Journal Pre-proof

Quantitative proteome dynamics across embryogenesis in a model chordate

Alexander N. Frese, Andrea Mariossi, Michael S. Levine, Martin Wühr

PII: S2589-0042(24)00576-5

DOI: https://doi.org/10.1016/j.isci.2024.109355

Reference: ISCI 109355

To appear in: *ISCIENCE*

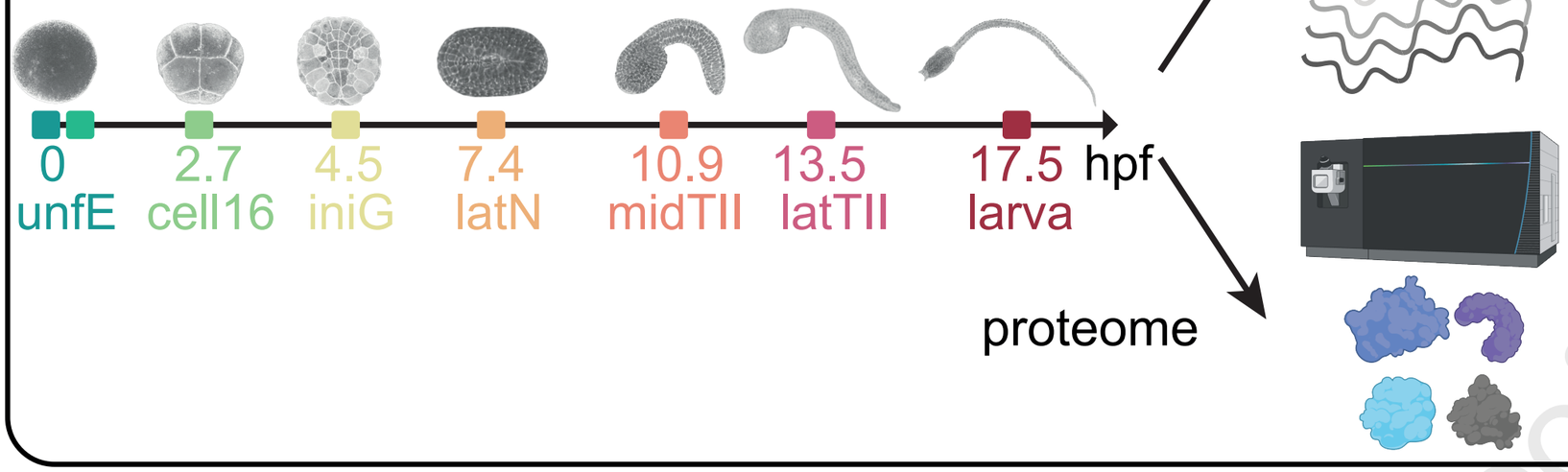
Received Date: 14 August 2023

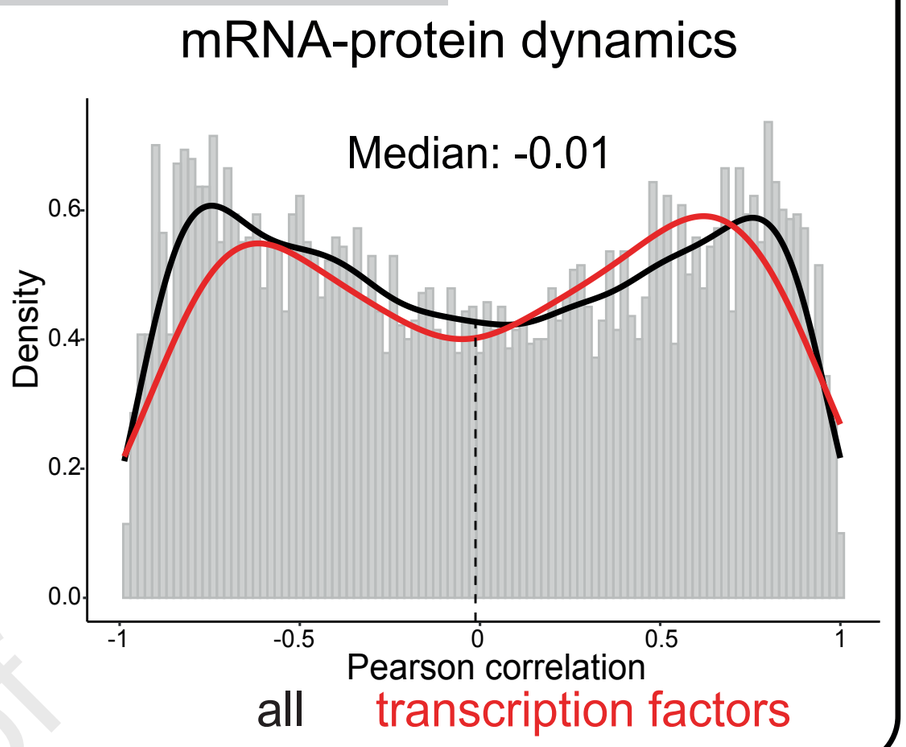
Revised Date: 11 December 2023

Accepted Date: 23 February 2024

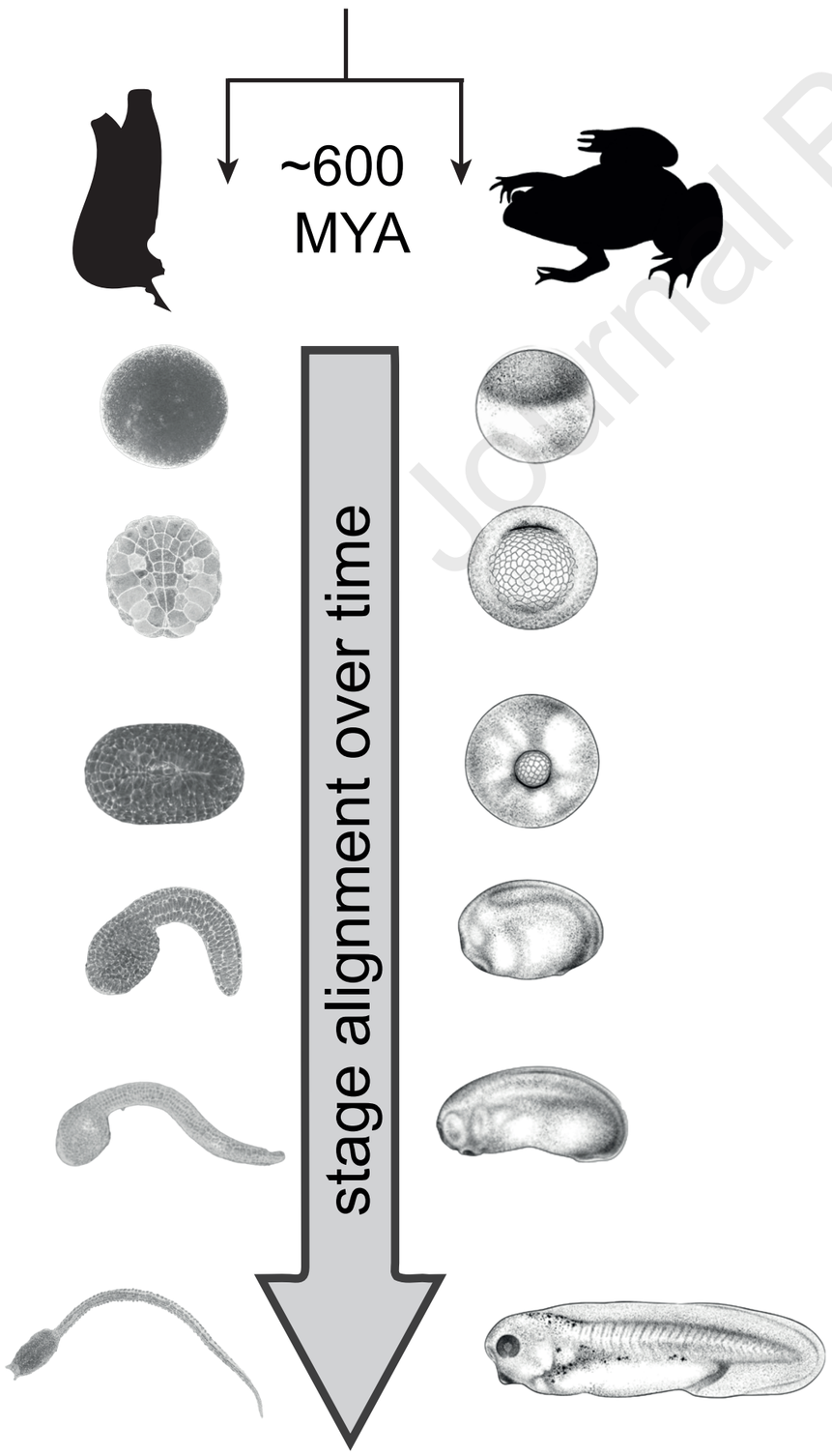
Please cite this article as: Frese, A.N., Mariossi, A., Levine, M.S., Wühr, M., Quantitative proteome dynamics across embryogenesis in a model chordate *ISCIENCE* (2024), doi: https://doi.org/10.1016/j.isci.2024.109355.

