z, Amit I, Hejnol A, Degnan BM, Tanay A: *Early metazoan cell type diversity and the evolution of multicellular gene regulation. Nature ecology & evolution* 2018, *2(7*):1176.

14.Kamm K, Osigus H-J, Stadler PF, DeSalle R, Schierwater B: *Trichoplax genomes reveal profound admixture and suggest stable wild populations without bisexual reproduction. Scientific reports* 2018, *8(*1):11168.

15.Andrews S: *FastQC: a quality control tool for high throughput sequence data. 2010.* In*.;* 2017.

16.Bolger AM, Lohse M, Usadel B: *Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics* 2014, *30(*15):2114–2120.

17.Gilbert D: *Gene-omes built from mRNA-seq not genome DNA.* 2013.

18.Xu H, Luo X, Qian J, Pang X, Song J, Qian G, Chen J, Chen S: *FastUniq: a fast de novo duplicates removal tool for paired short reads. PloS one* 2012, *7(*12):e52249.

19.Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q: *Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature biotechnology* 2011, *29(7*):644.

20.Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L: *Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature protocols* 2012, *7(*3):562.

21.Wu TD, Reeder J, Lawrence M, Becker G, Brauer MJ: *GMAP and GSNAP for genomic sequence alignment: enhancements to speed, accuracy, and functionality.* In: *Statistical Genomics.* Springer; 2016: 283–334.

22.Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL: *TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome biology* 2013, *14(*4):R36.

23.Langmead B, Salzberg SL: *Fast gapped-read alignment with Bowtie 2. Nature methods* 2012, *9(*4):357.

24.Robertson G, Schein J, Chiu R, Corbett R, Field M, Jackman SD, Mungall K, Lee S, Okada HM, Qian JQ: *De novo assembly and analysis of RNA-seq data. Nature methods* 2010, *7(*11):909.

25.Schulz MH, Zerbino DR, Vingron M, Birney E: *Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics* 2012, *28(*8):1086–1092.

26.Xie Y, Wu G, Tang J, Luo R, Patterson J, Liu S, Huang W, He G, Gu S, Li S: *SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. Bioinformatics* 2014, *30(*12):1660–1666.

27.Fu L, Niu B, Zhu Z, Wu S, Li W: *CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics* 2012, *28(*23):3150–3152.

28.Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL: *BLAST+: architecture and applications. BMC bioinformatics* 2009, *10(*1):421.

29.Bairoch A, Apweiler R: *The SWISS-PROT protein sequence data bank and its new supplement TREMBL. Nucleic acids research* 1996, *24(*1):21–25.

30.Pruitt KD, Tatusova T, Maglott DR: *NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic acids research* 2005, *33(*suppl_1):D501-D504.

31.Yang Y, Smith SA: *Optimizing de novo assembly of short-read RNA-seq data for phylogenomics. BMC genomics* 2013, *14(*1):328.

32.Eitel M, Francis WR, Varoqueaux F, Daraspe J, Osigus H-J, Krebs S, Vargas S, Blum H, Williams GA, Schierwater B: *Comparative genomics and the nature of placozoan species. PLoS biology* 2018, *16(7*):e2005359.

33.Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M: *Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics* 2005, *21(*18):3674–3676.

34.Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT: *Gene ontology: tool for the unification of biology. Nature genetics* 2000, *25(*1):25.

35.Li B, Dewey CN: *RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC bioinformatics* 2011, *12(*1):323.

36.Simons M, Raposo G: *Exosomes–vesicular carriers for intercellular communication. Current opinion in cell biology* 2009, *21(*4):575–581.

37.Käll L, Krogh A, Sonnhammer EL: *A combined transmembrane topology and signal peptide prediction method. Journal of molecular biology* 2004, *338(*5):1027–1036.

38.Viklund H, Bernsel A, Skwark M, Elofsson A: *SPOCTOPUS: a combined predictor of signal peptides and membrane protein topology. Bioinformatics* 2008, *24(*24):2928–2929.

39.Petersen TN, Brunak S, von Heijne G, Nielsen H: *SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature methods* 2011, *8(*10):785.

40.Krogh A, Larsson B, Von Heijne G, Sonnhammer EL: *Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. Journal of molecular biology* 2001, *305(*3):567–580.

41.Emanuelsson O, Nielsen H, Brunak S, Von Heijne G: *Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. Journal of molecular biology* 2000, *300(*4):1005–1016.

42.Kamm K, Schierwater B, DeSalle R: *Innate immunity in the simplest animals–placozoans. BMC Genomics* 2019, *20(*1):5.

43.Lindsay LL, Wieduwilt MJ, Hedrick JL: *Oviductin, the Xenopus laevis oviductal protease that processes egg envelope glycoprotein gp43, increases sperm binding to envelopes, and is translated as part of an unusual mosaic protein composed of two protease and several CUB domains. Biology of reproduction* 1999, *60(*4):989–995.

44.Wahli W: *Evolution and expression of vitellogenin genes. Trends in Genetics* 1988, *4(*8):227–232.

45.Eitel M, Guidi L, Hadrys H, Balsamo M, Schierwater B: *New insights into placozoan sexual reproduction and development. PLoS One* 2011, *6(*5):e19639.

46.Signorovitch AY, Dellaporta SL, Buss LW: *Molecular signatures for sex in the Placozoa. Proceedings of the National Academy of Sciences* 2005, *102(*43):15518–15522.

47.Ringrose JH, Van Den Toorn HW, Eitel M, Post H, Neerincx P, Schierwater B, Altelaar AM, Heck AJ: *Deep proteome profiling of Trichoplax adhaerens reveals remarkable features at the origin of metazoan multicellularity. Nature communications* 2013, *4:*1408.

48.Jovine L, Park J, Wassarman PM: *Sequence similarity between stereocilin and otoancorin points to a unified mechanism for mechanotransduction in the mammalian inner ear. BMC cell biology* 2002, *3(*1):28.

49.Sun M, Wartel M, Cascales E, Shaevitz JW, Mignot T: *Motor-driven intracellular transport powers bacterial gliding motility. Proceedings of the National Academy of Sciences* 2011, *108(*18):7559–7564.

50.Southey BR, Amare A, Zimmerman TA, Rodriguez-Zas SL, Sweedler JV: *NeuroPred: a tool to predict cleavage sites in neuropeptide precursors and provide the masses of the resulting peptides. Nucleic acids research* 2006, *34(*suppl_2):W267-W272.

51.Jeziorski MC, Green KA, Sommerville J, Cottrell GA: *Cloning and expression of a FMRFamide-gated Na+ channel from Helisoma trivolvis and comparison with the native neuronal channel. The Journal of physiology* 2000, *526(*1):13–25.

52.Perry SJ, Straub VA, Schofield MG, Burke JF, Benjamin PR: *Neuronal expression of an FMRFamide-gated Na+ channel and its modulation by acid pH. Journal of Neuroscience* 2001, *21(*15):5559–5567.

53.Lingueglia E, Champigny G, Lazdunski M, Barbry P: *Cloning of the amiloride-sensitive FMRFamide peptide-gated sodium channel. Nature* 1995, *378(*6558):730.

54.Assmann M, Kuhn A, Dürrnagel S, Holstein TW, Gründer S: *The comprehensive analysis of DEG/ENaC subunits in Hydra reveals a large variety of peptide-gated channels, potentially involved in neuromuscular transmission. BMC biology* 2014, *12(*1):84.

55.Moroz LL, Kocot KM, Citarella MR, Dosung S, Norekian TP, Povolotskaya IS, Grigorenko AP, Dailey C, Berezikov E, Buckley KM: *The ctenophore genome and the evolutionary origins of neural systems. Nature* 2014, *510(*7503):109.

56.Edgar RC: *MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research* 2004, *32(*5):1792–1797.

57.Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L: *InterPro: the integrative protein signature database. Nucleic acids research* 2008,

*37(*suppl_1):D211-D215.

58.Moroz LL, Kohn AB: *Unbiased view of synaptic and neuronal gene complement in ctenophores: are there pan-neuronal and pan-synaptic genes across metazoa? Integrative and comparative biology* 2015, *55(*6):1028–1049.