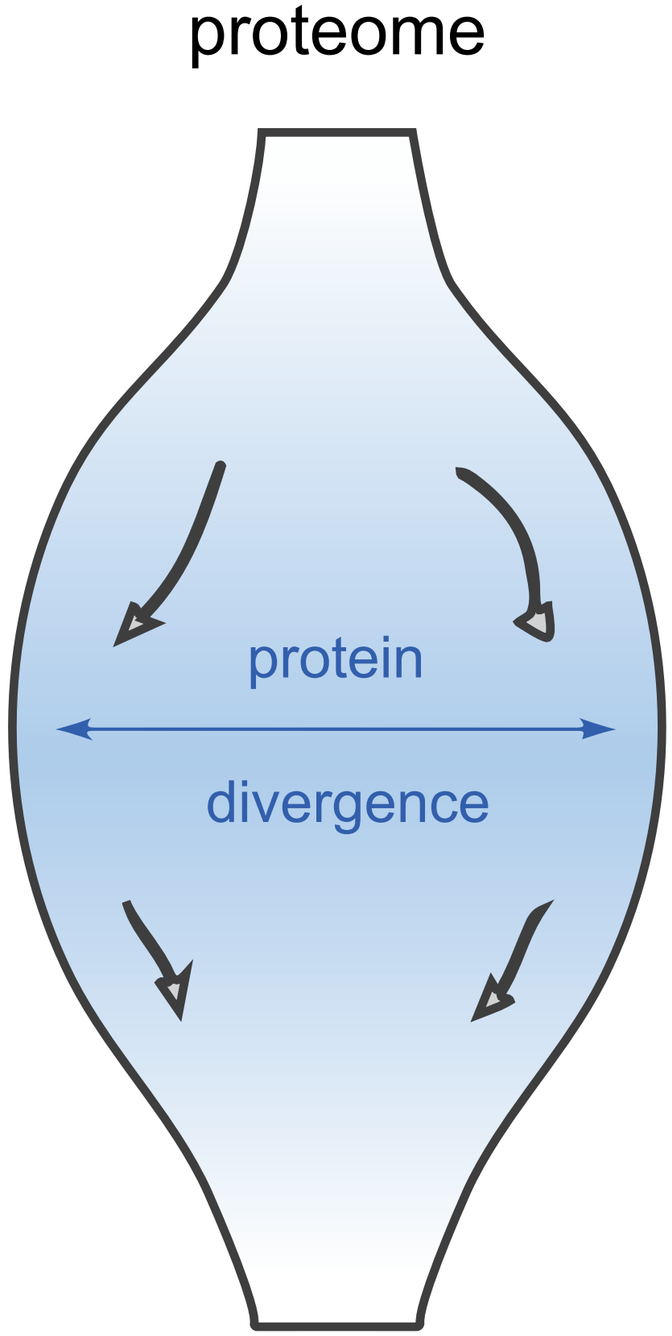
Multi-omics atlas





Comparing similarity of *Ciona* vs *Xenopus*





Quantitative proteome dynamics across embryogenesis in a model chordate

Alexander N. Frese[1,2,*], Andrea Mariossi[1,2,*], Michael S. Levine[1,2,#], Martin Wühr[1,2,3,#]
1 Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ, USA
2 Department of Molecular Biology, Princeton University, Princeton, NJ, USA
3 Lead Contact
* equally contributed
# correspondence: msl2@princeton.edu, wuhr@princeton.edu

Summary

The evolution of gene expression programs underlying the development of vertebrates remains poorly characterized. Here, we present a comprehensive proteome atlas of the model chordate *Ciona*, covering eight developmental stages and ~7,000 translated genes, accompanied by a multi-omics analysis of co-evolution with the vertebrate *Xenopus*. Quantitative proteome comparisons argue against the widely held hourglass model, based solely on transcriptomic profiles, whereby peak conservation is observed during mid-developmental stages. Our analysis reveals maximal divergence at these stages, particularly gastrulation and neurulation. Together, our work provides a valuable resource for evaluating conservation and divergence of multi-omics profiles underlying the diversification of vertebrates.

Keywords: Proteome, Transcriptome, Development, Evolution, Chordates, Vertebrates, *Ciona*, *Xenopus*

Introduction

Embryonic development progresses through a series of cellular states, each defined by distinct changes in mRNA and protein levels. Optimal cellular functionality depends on precise control of gene expression and correct protein concentrations [1–3]. However, (1) accurately measuring protein concentrations and (2) understanding the mechanisms governing cellular proteostasis remain a significant challenge.

While transcriptomic studies often rely on mRNA levels to predict protein concentrations, the key determinants of cellular functionality and phenotype, numerous studies have reported weak correlations between the two, challenging their reliability as proxies for each other [4–7]. This disparity is influenced by the stochastic nature of mRNA transcription, translation, and degradation and becomes particularly pronounced for dynamic cellular transitions during embryogenesis [8–11]. Thus, mRNA levels are not necessarily predictive of protein concentrations, which prompts a shift towards applying more comprehensive proteome-wide analyses.

Proteomic methods provide an accurate measurement of protein abundance but have been historically limited by technical challenges [12]. Recent advancements in quantitative multiplexed mass spectrometry (MS) have significantly enhanced the sensitivity and precision of these measurements, expanding our capacity to map the cellular proteome in detail [13–18]. Applying these techniques to the study of vertebrate embryos still presents considerable challenges. The system's complexity, high cell numbers, and substantial yolk content, which affects detection of moderate and low abundance proteins, have limited the coverage and scope of these analyses [4,19–23]. Urochordates are the nearest extant relatives to vertebrates and share several morphological and genomic traits [24]. In particular, *Ciona* has numerous experimental advantages like small size, low cell number, stereotyped cell lineages, rapid and comparatively simple development with experimental tractable embryogenesis, and a compact genome that is not complicated by the gene duplication events accompanying the advent of the vertebrates [25]. Additionally, *Ciona* retains conservation of non-coding elements, macrosynteny, and microsynteny with chordates, making it an ideal model for studying the evolution of vertebrate developmental processes [26–31].

While *Ciona* lacks the complex specializations and innovations characteristic of vertebrates, it has nonetheless advanced our understanding of the morphogenesis of basic chordate tissues such as the muscles, heart and notochord, as well as the evolution of key vertebrate processes such as neural crest [32–40]. The assembly of the *Ciona* genome [25] represented a significant landmark that enabled a variety of transcriptomics, epigenomics, and single-cell studies [41–47]. Here, we extend these large-scale datasets through the use of quantitative proteomics methods.

The evolution of gene expression and its role in morphological innovations have been studied primarily by comparative transcriptomics [48–51]. These studies point towards a 'phylotypic period' in vertebrates, whereby gene expression is most similar across different species during mid-embryogenesis or pharyngula stage, the “hourglass” model [52]. However, comparisons with non-vertebrate chordates such as tunicates and cephalochordates are not entirely consistent with the hourglass [53]. This suggests potential divergent developmental pathways or an earlier onset of conservation as compared with vertebrates. For example, in amphioxus this conservation aligns with the earlier neurula stage [54]. In fact, extending comparative analysis to invertebrates, suggests an inverse hourglass model with increased conservation during early and late developmental stages rather than in the middle of development [55–57]. This model implies a bottleneck in developmental pathways, potentially influencing the emergence of species-specific traits. The effectiveness of these comparative analyses require careful consideration of phylogenetic distances, species diversity, embryonic stages, and gene sets compared [58]. Several studies stress limitations of simplistic pairwise comparisons, robust testing of null hypotheses, and the challenge in balancing phylogenetic distances, which can be too short among closely related species or too extensive when the comparisons are made between vertebrates and invertebrates or across multiple phyla [51,53,58].

A major limitation of the earlier studies is the reliance of transcriptome datasets to infer the dynamics of gene activities [59]. Recent reports suggest significant disparities in mRNA and protein levels [4–6,60]. In this study we re-examine similarity of embryos at various developmental stages with comparisons of both transcriptome and proteome datasets. Proteomic studies offer a novel perspective in cross-species comparisons by quantifying protein conservation patterns, which are the primary executors of most cellular functions [61].

Here, we use state-of-the-art proteomics to quantify proteins in unfertilized *Ciona* eggs and to track proteomic changes throughout embryogenesis, revealing that the embryonic proteome accounts for at least half of the genome's protein-coding capacity. We create a detailed genome-wide dataset that shows precise measurement of protein kinetics and their association to key developmental processes such as fertilization, maternal-to-zygotic transition (MZT), gastrulation, and the formation of larval tissues. Further, we integrated these data with corresponding transcriptome information and carried out inter-species comparisons between *Ciona* and *Xenopus laevis*, the African clawed frog. We discuss the implications of these studies with respect to the conservation and divergence of genetic activities during chordate evolution and reconsider the hourglass model of development.

Results and Discussion

Adapting proteomics for the analysis of *Ciona* eggs and embryos

Mass spectrometry-based proteomics (MS) is a versatile tool for studying a variety of biological processes, although new model systems often require method adaptations. Key areas needing optimization include sample preparation and the reference proteome. Analyzing eggs and early embryos is often challenging due to the high yolk content. For instance, in *Xenopus*, yolk constitutes ~90% of egg protein content, limiting the depth of proteomics analyzes [62]. Researchers usually remove yolk through centrifugation after lysis of eggs or embryos [63–65]. However, when we analyzed *Ciona* egg lysates via Coomassie-stained gels, we found no exceptionally dominant protein band (Figure S1A), allowing us to analyze *Ciona* samples by MS without yolk removal. Another concern in proteomics is the quality of the protein reference database. For widely used models such as humans, mice, or yeast, this is typically derived from the genome. However, the quality of the genome for non-canonical model organisms is often poor, thereby severely limiting the proteins that can be identified via MS. A better reference database can be generated based on mRNA-seq data [64,66]. Accordingly, we first evaluated the quality of the latest *Ciona* genome by benchmarking it against a genome-free protein reference database, which we generated from available RNA-seq datasets (Figure S1B) [39,67–69]. Upon comparison, the RNA-seq based reference database clearly outperformed Uniprot [70] and the previous genome annotations (KH-2013 and KY19) [71], but increased peptide coverage by only 5% compared to the most recent KY21 annotation (Figure S1C,D) [72]. We decided to accept the modest decrease in identified peptides for the ease of annotation offered by the genome assembly and proceeded to use the KY21 genome as our primary reference for the remainder of this study.

Further examination of peptides identified using our genome-free database revealed mis-annotated gene coding sequences, mis-positioned intercistronic regions, and discrepancies in selenoprotein sequences present in the KY21 proteome (Figure S1E) [73–75]. We believe that our analysis is a step forward in improving the accuracy and completeness of the *Ciona* genome annotation and the potential of the proteome atlas to refine *Ciona* gene models and protein coding sequences. Collectively, our data reveals that the latest assembled *Ciona* genome, combined with the characteristics of its eggs and embryos, is highly suitable for proteomics studies, and supports *Ciona*'s potential as a valuable model system for proteomics investigation.

Absolute protein abundance measurements in the unfertilized egg

The mature egg contains an array of maternal proteins required for fertilization, transition to zygotic transcription, and the early stages of embryogenesis [76–79]. Given that many of

these proteins remain unidentified, incorporating a proteomic approach was the logical next step. We estimate the absolute concentrations of proteins in the unfertilized egg using MS1 precursor intensity in a deep label-free analysis [64]. Altogether, we quantified the abundance of 6,102 proteins, after collapsing isoforms (Figure 1A, Table S1), thereby expanding the number of known proteins by an additional 5,058 entries compared to the previous proteomic investigation of the *Ciona* egg [80]. Nearly 90% of identified proteins are supported by at least two peptides, and the mean sequence coverage is 21% (Figure S1F).

As expected, the most abundant protein is Vitellogenin (yolk protein), followed by ATP synthase subunits, actin, and a 60S ribosomal subunit (Figure 1B) [81]. The analysis spans approximately eight orders of magnitude, covering 95 transcription factors (TFs) and 46 signaling molecules (SMs) (Figure 1C). The median protein concentration is 22 nM. In contrast, the median concentrations of TFs and SMs are lower, 5.4 nM and 3.5 nM, respectively. Most of them are distributed towards the lower end of the concentration curve, aligning with reports from other systems, where it has been noted that these molecules can exert significant biological effects even at low concentrations, particularly in driving dynamic cellular processes such as differentiation [82]. Among the identified TFs in the egg are known maternal factors such as Gata.a, Prd-B/Prdtun2, and Zeb (also known as Zinc Finger (C2H2)-33 or Ci-ZF266) [83]. Among the SMs, known maternal factors include β-Catenin, Eph.a/Eph1, Eph.b/Eph2, Raf/Raf1, Tll/Tolloid, Notch, and Numb [83–85]. The interaction of these known maternal deposits have been reported to be essential to establish the first distinct spatial domains of gene expression that launch the gene regulatory networks controlling embryogenesis [86]. Alongside these molecules, the proteomic landscape is characterized by an abundance of kinases and phosphatases, common regulatory components controlling the cell cycle and proliferation. Proteins indicative of posterior end markers (PEM), which include germline determinants and positional cues for the axial development of the embryo, are conspicuous components of the maternal proteome [87,88]. These findings suggest a preparatory state for fertilization and subsequent developmental cascades. Furthermore, in addition to Vitellogenin, we observe a notable enrichment of metabolic components, emphasizing the importance of energy and nutritional reserve components supplied by the egg for the early stages of development. These proteins ensure that *Ciona* embryos, which do not feed before metamorphosis, have the necessary resources for successful settlement.

We next asked whether different subunits within the same protein complex are found at expected stoichiometric ratios. To this end, we mapped the proteins identified in the egg to known stable complexes from the CORUM database (Figure 1D) [89]. We observed overall comparatively tight distributions of subunits in most macromolecular complexes, such as MCM (involved in genomic DNA replication) [90], CCT (playing a significant role in

protein folding in the eukaryotic cytosol) [91], the HAUS complex (essential for mitotic spindle assembly) [92], and Prefoldin (chaperone proteins regulating correct protein folding) [93]. For all the complexes for which we detect more than two subunits, the distribution is significantly different from the distribution of the entire dataset ($P < 0.01$, two-way ANOVA with Tukey's multiple-comparisons test) (Figure 1D).

Altogether, the proteomics of the unfertilized egg highlight intricate networks that anticipate subsequent developmental processes such as fertilization, spatial patterning, and hatching. The consistency of values obtained for different subunits of stoichiometric protein complexes corroborates the reliability of our data, providing a robust platform for future studies.

A high-quality multi-omics atlas of *Ciona* development

We next measured the dynamics of protein and mRNA abundances as the egg develops into a swimming tadpole. For this relative comparison analysis we combined accurate multiplexed proteome analysis (TMTproC) [14] with RNA-seq on matching samples at eight key developmental stages. These stages span early embryonic development and include the maternal/zygotic transition, gastrulation, neurulation, tail elongation, and hatching of swimming tadpoles (Figure 2A), thereby encompassing all of the important developmental processes. Moreover, the parallel sampling of both modalities facilitates a direct comparison between RNA and protein expression.