59.Moroz LL, Kohn AB: *Independent origins of neurons and synapses: insights from ctenophores. Philosophical Transactions of the Royal Society B: Biological Sciences* 2016, *371(*1685):20150041.

60.Senatore A, Raiss H, Le P: *Physiology and evolution of voltage-gated calcium channels in early diverging animal phyla: Cnidaria, placozoa, porifera and ctenophora. Frontiers in physiology* 2016, *7:*481.

61.Currie KP: *G protein inhibition of CaV2 calcium channels. Channels* 2010, *4(*6):497–509.

62.Wistrand M, Käll L, Sonnhammer EL: *A general model of G protein-coupled receptor sequences and its application to detect remote homologs. Protein Science* 2006, *15(*3):509–521.

63.Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G: *InterProScan 5: genome-scale protein function classification. Bioinformatics* 2014, *30(*9):1236–1240.

64.Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T: *trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics* 2009, *25(*15):1972–1973.

65.Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ: *IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Molecular biology and evolution* 2014, *32(*1):268–274.

66.Ellwanger K, Eich A, Nickel M: *GABA and glutamate specifically induce contractions in the sponge Tethya wilhelma. Journal of Comparative Physiology A* 2007, *193(*1):1–11.

67.Moroz LL, Kohn AB: *Parallel evolution of nitric oxide signaling: diversity of synthesis & memory pathways. Frontiers in bioscience (Landmark edition)* 2011, *16:*2008.

68.Müller CS, Haupt A, Bildl W, Schindler J, Knaus H-G, Meissner M, Rammner B, Striessnig J, Flockerzi V, Fakler B: *Quantitative proteomics of the Cav2 channel nano-environments in the mammalian brain. Proceedings of the National Academy of Sciences* 2010, *107(*34):14950–14957.

69.Maximov A, Südhof TC, Bezprozvanny I: *Association of neuronal calcium channels with modular adaptor proteins. Journal of Biological Chemistry* 1999, *274(*35):24453–24456.

70.Südhof TC: *The presynaptic active zone. Neuron* 2012, *75(*1):11–25.

71.Liu KS, Siebert M, Mertel S, Knoche E, Wegener S, Wichmann C, Matkovic T, Muhammad K, Depner H, Mettke C: *RIM-binding protein, a central part of the active zone, is essential for neurotransmitter release. Science* 2011, *334(*6062):1565–1569.

72.Kiyonaka S, Nakajima H, Takada Y, Hida Y, Yoshioka T, Hagiwara A, Kitajima I, Mori Y, Ohtsuka T: *Physical and functional interaction of the active zone protein CAST/ERC2 and the β-subunit of the voltage-dependent Ca2+ channel. The Journal of Biochemistry* 2012, *152(*2):149–159.

73.Graf ER, Valakh V, Wright CM, Wu C, Liu Z, Zhang YQ, DiAntonio A: *RIM promotes calcium channel accumulation at active zones of the Drosophila neuromuscular junction. Journal of Neuroscience* 2012, *32(*47):16586–16596.

74.Kaeser PS, Deng L, Wang Y, Dulubova I, Liu X, Rizo J, Südhof TC: *RIM proteins tether Ca2+ channels to presynaptic active zones via a direct PDZ-domain interaction. Cell* 2011, *144(*2):282–295.

75.Paps J, Holland PW: *Reconstruction of the ancestral metazoan genome reveals an increase in genomic novelty. Nature communications* 2018, *9.*

76.Fenster SD, Chung WJ, Zhai R, Cases-Langhoff C, Voss B, Garner AM, Kaempf U, Kindler S, Gundelfinger ED, Garner CC: *Piccolo, a presynaptic zinc finger protein structurally related to bassoon. Neuron* 2000, *25(*1):203–214.

77.Wagh DA, Rasse TM, Asan E, Hofbauer A, Schwenkert I, Dürrbeck H, Buchner S, Dabauvalle M-C, Schmidt M, Qin G: *Bruchpilot, a protein with homology to ELKS/CAST, is required for structural integrity and function of synaptic active zones in Drosophila. Neuron* 2006, *49(*6):833–844.

78.Moran Y, Barzilai MG, Liebeskind BJ, Zakon HH: *Evolution of voltage-gated ion channels at the emergence of Metazoa. Journal of Experimental Biology* 2015, *218(*4):515–525.

79.Srivastava M, Simakov O, Chapman J, Fahey B, Gauthier ME, Mitros T, Richards GS, Conaco C, Dacre M, Hellsten U: *The Amphimedon queenslandica genome and the evolution of animal complexity. Nature* 2010, *466(*7307):720.

80.Riesgo A, Farrar N, Windsor PJ, Giribet G, Leys SP: *The analysis of eight transcriptomes from all poriferan classes reveals surprising genetic complexity in sponges. Molecular biology and evolution* 2014, *31(*5):1102–1120.

81.Conaco C, Bassett DS, Zhou H, Arcila ML, Degnan SM, Degnan BM, Kosik KS: *Functionalization of a protosynaptic gene expression network. Proceedings of the National Academy of Sciences* 2012, *109(*Supplement 1):10612–10618.

82.Ryan JF, Chiodin M: *Where is my mind? How sponges and placozoans may have lost neural cell types. Philosophical Transactions of the Royal Society B: Biological Sciences* 2015, *370(*1684):20150059.

83.Sheng M, Hoogenraad CC: *The postsynaptic architecture of excitatory synapses: a more quantitative view. Annu Rev Biochem* 2007, *76:*823–847.

84.Missler M, Südhof TC, Biederer T: *Synaptic cell adhesion. Cold Spring Harbor perspectives in biology* 2012, *4(*4):a005694.

85.Ryan JF, Pang K, Schnitzler CE, Nguyen A-D, Moreland RT, Simmons DK, Koch BJ, Francis WR, Havlak P, Smith SA: *The genome of the ctenophore Mnemiopsis leidyi and its implications for cell type evolution. Science* 2013, *342(*6164):1242592.

86.Choii G, Ko J: *Gephyrin: a central GABAergic synapse organizer. Experimental & molecular medicine* 2015, *47(*4):e158.

87.Ritter SL, Hall RA: *Fine-tuning of GPCR activity by receptor-interacting proteins. Nature reviews Molecular cell biology* 2009, *10(*12):819.

88.Zhang H, Maximov A, Fu Y, Xu F, Tang T-S, Tkatch T, Surmeier DJ, Bezprozvanny I: *Association of CaV1. 3 L-type calcium channels with Shank. Journal of Neuroscience* 2005, *25(*5):1037–1049.

89.Laplante M, Sabatini DM: *mTOR signaling at a glance. Journal of cell science* 2009, *122(*20):3589–3594.

90.Sun Y, Liu W-Z, Liu T, Feng X, Yang N, Zhou H-F: *Signaling pathway of MAPK/ERK in cell proliferation, differentiation, migration, senescence and apoptosis. Journal of Receptors and Signal Transduction* 2015, *35(*6):600–604.

91.Sakarya O, Conaco C, Eğecioğlu Ö, Solla SA, Oakley T, Kosik K: *Evolutionary expansion and specialization of the PDZ domains. Molecular biology and evolution* 2009, *27(*5):1058–1069.

92.Dehal P, Boore JL: *Two rounds of whole genome duplication in the ancestral vertebrate. PLoS biology* 2005, *3(*10):e314.

93.Gerdol M, Venier P, Pallavicini A: *The genome of the Pacific oyster Crassostrea gigas brings new insights on the massive expansion of the C1q gene family in Bivalvia. Developmental & Comparative Immunology* 2015, *49(*1):59–71.

94.Gao D, Ko DC, Tian X, Yang G, Wang L: *Expression divergence of duplicate genes in the protein kinase superfamily in Pacific oyster. Evolutionary Bioinformatics* 2015, *11:*EBO. S30230.

95.Hernandez-Nicaise M-L: *The nervous system of ctenophores III. Ultrastructure of synapses. Journal of neurocytology* 1973, *2(*3):249–263.

96.Wenderoth H: *Transepithelial cytophagy by Trichoplax adhaerens FE Schulze (Placozoa) feeding on yeast. Zeitschrift für Naturforschung C* 1986, *41(*3):343–347.