Using this framework, we detected 7,095 protein isoforms encoded by 7,057 genes (Figure 2A, Table S2), representing 38% of the protein-coding genes annotated in the latest *Ciona* genome assembly [72]. This accounts for approximately 50% of the expressed genes captured in RNA-seq analyses (Figure 2B, Table S3). This protein number is more than tenfold greater than that reported in an earlier study, which identified 695 proteins across three sampled stages using two-dimensional gel electrophoresis and MALDI-TOF/MS [81]. Our proteome marks a significant advancement in the quality of the UniProt database, which reports experimental evidence at the protein level (PE1) for less than 1 % (21 out of 17,311 records). We cover 55 % of the redundant UniProt entries, of which four had prior evidence at the PE1 level. Importantly, we confirmed protein products for an additional 9,415 entries previously undocumented at the protein level, categorized under evidence levels PE2–4 (Figure 2C). The new proteome dataset significantly expands the known proteomic landscape of *Ciona*.

Descriptive analysis of proteomic data and RNA-seq atlas

For MS data, we applied a 0.5 % false discovery rate (FDR) filter, integrating linear discriminant analysis (LDA) with a target-decoy strategy [94,95] (Figure S2A). We quantify a total of 62,471 peptides, the proteins with most identified peptides are Vitellogenin and Titin (Figure S2B). The median number of peptides per quantified protein is 5, with 84 %

of the proteome showing more than two peptides per protein (Figure S2B). The identified peptides correspond to 7,095 protein isoforms matching 7,057 unique proteins (Figure 1A). In 35 instances, the dataset enabled differentiation between 2 to 4 splice variants (Figure S2C). The poly(A) pulldown RNA-seq datasets cover an average of 10,727 ± 1,007 genes (mean ± s.d.), with high reproducibility of the biological replicates (Figure S3A,B,C). The number of detected genes steadily increases as development proceeds, reflecting an expanding gene expression repertoire (Figure S3D). However, post-zygotic genome activation (ZGA) at the 16-cell stage did not result in an increase in gene counts, likely due to the degradation of maternal mRNAs as previously observed in zebrafish development [96]. The distribution of expression levels (transcripts per million, TPM) initially exhibited a bimodal pattern with peaks at very low and higher levels. As embryonic development proceeded, this distribution evolved into a more normal distribution (Figure S3E). These observations are consistent with the transition of bimodal distributions seen for homogenous cell populations to a unimodal distribution for heterogeneous cell populations [97].

Temporal dynamics and tissue-specific patterns in the proteome atlas

In order to extend our analysis and systematically identify proteins that may influence differentiation programs, we categorized the proteins into eight distinct clusters based on their activity at various stages (Figure S4A) and performed gene ontology (GO) enrichment analysis on each gene cluster (Table S2). Cluster 1 genes exhibited the most stable dynamics, with proteins involved in translation, RNA processing, cell division, DNA organization, ribonucleoprotein complex formation, ribosome biogenesis, and transfer RNA (tRNA) activity. These are indicative of housekeeping functions. Cluster 2 genes, most abundant in unfertilized eggs, rapidly degrade following fertilization and are enriched for mRNA processing, single fertilization proteins, and small GTPase-mediated signal transduction, aligning with spindle assembly roles post-fertilization. They also have an abundance of maternal ribosomes preparing embryos for future development. Proteins in Cluster 3, abundant in both fertilized and unfertilized eggs but rapidly degrading before MZT, are linked to cell division and protein degradation, facilitating rapid embryonic development during the first four hours post-fertilization (hpf). Notably, the Gata4 TF is an early determinant of dorsal-ventral patterning and it makes sense that it is a constituent of Cluster 3 [86]. Cluster 4 proteins, peaking during gastrulation and neurulation, are associated with cell division, translation elongation, embryonic organ development, and chromatin modification. This reflects the shift from maternal to zygotic production, high translational activity, cell division, and the onset of tissue differentiation. Clusters 5 to 8 exhibit a monotonous growth pattern during MZT, gastrulation, neurulation, and tailbud stages. In later stages, the focus shifts to energy generation, transport, metabolic processes, and tissue morphogenesis. These clusters are enriched with cofactors, coenzymes involved in metabolism, and actin filament organization, correlating with

metabolic preparation for swimming tadpoles. Collectively, these analyses revealed proteome dynamics during development, mirroring various aspects of tissue differentiation and morphogenesis.

Next, we evaluated the utility of the proteome atlas as a tool to analyze the expression of tissue-specific marker genes, including those representing the major lineages/germ layers (Figure S4B). This revealed a series of staggered progression waves in protein expression across different tissue types. In line with existing literature [98], we observe that the onset of most tissue differentiation began with gastrulation at the 110-cell stage (epidermis, and endoderm). In the case of the notochord (SEC31b) and mesenchyme (Ci-Psl3), some markers emerge as early as the 16-cell stage, underscoring the unique aspects of *Ciona* embryogenesis where most cells are restricted to a single tissue fate by the start of gastrulation [99]. Markers of differentiating neurons associated with the dorsal and lateral regions of the brain such as Synaptotagmin 1 (syt) [100], Cel3/4/5 (also known as Etr-1, cel3.a) [101], and Rlbp1 (also known as Cralbp) [100] are also identified at relatively early stages of embryogenesis. For the muscle lineage, we observe multiple proteins expressed contemporaneously starting from the mid-tailbud II stage (Figure S4B) [38]. These examples highlight a developmental progression in protein expression patterns and how the proteome atlas effectively mirrors the establishment of definitive cellular phenotypes, in this case elongated muscles.

To further evaluate the utility of the proteome atlas, we explored aspects of temporal fate patterning, focusing on TFs and SMs that are critical for cell specialization during embryogenesis. The data cover approximately 40 % of all annotated TFs and ~60 % of all SMs, kinases and phosphatases (Figure S4C). Principal component analysis (PCA) shows a smooth transition from one stage to the next, with the first two principal components accounting for over 80 % of the proteome's variance. A striking 'salt and pepper' pattern emerged when overlaying transcriptional regulators across the proteome's development. The observed expression dynamics likely reflect a combination of tissue composition and protein accumulation, effectively separating early and late expression protein along a spatial developmental continuum (Figure S4D).

We also ranked protein changes across consecutive developmental stages to identify stage-specific proteins. This analysis highlights significant changes in protein abundance at three key stages: post-fertilization, the maternal-to-zygotic transition (MZT), and the onset of metamorphosis. Post-fertilization, the egg's proteome exhibits substantial alterations of proteins involved in calcium signaling, mitochondrial function, and translation. The MZT phase shows a surge in proteins related to organogenesis. As swimming tadpoles transition towards metamorphosis there is an increase in proteins associated with tail reabsorption. Examples include the TF Hox10 [102] (Figure S4E).

Quantitative mRNA-protein expression landscapes

Cellular protein concentrations are modulated via transcriptional and translational mechanisms [103]. By integrating transcriptomic and proteomic data from stage-specific embryos, we can explore the extent to which RNA signatures explain protein dynamics. First, we observe that protein and transcript expression vary significantly, spanning different orders of magnitude (Figure S5A). Moreover, consistent with existing literature [104,105], proteins encoded by low-abundance genes are underrepresented, indicating proteome coverage is not yet exhaustive (Figure S5B). We also notice strong variations in quantitative levels at each developmental stage, evident at both the protein and gene levels. There is little overlap in the rank order or even the identity of the most abundant proteins and mRNAs at any given stage (Figure S5C).

The overall correlation between the 7,021 mRNA and protein pairs is low, with a median Pearson correlation of -0.012 (Figure 2D), similar to previous studies (Table S5) [4–7]. This is partly due to the limitations of these methods in accounting for the orientation of vectors in multi-dimensional space. Furthermore, our approach assesses how mRNA and protein pairs change over the developmental timeline rather than a snapshot of a specific stage. Figure 2E illustrates an example of TF with high Pearson correlation between RNA-protein dynamics. Additionally, Figure S6 presents a selection of TFs known to play significant roles in the early development of *Ciona* [98].

Using k-means co-clustering of mRNA and protein pairs, we identified 5 distinct cluster dynamics (Figure 2F). We found that the genes involved in DNA replication/repair, centriole elongation/replication, rRNA processing, and protein localization to the nucleus have maternally loaded RNA and the most static protein dynamics. Metabolic processes broadly span all of the clusters, implying that metabolic processes are not categorized by a specific dynamic pattern. Axon development, heart development, and muscle filament sliding/contraction genes are expressed at the transcript and protein level during the tailbud and larval stages of development. These data suggest that the genes in the more dynamic clusters are preferentially associated with organogenesis while the genes in the less dynamic clusters tend to drive housekeeping or cell cycle functions (Figure 2G).

In summary, we profiled *Ciona*'s proteome and transcriptome across key developmental stages, resulting in an atlas of 7,021 protein-mRNA pairs, underscoring the complementary nature of mRNA and protein data in understanding cellular mechanisms. The dataset shows how mRNA and protein profiles can diverge and decouple due to translational regulation, demonstrating that transcriptional changes can be modified or overridden. This atlas, enriched with existing genomic and epigenomic data, provides a basis for further exploring RNA-protein dynamics during embryogenesis and systematically assessing adaptive expression of both RNAs and proteins.

Conserved and divergent features of the *Ciona* and *Xenopus* proteomes

Embryogenesis progresses through distinct stages, but it remains unclear if the regulatory mechanisms guiding these transitions are conserved across species. In particular, how well are the protein dynamics of orthologues conserved over significant evolutionary distances? Is there a conservation of protein abundances in relation to the levels of their corresponding mRNAs? With these questions in mind, we compare the proteome of *Ciona* development with that of a vertebrate. We focused on the African clawed frog *Xenopus laevis*, which is very attractive for proteomics analysis [4,5,63,64,106,107] resulting in one of the best characterized vertebrate proteomes throughout embryogenesis. *Xenopus* and *Ciona* diverged approximately 500-600 million years ago [108], providing a significant evolutionary distance for comparison (Figure 3A).

We applied k-means clustering to classify 3,350 one-to-one orthologous protein pairs into 5 distinct clusters, using the frog proteome time series data from Sonnet et al. [106] (Table S7), and we identified significant similarities in proteome dynamics between these two species (Figure 3B). More than half of the shared proteins are stably expressed in both species throughout development (blue cluster, Figure 3B,C). This cluster is enriched for proteins involved in DNA replication, spindle formation, and chromosome movements. Clusters that capture the activity of genes involved in rRNA processing, tRNA processing, and mRNA splicing via the spliceosome show an increase in expression throughout embryogenesis in both organisms. Genes involved in metabolic and catabolic processes also shared an increase in expression throughout embryogenesis in both organisms, however with a more pronounced increase in *Ciona* (Figure 3B,C). Basement membrane assembly and muscle differentiation genes have similarly high expression throughout embryogenesis in both organisms (Figure 3B,C), including those known to have roles in late development such as Lamα5 and Smyd1 [109,110]. These results highlight the similarities of orthologous protein dynamics during the development of these highly divergent species.

We next shifted our focus to the dynamics of orthologous TFs during development. We looked at the relative expression of these proteins in swimming tadpoles over their relative expression levels in the eggs of each organism (Figure 3D). Overall, TFs that showed the most pronounced changes in *Ciona* tended to also increase their expression in *Xenopus*. Notably, Smyd1, Tfap2-r.b, and Arid3, which are known transcriptional regulators of muscle [83,99,109], ectoderm/neural crest development [99], and chromatin remodeling [83], respectively, exhibited similar patterns of expression in both species (Figure 3E). Importantly, we observed TFs that showed different expression dynamics between the two species. The Y-box binding protein, Ybx, exhibited inverse behavior between the two organisms. In *Ciona*, Ybx mRNA [83] and protein are maternally deposited, whereas in *Xenopus*, it is strictly expressed after fertilization and plays a crucial role in muscle and

vascular development [111,112]. Ybx is a highly conserved protein involved in transcriptional regulation and is a component of messenger ribonucleoprotein complexes [113]. Notably, in zebrafish, both mRNA and protein are maternally deposited and are essential for activating maternal Nodal signaling [114]. Understanding the underlying reasons for the differential behavior of Ybx in *Ciona* and *Xenopus* requires further investigation. Despite many similarities, there are numerous differences that probably reflect species-specific functions.

We have identified conserved and unique protein dynamics across *Ciona* and *Xenopus* through comparison for more than ~3,000 orthologous proteins. Overall, we find strikingly high conservation of protein dynamics between the two organisms even though they are separated by ~600 million years of evolution. This analysis therefore presents an exciting opportunity to shed light on conserved regulatory processes in chordate development.

An inverse hourglass model for proteome evolution between *Ciona* and *Xenopus*

Cross-species embryonic development is typically aligned at the transcriptome level [48–51,53,55]. We therefore used developmental proteomes to establish stage correspondences between *Ciona* and *Xenopus* species throughout embryogenesis. We identified 7,636 one-to-one orthologs at the gene level (Table S8, S9) [53,115]. At the proteome level, we complemented the time series data from Sonnet et al. [106] (comprising 3,350 one-to-one orthologs, Table S7) by using an additional independent proteome time series from Itallie et al. [107], which included 5,376 one-to-one protein pairs (Table S10).

Starting at the transcriptome level, we observed that 60 % of the orthologs are commonly expressed in both species during the early stages, before gastrulation. This shared expression decreased to 55 % during the mid-developmental transition (gastrulation and neurulation) and reached 50 % in the late phase (tailbud, larva, juveniles), with the highest proportion detected in early development (Figure S7A). We next sought to determine how changes in gene expression mark different developmental stages. We found that gene expression patterns between the two species do not show abrupt changes between stages but rather change gradually and continuously throughout embryonic development. This indicates a single continuum of differentiation, rather than distinct subsets, with smooth transitions across consecutive stages. The greatest transcriptomic similarity occurs at hatching, when excluding *Ciona* metamorphosis stages (Figure S7B, Table S8, S9).

Comparison of the shared proteome reveals striking differences with the analysis of transcriptomes. The proteomes exhibit distinct phases of shared expression, one early and one late, which are divided by a sharp mid-developmental transition (Figure 4A). The

two species showed increasing proteome divergence with each other as they undergo neurulation. This pattern is consistent with an inverse hourglass model with the highest divergence during gastrulation and neurulation (Figure 4A, S8, S9). The early developmental phase may be subject to more functional constraints and less refractory to change, while the larval stage, crucial for forming a swimming tadpole in both species, shows overlapping protein functions and similar phenotypes.

The proteogenomic patterns revealed by this study remain consistent across various types of comparisons and are robust against different parameters used in constructing the correlation matrix (Pearson (r), Spearman (ρ), Cosine) (Figure S7C), and potential stage sampling biases (Figure S7B, S8, S9). For example, extending the *Ciona* time series from 8 to 20 stages (from egg to juveniles,Table S8) [53] and the *Xenopus* series to 17 distinct time points (from egg to swimming and feeding tadpoles, Table S9) [53,115] again showed maximal transcriptomic similarity at hatching (Figure S7B). Similarly, when analyzing a different proteome dataset for inter-species comparison [106,107], the dual-phase pattern is still evident. This *Xenopus* time series included two additional time points beyond those previously analyzed, effectively spanning the first 120 (hpf) of embryogenesis (Figure S8, S9).

To map stage transitions in the embryonic timeline, we classified stages with similar morphological events in both species, including cleavage, blastula formation, gastrulation, neurulation, tailbud and swimming larva. We determined the highest correlation points for each stage using both transcriptome and proteome data. By connecting these points (shown as a black line in Figure 4A), we assessed whether mRNA or protein expression better matched the known phenotypic stages. This analysis revealed that protein correlations more closely followed the established mapping of equivalent developmental stages (Figure 4A), indicating that proteomes provide a more accurate representation of embryonic stages compared to transcriptomes (Figure 4A).