97.Corona M, Velarde RA, Remolina S, Moran-Lauter A, Wang Y, Hughes KA, Robinson GE: *Vitellogenin, juvenile hormone, insulin signaling, and queen honey bee longevity. Proceedings of the National Academy*

of Sciences 2007, *104(*17):7128–7133.

98.Nelson CM, Ihle KE, Fondrk MK, Page Jr RE, Amdam GV: *The gene vitellogenin has multiple coordinating effects on social organization. PLoS biology* 2007, *5(*3):e62.

99.Burkhardt P, Grønborg M, McDonald K, Sulur T, Wang Q, King N: *Evolutionary insights into premetazoan functions of the neuronal protein homer. Molecular biology and evolution* 2014, *31(*9):2342–2355.

100.Driscoll T, Gillespie JJ, Nordberg EK, Azad AF, Sobral BW: *Bacterial DNA sifted from the Trichoplax adhaerens (Animalia: Placozoa) genome project reveals a putative rickettsial endosymbiont. Genome biology and evolution* 2013, *5(*4):621–645.

101.Grell KG, G B: *Die Ultrastruktur von Trichoplax adhaerens F. E. Schulze. Cytobiologie* 1971, *4:*216–240.

102.Chandana T, P Venkatesh Y: *Occurrence, functions and biological significance of arginine-rich proteins. Current Protein and Peptide Science* 2016, *17(*5):507–516.

103.Leys SP: *Elements of a 'nervous system'in sponges. Journal of Experimental Biology* 2015, *218(*4):581–591.

104.Prosser CL: *Ionic analyses and effects of ions on contractions of sponge tissues. Zeitschrift für vergleichende Physiologie* 1967, *54(*2):109–120.

105.de Ceccatty MP: *Effects of drugs and ions on a primitive system of spontaneous contractions in a sponge (Euspongia officinalis). Experientia* 1971, *27(*1):57–59.

106.Morgan AJ, Jacob R: *Ionomycin enhances Ca2+ influx by stimulating store-regulated cation entry and not by a direct action at the plasma membrane. Biochemical Journal* 1994, *300(*3):665–672.

107.Thiemann M, Ruthmann A: *Microfilaments and microtubules in isolated fiber cells of Trichoplax adhaerens (Placozoa). Zoomorphology* 1989, *109(*2):89–96.

108.Smith CL, Abdallah S, Wong YY, Le P, Harracksingh AN, Artinian L, Tamvacakis AN, Rehder V, Reese TS, Senatore A: *Evolutionary insights into T-type Ca2+ channel structure, function, and ion selectivity from the Trichoplax adhaerens homologue. The Journal of general physiology* 2017, *149(*4):483–510.

109.Nordström KJ, Sällman Almén M, Edstam MM, Fredriksson R, Schiöth HB: *Independent HHsearch, Needleman–Wunsch-based, and motif analyses reveal the overall hierarchy for most of the G protein-coupled receptor families. Molecular biology and evolution* 2011, *28(*9):2471–2480.

110.Krishnan A, Schiöth HB: *The role of G protein-coupled receptors in the early evolution of neurotransmission and the nervous system. Journal of Experimental Biology* 2015, *218(*4):562–571.

111.Simpson TL: *The cell biology of sponges:* Springer Science & Business Media; 2012.

112.Krishnan A, Dnyansagar R, Almén MS, Williams MJ, Fredriksson R, Manoj N, Schiöth HB: *The GPCR repertoire in the demosponge Amphimedon queenslandica: insights into the GPCR system at the early divergence of animals. BMC evolutionary biology* 2014, *14(*1):270.

113.De Mendoza A, Sebé-Pedrós A, Ruiz-Trillo I: *The evolution of the GPCR signaling system in eukaryotes: modularity, conservation, and the transition to metazoan multicellularity. Genome biology and evolution* 2014, *6(*3):606–619.

114.Hille B: *Modulation of ion-channel function by G-protein-coupled receptors. Trends in neurosciences* 1994, *17(*12):531–536.

115.Sassone-Corsi P: *The cyclic AMP pathway. Cold Spring Harbor perspectives in biology* 2012, *4(*12):a011148.

116.Gao T, Yatani A, Dell'Acqua ML, Sako H, Green SA, Dascal N, Scott JD, Hosey MM: *cAMP-dependent regulation of cardiac L-type Ca2+ channels requires membrane targeting of PKA and phosphorylation of channel subunits. Neuron* 1997, *19(*1):185–196.

117.Shearman MS, Sekiguchi K, Nishizuka Y: *Modulation of ion channel activity: a key function of the protein kinase C enzyme family. Pharmacological Reviews* 1989, *41(*2):211–237.

118.Gamper N, Shapiro MS: *Target-specific PIP2 signalling: how might it work? The Journal of physiology* 2007, *582(*3):967–975.

119.Barry MF, Ziff EB: *Receptor trafficking and the plasticity of excitatory synapses. Current opinion in neurobiology* 2002, *12(*3):279–286.

120.Prole DL, Taylor CW: *Structure and Function of IP3 Receptors. Cold Spring Harbor perspectives in biology* 2019:a035063.

121.Dascal N: *Ion-channel regulation by G proteins. Trends in Endocrinology & Metabolism* 2001, *12(*9):391–398.

122.Alié A, Manuel M: *The backbone of the post-synaptic density originated in a unicellular ancestor of choanoflagellates and metazoans. BMC evolutionary biology* 2010, *10(*1):34.

123.Tonikian R, Zhang Y, Sazinsky SL, Currell B, Yeh J-H, Reva B, Held HA, Appleton BA, Evangelista M, Wu Y: *A specificity map for the PDZ domain family. PLoS biology* 2008, *6(*9):e239.

124.Kim J, Kim I, Yang J-S, Shin Y-E, Hwang J, Park S, Choi YS, Kim S: *Rewiring of PDZ domain-ligand interaction network contributed to eukaryotic evolution. PLoS genetics* 2012, *8(*2):e1002510.

125.Fernandez-Valverde SL, Calcino AD, Degnan BM: *Deep developmental transcriptome sequencing uncovers numerous new genes and enhances gene annotation in the sponge Amphimedon queenslandica. BMC genomics* 2015, *16(*1):387.

126.Ereskovsky AV, Richter DJ, Lavrov DV, Schippers KJ, Nichols SA: *Transcriptome sequencing and delimitation of sympatric Oscarella species (O. carmela and O. pearsei sp. nov) from California, USA. PloS one* 2017, *12(*9):e0183002.

127.King N, Westbrook MJ, Young SL, Kuo A, Abedin M, Chapman J, Fairclough S, Hellsten U, Isogai Y, Letunic I: *The genome of the choanoflagellate Monosiga brevicollis and the origin of metazoans. Nature* 2008, *451(*7180):783.

128.Fredman D, Rentzsch; MSF, Technau U: *Nematostella vectensis transcriptome and gene models v2.0. Figshare Dataset* 2013.

129.Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M: *De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature protocols* 2013, *8(*8):1494.

Tables

Table 1. Table denoting identified secretome genes with BLASTp annotation indicative of roles in digestion.

Gene ID	AVG TPM	Blast Homology	Accession	Species	BitScore	E-Value
evg43312	7901	Trypsin-1	P07477	*Human*	99.8	3.57E-23
evg958433	5271	Phospholipase A2	P00592	*Sus scrofa*	77.4	5.26E-17
evg52845	4969	Phospholipase A1 2	P53357	*Dolichovespula maculata*	99.8	1.67E-22
evg1030094	3401	Kallikrein 1-related peptidase b16	P04071	*Mus musculus*	99.8	1.17E-22
evg1172676	2585	Cathepsin B-like CP1	P92131	*Giardia intestinalis*	99	1.78E-22
evg1236652	1838	Chymotrypsinogen A	P00766	*Bos taurus*	148	2.35E-40
evg1172592	1611	Lysozyme g-like protein 2	Q86SG7	*Human*	94.7	9.97E-22
evg93589	940	Penicillopepsin-1	B8MF81	*Talaromyces stipitatus*	99	6.6E-22
evg1106662	342	Alpha-amylase 4	O42918	*Schizosaccharomyces pombe (Fission yeast)*	97.8	2.5E-20
evg1036213	133	Trypsin-3	P35030	*Human*	99.8	5.72E-23
evg384	132	Cathepsin C	Q26653	*Schistosoma mansoni (Blood fluke)*	98.6	1.21E-20
evg687502	120	Putative phospholipase B-like 2	Q3TCN2	*Mus musculus*	508	1.5E-172
evg1240777	95	MAM and LDL-receptor class A domain-containing protein 2	B3EWZ6	*Acropora millepora*	99	2.37E-20
evg5390	83	Chitobiase	Q54468	*Serratia marcescens*	84.7	5.81E-16

Table 2. Table denoting identified secretome genes with BLASTp annotation indicative of roles in immunity and defense.

Gene ID	AVG TPM	Blast Homology	Accession	Species	BitScore	E-Val
evg958433	5,271	Phospholipase A2	P00592	*Sus scrofa*	77.4	5.26E-17
evg1226704	3,885	Granzyme K	O35205	*Mus musculus*	99.8	3.25E-23
evg588262	1,625	Pro-cathepsin H	P00786	*Rattus norvegicus*	99.8	1.20E-22
evg1172592	1,611	Lysozyme g-like protein 2	Q86SG7	*Homo sapiens*	94.7	9.97E-22
evg107861	1,120	Complement C3	P01024	*Homo sapiens*	97.1	9.59E-19
evg1224954	1,051	Granulins	P23785	*Rattus norvegicus*	95.5	1.57E-19
evg1106487	664	Prosaposin	O13035	*Gallus gallus*	99.4	1.01E-21
evg772045	228	Protein disulfide-isomerase A6 homolog	Q9V438	*Drosophila melanogaster*	99	5.50E-21
evg1030832	187	Prosaposin	P07602	*Homo sapiens*	90.9	1.61E-18
evg1173372	149	Variant surface antigen F	Q49538	*Mycoplasma hyorhinis*	49.3	3.81E-06
evg772232	124	Peroxiredoxin	Q57109	*Methanothermobacter marburgensis*	99.8	1.09E-23
evg11227	114	Protein disulfide-isomerase A6	Q15084	*Human*	99.4	3.62E-21
evg1107447	101	Uncharacterized protein D2005.3	Q93408	*Caenorhabditis elegans*	96.7	3.70E-24

Table 3. Table denoting identified secretome genes with BLASTp annotation indicative of roles in development and reproduction.

Gene ID	AVG TPM	Blast Homology	Accession	Species	BitScore	E-Val
evg1106650	690	Polycystin-1	O08852	*Mus musculus*	65.9	5.72E-10
evg1106594	611	Kielin/chordin-like protein	Q9IBG7	*Xenopus laevis*	74.7	5.66E-12
evg7638	602	Ovochymase-2	Q66TN7	*Rhinella arenarum*	79.7	2.54E-15
evg1338849	295	Golgi-associated plant pathogenesis-related protein 1	Q9H4G4	*Homo sapiens*	87.8	1.07E-19
evg1030554	267	Vitellogenin-2	P02845	*Gallus gallus*	99.8	5.71E-19
evg2802	191	Multiple coagulation factor deficiency protein 2 homolog	Q8K5B3	*Rattus norvegicus*	87.8	5.57E-21
evg1348625	187	Cadherin-related tumor suppressor	P33450	*Drosophila melanogaster*	99.8	8.61E-20
evg1030832	187	Prosaposin	P07602	*Homo sapiens*	90.9	1.61E-18
evg1172861.1	179	Proactivator polypeptide-like 1	Q6NUJ1	*Homo sapiens*	84.7	5.26E-16
evg102425	174	Glutathione S-transferase P	Q8JFZ2	*Xenopus laevis*	95.1	6.38E-22
evg1409072	171	Prosaposin	P26779	*Bos taurus*	99.8	4.00E-21
evg661	157	Oncoprotein-induced transcript 3 protein	Q8R4V5	*Mus musculus*	91.7	2.98E-19
evg17156	133	Laminin subunit beta-1	P07942	*Homo sapiens*	53.5	1.64E-06
evg1330357	119	Neurogenic locus notch homolog protein 3	Q9UM47	*Homo sapiens*	99.8	9.87E-20
evg1326682	95	Prosaposin	P26779	*Bos taurus*	70.5	2.37E-12
evg1789636	80	Leucine-rich repeats and immunoglobulin-like domains protein 3	Q6UXM1	*Homo sapiens*	99.8	9.73E-20
evg1789636	80	Leucine-rich repeats and immunoglobulin-like domains protein 3	Q6UXM1	*Homo sapiens*	99.8	9.73E-20

Table 4. Table denoting identified secretome genes with BLASTp annotation indicative of roles in the cell matrix and cell adhesion.

Gene ID	AVG TPM	Blast Homology	Accession	Species	BitScore	E-Val
evg1173237	779	Mammalian ependymin-related protein 1	Q9N0C7	*Macaca fascicularis*	48.9	1.04E-05
evg207671	466	Ependymin-related protein 2	P86729	*Haliotis asinina*	54.7	6.12E-08
evg267926	297	Fibropellin-3	P49013	*Strongylocentrotus purpuratus*	99.8	2.62E-20
evg1253428	269	Ependymin-related protein 1	P86734	*Haliotis asinina*	48.9	9.68E-06
evg6706	144	Stereocilin	Q8VIM6	*Mus musculus*	91.7	3.27E-17
evg1107223	135	Adventurous-gliding motility protein Z	Q1D823	*Myxococcus xanthus*	50.1	6.99E-05
evg848790	131	Collagen-like protein 2	Q5UQ13	*Acanthamoeba polyphaga mimivirus*	84.3	4.93E-16
evg1027559	101	F-actin-capping protein subunit beta isoforms 1 and 2	P14315	*Gallus gallus*	459	3.47E-162
evg841196	80	Ependymin-related protein 2	P86729	*Haliotis asinina*	62.8	4.54E-11