Our results are consistent with an inverse hourglass model for protein conservation whereby protein activity is most divergent at mid-developmental stages and the molecular components that comprise early and late embryogenesis are more conserved (Figure 4B, S8, S9). We hypothesize that this divergence might represent the distinct mechanisms of gastrulation and neurulation in the two species. In *Ciona*, gastrulation takes place via a cup-shaped gastrula driven by invagination of the endoderm, whereas in *Xenopus*, convergent extension of mesoderm and epidermal epiboly play important roles. Most importantly, *Ciona* differs temporally from its vertebrate cousin by specifying its axis at the neurula stage, rather than at gastrulation [116]. In frog development, Stage 9 signifies the beginning of gastrulation. Maternal deposits and translation play a significant role in shaping early embryogenesis. It's likely that similar proteins and pathways are conserved

across species for timing and initiating this crucial phase, as evidenced by the high conservation observed in the proteome during this period. However, as gastrulation begins, the dynamics of embryogenesis shift, the mechanisms underlying this process start to differ significantly among species, setting the stage for the zygotic genome to take over gradually. This divergence is reflected in low or negligible signals of conservation observed in the blastula stage transcriptome among different species. New genes need to be expressed becoming more diverse and species-specific to evolutionary adaptations. The highest similarity between the species proteomes is observed at the larval stage, likely due to shared structural and ecological needs of swimming larvae.

Throughout all stages, we noticed that the proteome correlations were always higher than the transcriptome correlations (Figure 4A,B). This suggests that protein behavior is more evolutionarily conserved over time than mRNA behavior, likely because proteins are directly responsible for carrying out functions [61,117]. It is possible that post-transcriptional mechanisms, such as variations in translation or protein degradation rates, have evolved to offset differences in mRNA dynamics.

The proteome closely reflects an organism's physical traits, offering a more accurate measure of developmental and evolutionary differences within chordates. This underscores the importance of proteomics for evolutionary studies across species. However, previous gene ontology analysis linked variations in the transcriptome to specific biological functions. Regulatory mechanisms, including post-transcriptional, translational, and protein-degradation processes, appear to compensate for mRNA levels dissimilarity, aligning protein abundances with evolutionarily preferred levels [61,118,119]. This suggests a synergy between genetic drift and regulatory mechanisms in chordate evolution, focusing on key regulatory genes essential for developmental processes and post-translational regulation. Our study highlights the significance of the simple chordate *Ciona* in understanding chordate development, proving its worth as a model for future comparative research, particularly in studying proteome stability and its evolutionary implications.

Limitations of the study

Our analysis is subject to certain limitations. The proteome atlas identifies ~15,000 expressed genes and ~7,000 proteins. Nearly 40% of the proteome remains uncharacterized, likely missing proteins expressed during later stages, such as metamorphosis, which our embryo-centric analysis does not cover. It is also possible that a number of RNAs and proteins are exclusively expressed in juveniles or adults, representing another gap yet to be addressed. Additionally, the detection of certain proteins is challenged by their incompatibility with standard proteomics methods, including precipitation and digestion steps, or due to their low abundance [13,120]. Our

analysis, based on whole embryos, inherently reflects average protein levels across diverse cell types. Our study includes the analysis of different stages of *Ciona* embryogenesis. We would like to point out that there is a comparative under-representation of metamorphosis and juvenile stages.

Acknowledgments

We thank all members of the Wühr laboratory for helpful discussion, particularly to Felix Keber, Edward Cruz and Alex Johnson. We also thank members of the Levine laboratory, especially Pavan Choppakatla for insightful inputs. We thank Nicholas Treen for providing unfertilized *Ciona* eggs and Lillia Ryazanova for assistance in *Ciona* protein sample preparation. We thank Elizabeth Van Itallie for sharing Xenopus proteomics data.

This study was funded by NIH grant (T32GM007388) to Princeton University, NIH grant (NS076542) to M.S.L., NIH grant (R35GM128813) to MW, Eric and Wendy Schmidt Transformative Technology Fund to M.W., Diekman collaboration fund to M.S.L. and M.W. and Princeton Catalysis Initiative to M.S..L and M.W..

Author contributions

Conceptualization, A.N.F., A.M., M.S.L. and M.W.; Methodology, A.N.F., A.M., M.S.L. and M.W.; Investigation, A.N.F., A.M.; Writing – Original Draft, A.N.F., A.M., M.S.L. and M.W.; Writing – Review & Editing, A.M., M.S.L. and M.W.; Funding Acquisition, M.S.L. and M.W.; Supervision, M.S.L. and M.W.

Declaration of interests

The authors declare no competing interests.

Main Figure titles and legends

Figure 1 Absolute proteomics of the *Ciona* egg.

A, Schematic of label-free proteomics utilized to determine absolute protein concentrations. Unfertilized *Ciona* eggs were lysed, and human proteins of known concentrations (UPS2) were added to the lysate as a reference standard. Following normalization as outlined in the materials and methods, we detect ~195,000 peptides and estimate protein concentrations for ~6,000 proteins.

B, Table of selected proteins in the unfertilized egg including the top 5 most abundant and transcription factors important to embryonic development.

C, Histogram of all quantified proteins in the *Ciona* egg (gray) with superimposed kernel density estimates (KDE) of transcription factors (TFs - red) and signaling molecules (SMs

- blue). Both TFs and SMs follow a distribution similar to the global egg proteome (black) but with a lower median concentration. The complete data is provided in Table S1.
D) Stoichiometries of protein complexes. Concentrations of subunits from a shared protein complex display comparable values and show typically a statistically different distribution than the entire proteome (*P < 0.01, two-way ANOVA with Tukey's multiple-comparisons test).

Figure 2 Proteome and RNA analyses during *Ciona* embryogenesis

A, Overview of the transcriptome and proteome time-course experiments. Staged embryos were collected at eight developmental stages, beginning with unfertilized egg (unfE), fertilized egg (fertE), 16-cell stage (cell-16), initial gastrula (iniG), late neurula (latN), middle tailbud II (midTII), late tailbud II (latTII), and hatching tadpole (larva). Each stage is represented by a unique color code, and abbreviation, both are kept consistent throughout the figures. Time indicates hours post-fertilization (hpf).
B, Number and overlap of identified protein-coding genes in the transcriptome and proteome datasets.
C, Donut plot with the percentage of protein evidence categories from UniProt that are identified at the proteome level (9,419 entries). Evidence level: (1) protein evidence; (2) transcript evidence; (3) homology; (4) predicted.
D, Histogram of Pearson correlations between RNA and corresponding protein dynamics throughout *Ciona* development (gray). The lines represent kernel density estimates (KDE) for all genes (black), transcription factors (red), and signaling molecules (blue). Notably, mRNA dynamics correlate poorly with protein dynamics. n = 7021 pairs.
E, Example of high Pearson correlation between RNA and protein dynamics for the TF Hox10.
F, K-means clustering used to classify RNA (left) and protein (right) dynamics for each gene during *Ciona* development. The thickness of the lines scales with the number of proteins represented in each cluster, as indicated in the legend.
G, GO term analysis used to discern the functional relevance of each of the clusters (indicated by matching colors) identified in F.

Figure 3: Comparison of development between chordate and vertebrate

A, Experimental design of the inter-species comparative developmental transcriptome and proteome time courses. Full circles highlight stages of development sampled for RNA-seq and proteomics.
B, K-means co-clustering of the dynamics of orthologs (3,325) between *Ciona* and *Xenopus* development. The thickness of the lines scales with the number of proteins represented in each cluster. The number of proteins in each cluster are quantified in the legend. *Xenopus* proteome time series from Sonnett et al. [106].

C, GO term analysis identifying the functional significance of each of the clusters from B. The color of the clusters in B is kept consistent.
D, The log2 fold change (FC) protein correlation between *Ciona* and *Xenopus* TFs. Here, FC is defined as the ratio of relative protein abundance in the larva stage compared to the egg. Most TFs show similar behavior with the notable exception of Ybx.
E, Relative protein dynamics of TFs Ybx, Smyd1, Tfap2-r.b, Arid3, and E2f4/5. Each exhibit large fold changes in both organisms. Colors are preserved in these five proteins from the plotting in D. These TFs are canonically important for organism development by regulating transcriptional activation during the cell cycle, early muscle development, ectoderm development, gene activation through chromatin remodeling, and Nodal signaling respectively. Ybx exhibits signs of being maternally deposited in Ciona, but not in *Xenopus*, suggesting functional evolutionary divergence of this ortholog from chordate to vertebrate. *Xenopus* illustrations © Natalya Zahn (2022)

Figure 4 The protein anti-hourglass model
A, Similarity heatmaps showing Pearson similarity between the two species for each investigated time point. Developmental stages are color-coded as defined in Figure 3A. The black line follows the highest correlation of the Xenopus time-point for each *Ciona* stage (n= 3,350, *Xenopus* transcriptome from Hu et al. [53], Session et al. [115], *Xenopus* proteome from Sonnett et al. [106]).
B, Temporal divergence of gene and protein expression from *Xenopus* embryogenesis to each *Ciona* stage. Maximal similarity is represented by the smallest distance from the center line, revealing a nested hourglass model in which the proteome exhibits more evident bottlenecks at early and later stages. Gray boxes outline these periods of minimal divergence. Regardless of stage, proteins show higher similarity between the two species' developmental mapping than RNA-seq, suggesting that protein dynamics are evolutionarily more conserved than mRNA dynamics (n= 3,350, *Xenopus* transcriptome from Hu et al. [53], Session et al. [115], *Xenopus* proteome from Sonnett et al. [106]).