Figures

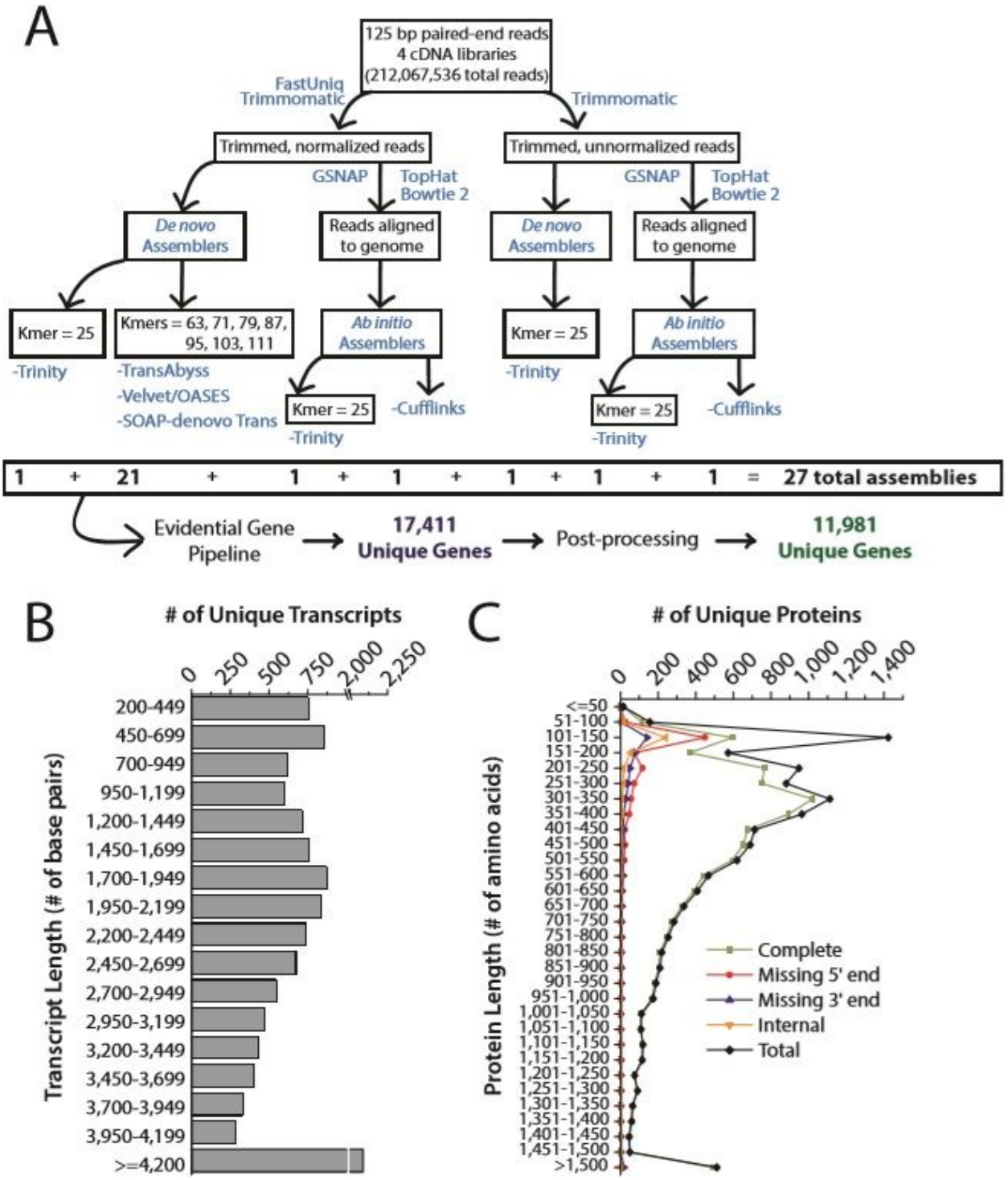

Figure 1

(A) Flow chart of the combined de novo and ab initio assembly pipeline for the Trichoplax whole animal mRNA transcriptome. (B) A histogram of assembled transcripts reveals a normal distribution with minimal enrichment of small, likely fragmented sequences. (C) Histogram of unique translated proteins predicted from the transcriptome with TransDecoder[129] reveals that the vast majority of transcripts contain complete open reading frames (green line), and very few are fragmented by lacking the 5’ end of

the coding sequence (i.e. start codon; red line), the 3’ end (stop codon; blue line) or both (internal; yellow line).

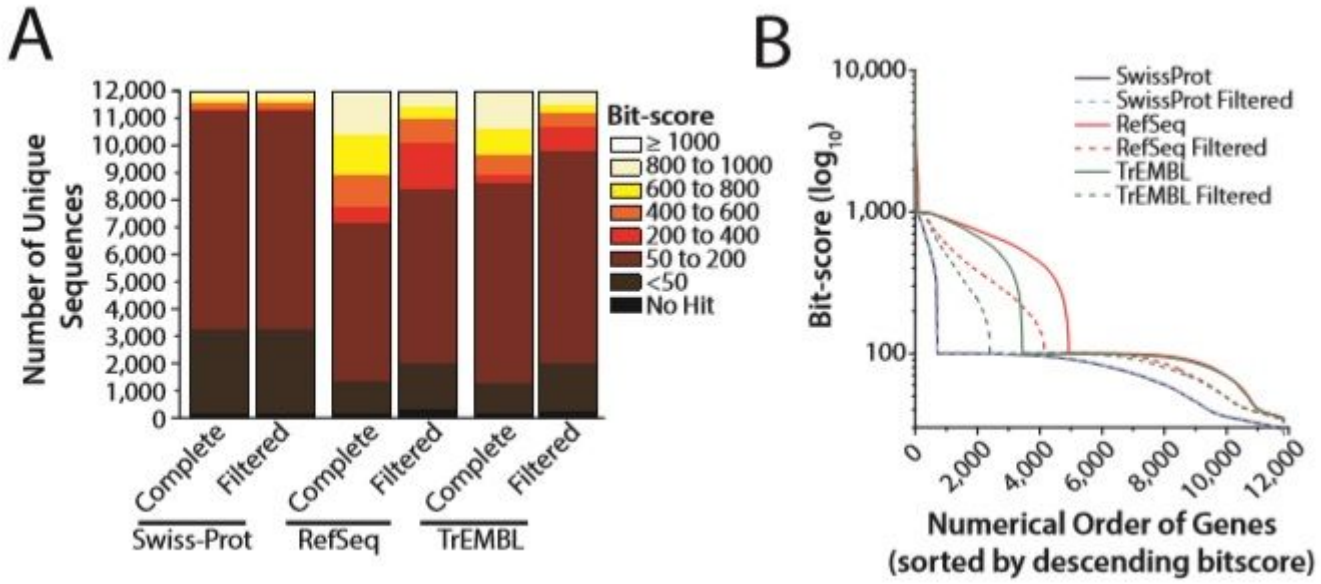

Figure 2

(A) Bar graph illustrating bit-scores for translated protein BLAST alignments of the Trichoplax mRNA transcriptome with the Swiss-Prot, RefSeq and TrEMBL protein databases. Trichoplax transcriptome query sequences were aligned to the three complete protein databases, as well as filtered versions of these databases where Trichoplax gene sequences had been removed. (B) Plot of Trichoplax transcriptome sequences aligned to the complete and filtered Swiss-Prot, RefSeq and TrEMBL protein databases, sorted numerically by decreasing bit-score, against their corresponding bit-scores.

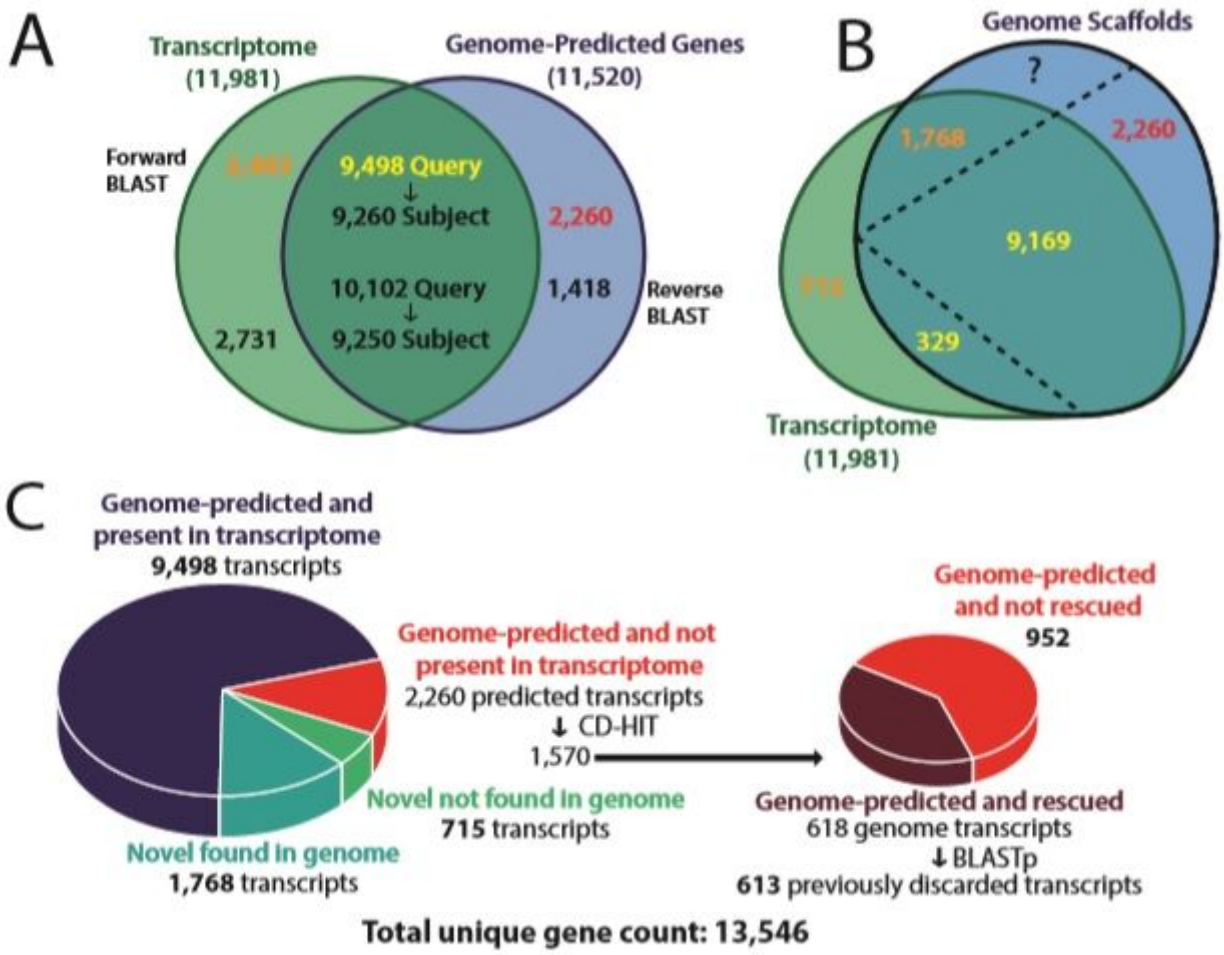

Figure 3

(A) Venn diagram illustrating reciprocal BLAST comparisons between the Trichoplax transcriptome sequences and genes predicted from the published genome. (B) GMAP mapping of Trichoplax transcriptome sequences to the genome scaffolds identifies 715 completely novel genes, plus 1,768 genes present in the genome but not predicted in the initial genome study. (C) Search for the 2,260 genome predicted genes that failed to find a match in the transcriptome identifies an additional 615 transcriptome sequences that were previously discarded during assembly and processing

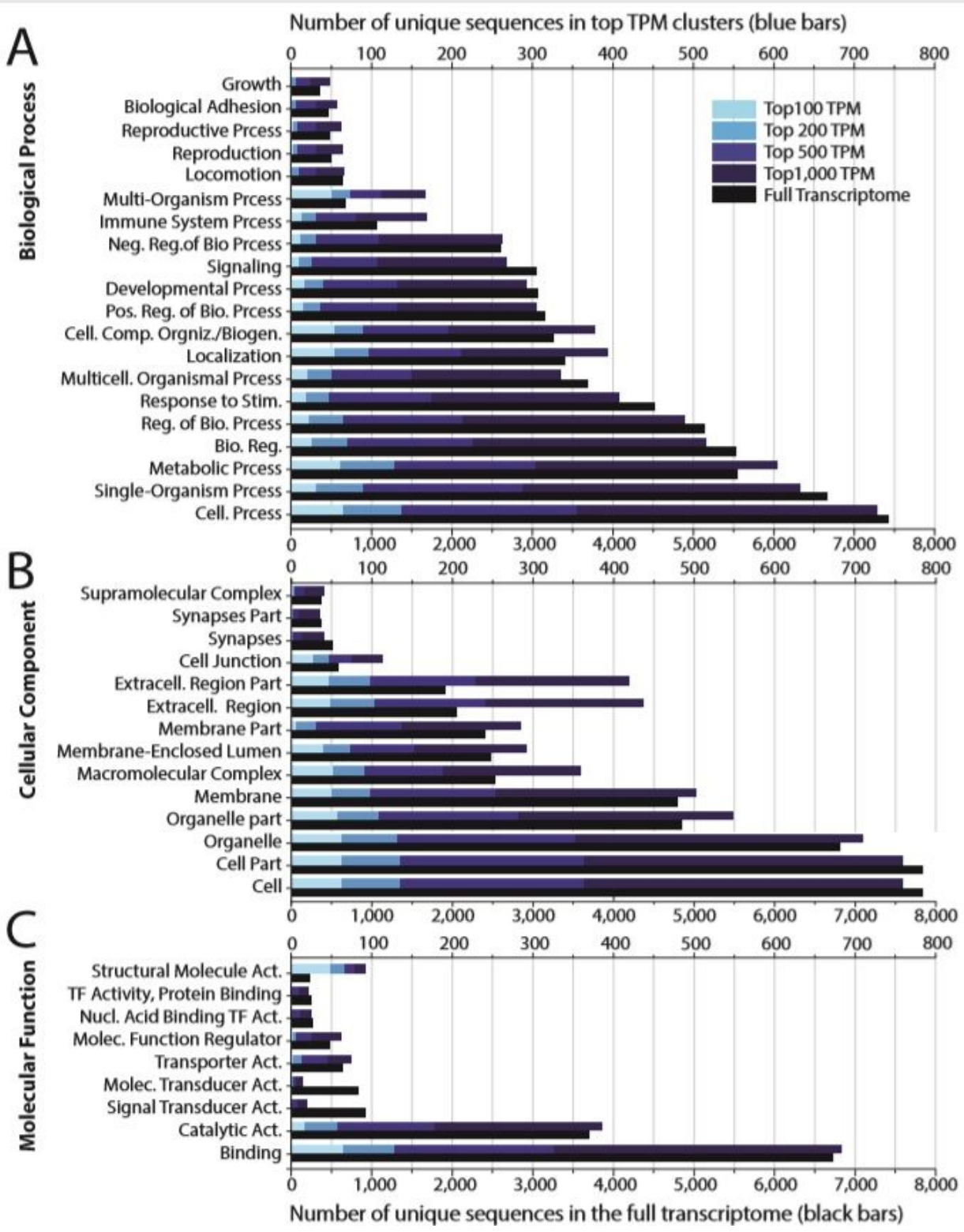

Figure 4

(A) A plot of the number of sequences in the Trichoplax transcriptome receiving BLAST2GO level 2 gene ontology terms in the category of Biological Processes. Black bars represent gene counts among the entire transcriptome (bottom axis), while the blue bars represent subsets of the transcriptome partitioned according to decreasing average transcripts per million (TPM) mRNA expression levels (i.e. top 100 most expressed genes to top 1,000 most expressed genes). (B) Level 2 gene ontology annotation in the

category of Cellular Component. (C) Level 2 gene ontology annotation in the category of Molecular Function.