Supplemental Tables Titles

Table S1. *Ciona* absolute protein abundance in unfertilized egg, related to Figure 1.
Table S2. *Ciona* relative protein abundance time series, related to Figure 2.
Table S3. *Ciona* TPM RNA-seq time series, related to Figure 2.
Table S4. *Ciona* RNA-seq alignment statistics, related to Figure 2.
Table S5. *Ciona* relative mRNA-protein dynamics, related to Figure 2.
Table S6. *Ciona-Xenopus* one-to-one orthologs, related to Figure 3.
Table S7. *Ciona-Xenopus* protein dynamics from Sonnett et al. 2018, related to Figure 3.
Table S8. *Ciona* TPM RNA-seq time series from Hu et al. 2017, related to Figure 4.
Table S9. *Xenopus* TPM RNA-seq time series from Session et al. 2016 and Hu et al. 2017, related to Figure 4.

608 Table S10. *Ciona-Xenopus* protein dynamics from Itallie et al. 2021, related to Figure 4.

STAR★Methods

RESOURCE AVAILABILITY

Lead contact

Further information and requests should be directed to the lead contact, Martin Wühr (wuhr@princeton.edu).

Materials availability

Materials generated for this study are available on request from Martin Wühr (wuhr@princeton.edu).

Data and code availability

- Data: The raw data associated with the RNA-seq experiments and gene expression matrices are available in GEO under the accession number: GSE237005. The mass spectrometry experiments presented in this study have been deposited to the ProteomeXchange Consortium (http://www.proteomexchange.org/). Embryo developmental proteome (deposited via the PRIDE partner repository) with accession number: PXD043619. Genome annotation files, transcription factor and signaling molecules databases used for RNA-seq and proteomics analyses, alignment files used in orthology assignment, and additional files are publicly available on GitHub (https://github.com/andreamariossi/proteome_ciona)
- Code: All code to reproduce this study is publicly available on GitHub (https://github.com/andreamariossi/proteome_ciona).
- Other items: Additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Ciona handling and embryos collection

Wild type adult hermaphrodite *Ciona robusta* (formerly known as *Ciona intestinalis* Type A) [121] were obtained from M-Rep located in San Diego, CA and maintained in artificial seawater (Instant Ocean) at 18°C, under continuous illumination. Dechorionation and in vitro fertilization procedures were conducted following the protocol described in [122]. For each time point in the time series, embryos were staged and collected according to [123] at approximately 18°C and a total of 150 embryos were placed in Trizol for RNA extraction, while approximately 3,000 embryos were rapidly frozen in liquid nitrogen for protein TMTproC sample preparation. All samples were then stored at -80 °C until further use.

For absolute mass spectrometry analysis, approximately 5,000 unfertilized dechorionated eggs were directly snap-frozen.

METHOD DETAILS

SNP prevalence between *Ciona* batches

One concern is the presence of single nucleotide polymorphisms (SNPs), a characteristic feature of ascidian evolution [73,124], which can cause protein sequence polymorphisms and lead to incorrect peptide inference during the processing of MS data. We evaluated the potential influence of SNPs on peptide quantification accuracy. We obtained bulk RNA-seq data from two batches of 16-cell Ciona embryos. Each batch was assembled via Trinity, then translated into protein reference databases with the mass spec protein reference tool (https://kirschner.med.harvard.edu/tools/mz_ref_db.html) [64]. We reciprocally BLASTed each database against the other and found 16,037 shared proteins. These shared proteins were trypsin digested in silico. 98.8 % of the resulting peptides were identical between these batches while only 1.2 % were wholly unique to one batch or the other indicating minimal influence of intra-specific genetic variability on peptide recognition.

Generating protein reference database

The protein reference database, a FASTA file containing all potential proteins from the species under study, was used to generate *in silico* tryptic peptides and reference MS/MS spectra for peptide identification. 1,222,451,669 *Ciona* bulk RNA-seq reads from numerous studies [39,67–69] were assembled *de novo* via Trinity (version 2.11) into 2,328,005 transcripts [125]. The 55,974 transcripts making up the KH *Ciona* transcriptome (KHNCBI.Transcript.2018.fasta, retrieved from ANISEED) [126] were integrated alongside our de novo transcripts. The transcripts were cleaned and trimmed via SeqClean (http://compbio.dfci.harvard.edu/tgi/software/), then masked for common repeat motifs via RepeatMasker (version 4.1) [127]. The masked transcripts were clustered via TGICL (version 2.1) and assembled via CAP3 [128,129]. The resulting contigs and singletons were searched against a database of model organism containing human (*Homo sapiens*), Red junglefowl (*Gallus gallus*), Western clawed frog (*Xenopus tropicalis*), zebrafish (*Danio rerio*), Florida lancelet (*Branchiostoma floridae*), Pacific purple sea urchin (*Strongylocentrotus purpuratus*), and urochordate (*Ciona robusta*) using BLASTX (version 2.10.1) [130]. The BLASTX report was parsed and the transcripts were translated into proteins. The translated proteins were processed to remove redundancies with a CD-HIT (version 4.8.1) threshold of 95 % [131,132].

Proteomics sample preparation

Samples were prepared by lysing frozen embryos in lysis buffer (50 mM HEPES pH 7.2, 2% SDS, and 1x protease in artificial saltwater) followed by clarification via centrifugation. Lysates were diluted to 2 ug/µL with 100 mM HEPES (pH 7.2). DTT was added to a concentration of 5 mM and samples incubated for 20 mins at 60 °C. After cooling to RT, N-ethylmaleimide (NEM) was added to a concentration of 20 mM and samples incubated for 20 mins at RT. 10 mM DTT was added and samples incubated for 10 mins at RT to quench NEM. To 200 µL of each sample were brought up to 2 mL with 800 µL MeOH, 400 µL chloroform, and 600 µL water. Samples were centrifuged at 20,000 g for 2 minutes at RT. Upper layer was discarded and 600 µL MeOH was added. Samples were centrifuged at 20,000 g for 2 minutes at RT. Supernatant was discarded and 500 µL MeOH was added [133]. Samples were centrifuged at 20,000 g for 2 minutes at RT. Supernatant was discarded and the pellet was air dried. Pellet was resuspended in 6 M GuaCl, 10 mM EPPS pH 8.5 to ~5 mg/mL.

For the label-free samples, UPS2 standards (Sigma-Aldrich) were added to a final concentration of 27 ng/µL in the 450 µg protein samples. Samples were diluted with 10 mM EPPS pH 8.5 to 2 M guanidine hydrochloride. Samples were digested overnight at RT in LysC (Wako) at a concentration of 20 ng/µL. Samples were further diluted with 10 mM EPPS pH 8.5 to 0.5 M guanidine hydrochloride. 20 ng/µL LysC and 10 ng/µL Trypsin (Promega) were added to each sample and incubated for 16 hours at 37 °C. Peptide supernatant was cleared by ultracentrifugation at 100,000 g for 1 hour at 4 °C (Beckman Coulter, 343775), then vacuum-dried overnight.