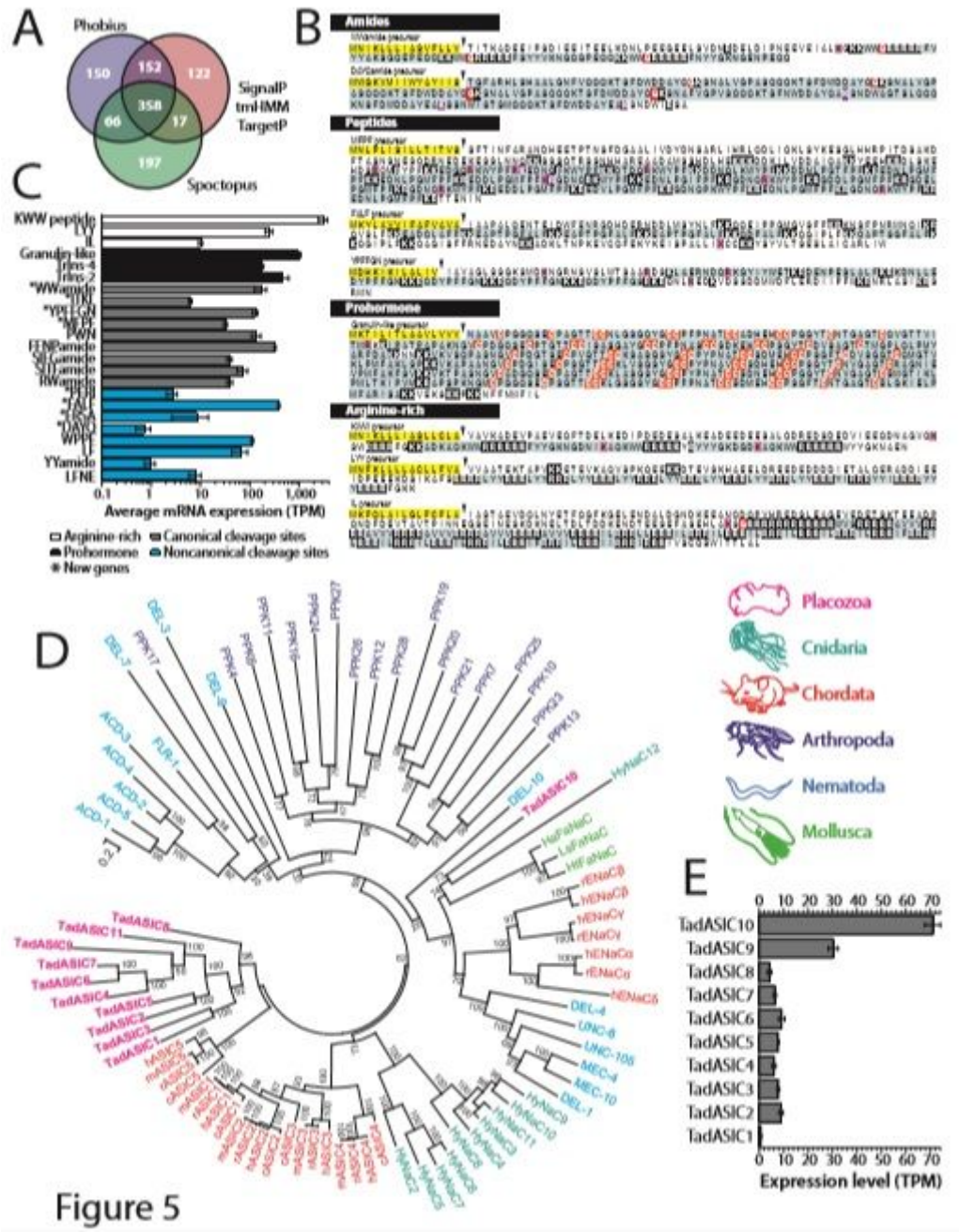

Figure 5

(A) Venn diagram illustrating the numbers of Trichoplax proteins predicted to be secreted by the three separate pipelines Phobius, Spoctopus and SignalP/tmHMM/TargetP. (B) Sequences of some of the newly identified regulatory proteins within the secretome. Signal peptides are depicted in yellow, and the predicted cleavage sites demarked by black arrows. Blue sequences depict predicted mature peptides, putative amidated glycines are in red, and NeuroPred predicted convertase cleavage sites are in black. Basic residues that are likely involved in convertase recognition but are not cleaved are in pink, basic residues we suspect to be cleavage sites but that did not meet the prediction thresholds of NeuroPred are in purple, and cysteine residues are in orange. (C) Bar graph denoting average whole animal TPM

expression levels for the various Trichoplax regulatory peptide and pro-hormone mRNA precursors, ±SE. Asterisks denote novel peptides identified in this study. (D) A maximum likelihood phylogenetic tree of various ASIC channels from Trichoplax, the cnidarian Hydra magnipapillata, the chordate Mus musculus, the arthropod Drosophila melanogaster, the nematode Caenorhabditis elegans, and the gastropod molluscs Helix aspera, Lymnaea stagnalis, and Heliosoma trivolvis. The tree was inferred using IQ-TREE from a protein alignment trimmed with trimAl using a gap threshold of 0.6, with 1,000 ultrafast bootstrap replicates and the substitution model WAG+F+I+G4. (E) Bar graph denoting average TPM expression levels for the 10 ASIC channels identified in the Trichoplax transcriptome, ±SE. Accession numbers for phylogenetic tree: ACD-1:NM_058894, ACD-2:NM_058892, ACD-3: NP_001257250.1, ACD-4:NM_072829, ACD-5:NM_058795, cASIC1:AY956393, cASIC2:NM_001040467, cASIC3:XP_025003153.1, cASIC4:XP_015145601.2, cASIC5: XP_015140827.1, hASIC1a:U78180, hASIC2a:U50352, hASIC2b:NM_18337, hASIC3:AF095897, hASIC4:AJ271643, rASIC1a:U94403, rASIC1b:AJ309926, rASIC2a:U53211, rASIC2b:AB049451, rASIC3:AF013598, rASIC4:AJ271642, hBASIC:AJ252011, mBASIC:Y19035, rBASIC:Y19034; degenerin-like (DEL)-1:U76403, DEL-4:NM_059829, DEL-7:NM_068875, DEL-9:NM_076221, DEL-10:NM-062901, αhENaC:X76180, βhENaC:X87159, γhENaC:X87160, δhENaC:U38254, αrENaC:X70521, βrENaC:X77932, γrENaC:X77933, HtFaNaC:AF254118, LsFaNaC:AF335548, HaFaNaC:X92113, fluoride-resistant (FLR)-1:AB012617, HyNaC2:AM393878, HyNaC3:AM393880, HyNaC4:AM393881, HyNaC5:FN257513, HyNaC6:HG422725, HyNaC7:HG422726, HyNaC8:HG422727, HyNaC9:HG422728, HyNaC10:HG422729, HyNaC11:HG422730, HyNaC12:HG422731, mechanosensitive (MEC)-4:X58982, MEC-10:L25312, pickpocket (PPK):Y16225, PPK4:NM_206137, PPK6:NM_137617, PPK7:NM_135172, PPK10:NM_001038805, PPK11:NM_001038798, PPK12:NM_137828, PPK13:NM_001014495, PPK16:NM_001038797, PPK17:NM_135965, PPK19:AY226547, PPK20:NM_143448, PPK21:NM_143447, PPK23:NM_001014749, PPK24:NM_143603, PPK25:NM_206044, PPK26:NM_139868, PPK27:NM_139569, PPK28:NM_001014748, ripped pocket (RPK):Y12640, uncoordinated (UNC)-8:U76402, U¬NC-105:NM_063301.

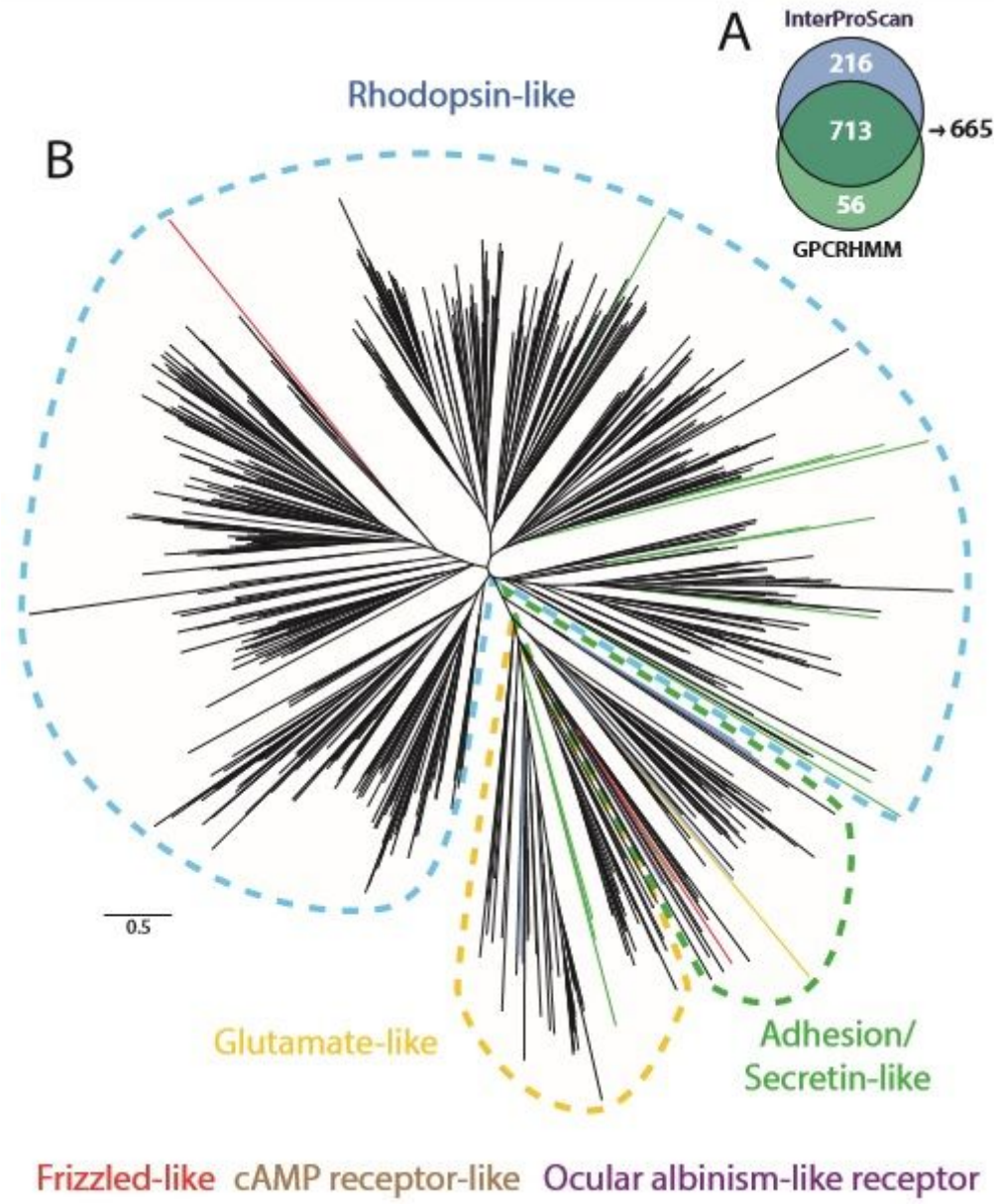

Figure 6

(A) Venn diagram illustrating the number of predicted Trichoplax G-protein coupled receptors using a combined InterProScan and GPCRHMM approach. (B) A maximum likelihood phylogenetic tree of the 665 identified Trichoplax G-protein coupled receptors, inferred with IQ-TREE (model VT+F+G4) from a MUSCLE protein alignment trimmed with trimAl with a gap threshold of 0.5. Colored bubbles demark clades of GPCRs that received a majority of InterPro domain annotations for distinct classes of GPCRs, while colored branches depict GPCRs of different classes that clustered outside of their group.

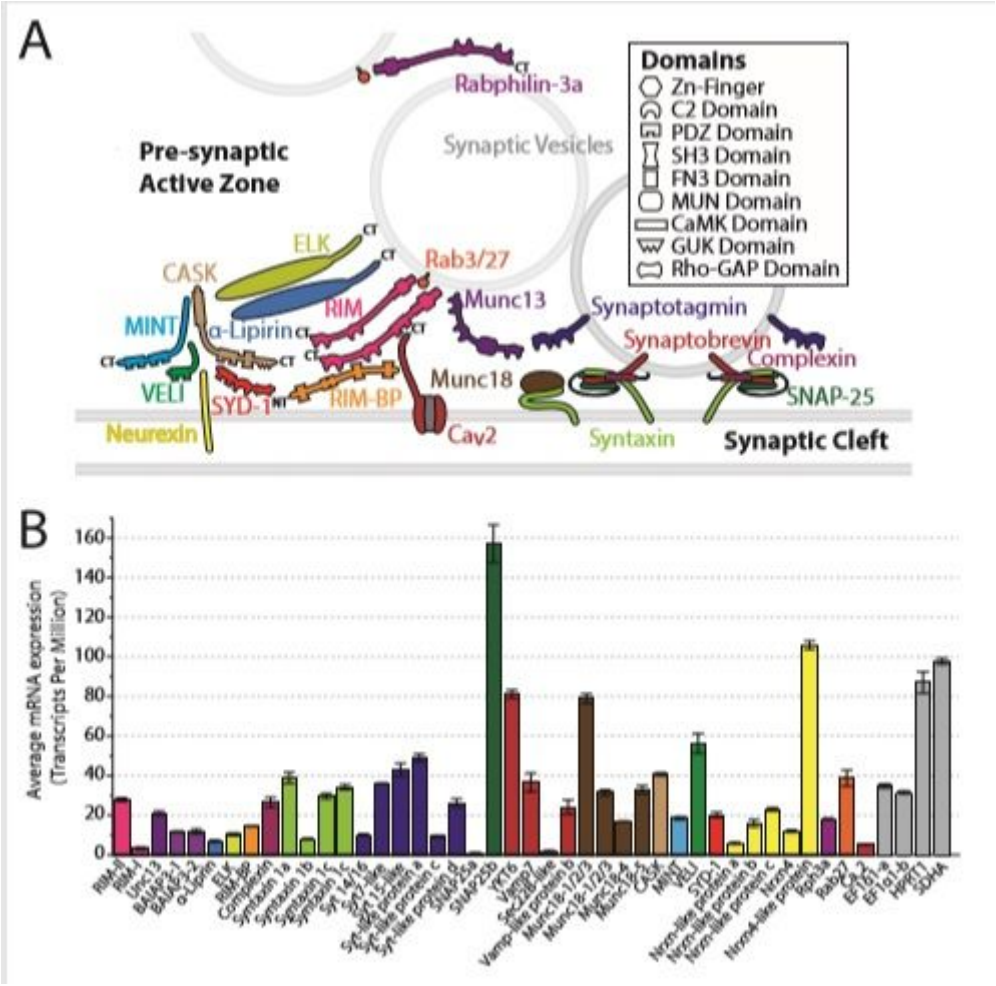

Figure 7

(A) Illustration of core pre-synaptic proteins identified in the Trichoplax transcriptome. The legend on the right depicts protein interaction domains found within the Trichoplax pre-synaptic protein homologues shown in the illustration. (B) Bar graph denoting the average whole animal TPM mRNA expression levels, ± SE, for identified genes shown in panel A.

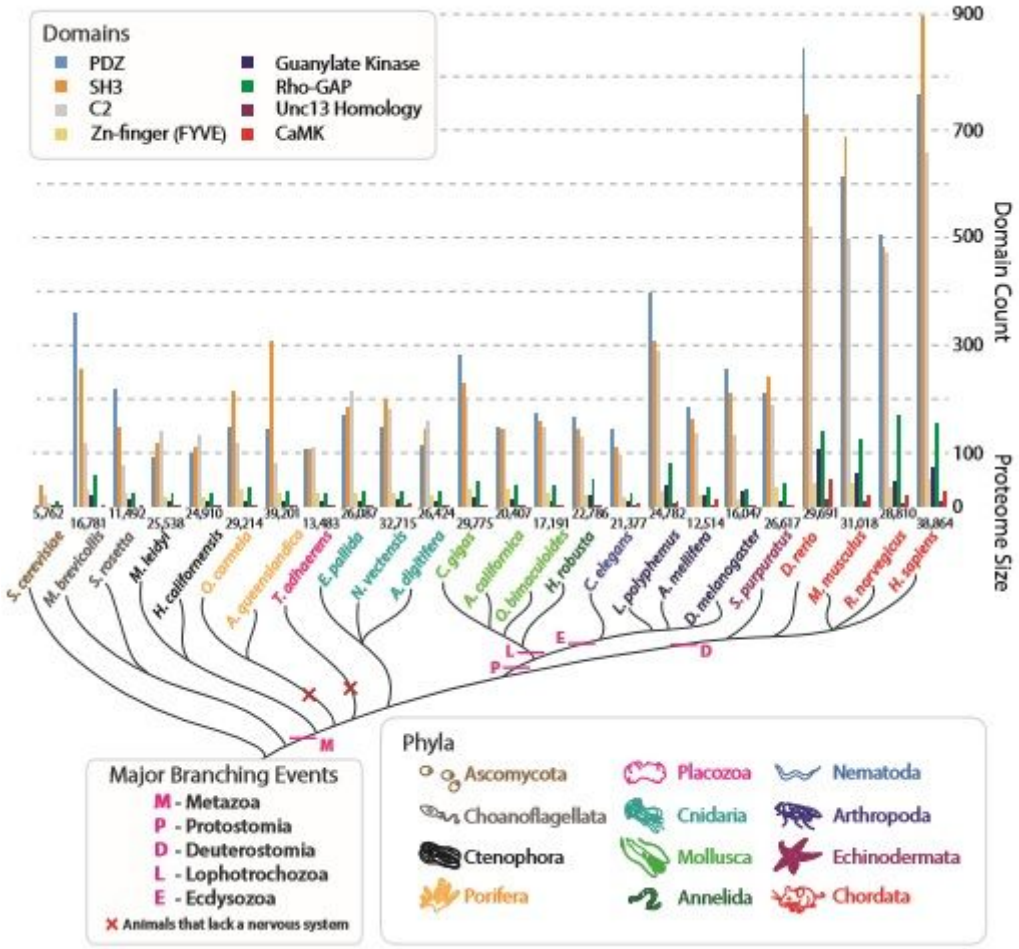

Figure 8

Bar graph denoting counts of select protein interactions domains predicted within respective proteomes of numerous metazoan species with InterProScan. Numbers at the base of the bar graphs denote the total number of non-redundant proteins contained within the proteome of each species.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFiles.docx
- SupplementaryTables.xlsx