For TMTpro-labeling, samples were digested with LysC and Trypsin as above, then resuspended in 200 mM EPPS pH 8.0. pre-mixed TMTpro tags (8-plex Thermo Fisher Scientific 20 µg/µL in dry acetonitrile stored at -80 °C) at a 5 µg TMTpro: 1 µg peptide ratio. To cover the eight developmental time series samples, tags are as follows: 126 - unfertilized egg; 128C – fertilized egg; 129N – 16-cell; 130C – 110-cell; 131N – late neurula; 131C – mid tailbud II; 133C – late tailbud II; 134N – larva. Samples were incubated for 2 hours at RT. Reactions were quenched by addition of hydroxylamine (Sigma, HPLC grade) to a final concentration of 0.5 % for 30 minutes at RT. Samples were pooled into a single tube, cleared by ultracentrifugation at 100,000 g for 1 hour at 4 °C (Beckman Coulter, 343775), then and vacuum-dried overnight.

For either label-free or TMTpro-labeled, samples were resuspended with 10 mM ammonium bicarbonate (pH 8.0) with 5 % acetonitrile to 1 µg/µL. Samples were separated by medium pH reverse phase HPLC (Zorbax 300Extend C18, 4.6 x 250 mm column) into 96 fractions [14,134]. The fractions were then pooled into 24 fractions [135], dried, and resuspended in HPLC grade water. Samples were then desalted via homemade stage tips with C18 material (Empore) and resuspended to 1 µg/µL in 1 % formic acid [136].

Drawings

Ciona schematics are adapted from FABA (FABA Four-dimensional Ascidian Body Atlas) [123] and *Xenopus* illustrations from Xenbase (www.xenbase.org RRID:SCR_003280) and Natalya Zahn [137]. Source icons with BioRender.com.

QUANTIFICATION AND STATISTICAL ANALYSIS

Proteomics analysis

Approximately 1 µg per sample was analyzed by LC-MS, as previously described [134]. LC-MS experiments were analyzed on an nLC-1200 HPLC (Thermo Fisher Scientific) coupled to an Orbitrap Fusion Lumos MS (Thermo Fisher Scientific). Peptides were separated on an Aurora Series emitter column (25 cm × 75 µm ID, 1.6 µm C18) (Ionopticks), held at 60 °C during separation by an in-house built column oven. Separation was achieved by applying a 12 % to 35 % acetonitrile gradient in 0.125 % formic acid and 2 % DMSO over 90 minutes for fractionated samples. Electrospray ionization was enabled by applying a voltage of 2.6 kV through a MicroTee at the inlet of the microcapillary column. For the label-free samples, we used the Orbitrap Fusion Lumos with the label-free method with data-dependent acquisition (DDA) previously described [64]. For the TMTpro samples, we used the Orbitrap Fusion Lumos with the TMTproC method previously described [14].

Mass spectrometry data analysis was performed essentially as previously described [106] with the following modifications. The raw MS files were analyzed using the GFY software licensed through Harvard University. MS2 spectra assignment was performed using the Sequest algorithm [138] by searching the data against either our reference protein dataset described above, the KY21 *Ciona* proteome [72], or the Uniprot *Ciona* proteome [70].

For label-free analysis, these proteomes were merged with the UPS2 proteomics standards FASTA file (Sigma-Aldrich) along with common contaminants. Peptides that matched multiple proteins were assigned to the proteins with the greatest number of unique peptides. To control for peptide false discovery rate, target-decoy search strategy was used where reverse sequences were searched in parallel with forward sequences [94]. Filtering was performed using a linear discriminant analysis (LDA) that accounts for parameters from Sequest's database search output, such as XCorr, deltaCorr, missed cleavages, charge state, peptide length, and the fraction of matched ions was also implemented to distinguish genuine peptide spectral matches (PSMs) from reverse hits. The data were then filtered to 0.5 % FDR on the peptide level and 1 % FDR on the protein level [95,139].

Absolute protein concentration estimates in unfertilized egg

Protein concentration in the label-free egg sample was calculated by building a standard curve of MS signal to UPS2 standard concentration. The UPS2 known standard concentrations were obtained from Sigma Aldrich and concentrations were converted to log space. The MS signal area was also converted to log space and Thiel regression was performed to obtain a standard curve. Signal area was then converted to concentration and scaled to a total protein concentration of 2 mM. A cutoff of 0.01 μM was applied for low concentration protein. Information on known protein complexes was obtained from the CORUM Protein Complexes dataset [89]. A two-way ANOVA, followed by a post-hoc Tukey HSD test, was applied to assess the distribution of protein concentrations.

Proteomics data processing

GFY output tables for TMTcPro MS were filtered for human protein contaminants, reversed sequences and proteins which were only identified based on modified peptides as previously described [14].
Annotations and classifications of transcription factors, signaling molecules, kinases, and phosphatases are based on data merged from the Ghost website [140] and [141]. The proportional coverage of these families within our dataset was determined by counting the number of members that could be identified at the protein level.

K-means clustering was performed using the kmeans function in R with nstart = 100. The number of clusters was selected to 8 to capture overall protein dynamics. Further cluster increases did not reveal new cluster dynamics. GO enrichment analyses were used to assign categories to each cluster using gProfiler [142].

Principal component analysis (PCA) and was performed in R with prcomp function from the stats package. Annotations for families of transcription factors, signaling molecules, kinases, and phosphatases were then overlaid on the graphs.

For the calculation of cumulative abundance, proteins and genes were initially ranked from highest to lowest. The total expressed as a percentage is plotted against their rank order. The names or identifiers of the seven most abundant transcripts or proteins (rank 1 to 7) are listed in descending order for the respective stage.

To measure the similarity between the proteome and transcriptome datasets, Pearson’s correlation coefficient (r), Spearman's rank correlation coefficient (ρ), and Cosine distance were calculated for each individual gene-protein pair across all stages. These coefficients were then plotted as histogram distributions.

Ciona and *Xenopus* protein orthologs

Reciprocal protein-protein BLAST (RHB) (BLASTP, version 2.10.1) was used to identify orthologs between *Ciona* and *Xenopus* [130]. *Ciona* and *Xenopus* alternated as query and reference. For each BLASTP, the max target sequence was set to 1, e-value threshold was set to 0.01, and the matrix set to BLOSUM45. The query ID, reference ID, e-value, and bit score were logged for each match. “best-match” protein orthologs between *Ciona* and *Xenopus* based on the criteria of (1) lowest e-value and (2) highest bit score. Only proteins confirmed in both directions as “best-match” were used in the cross-species proteomic analysis (Table S6).

Comparative proteomics

The extent of conservation or divergence in protein expression among chordates and vertebrates was assessed by comparing the proteome of *Xenopus laevis* with that of *Ciona*. Two independent frog time series were reanalyzed: one comprising 8 time points from Sonnet et al. [106] (Table S7) and another with 10 time points from Itallie et al. [107] (Table S10). These series collectively cover frog embryogenesis comprehensively, overlapping at three stages (St1, St12, and St30). All three datasets were first subjected to median-based normalization. Then, the dynamics of proteins in each dataset were scaled to sum to 1 across the time series, allowing for comparison of expression across species. Correlation coefficients, including Pearson (r), and Spearman (ρ), were calculated using pairs of orthologs. These orthologs were identified based on RHB methods as explained earlier, for all pairwise combinations of developmental stages. To co-cluster the *Ciona*-*Xenopus* proteomes across developmental stages (data from Sonnet et al. [106]), we used k-means clustering. Each cluster was then assigned to a functional category, based on its overall gene expression and GO enrichment profiles.

RNA sequencing

For each of the eight embryonic stages, a total of 150 embryos were collected and stored at -80 °C in Trizol (Thermo Fisher Scientific). We prepared two biological replicates, one replicate consisted of embryos from the same *in vitro* fertilization batch for proteomic analysis. The other replicate was collected from an independent developmental time course. Total RNA was isolated using the Clean and Concentrator Zymo kit (Zymo), with genomic DNA (gDNA) removal achieved through on-column treatment with Turbo DNase (Invitrogen) at room temperature for 10 minutes. The resulting RNA was re-suspended in 15 µl of DEPC-treated water and quantified using a NanodropTM and Qubit (Thermo Fisher Scientific), while its quality was assessed using a Bioanalyzer 2100 (Agilent Technologies). The RNA integrity number (RIN) values ranged between 8 and 10. cDNA libraries were prepared using the PrepX RNA-seq directional protocol (Takara Bio) following the manufacturer's instructions and utilizing an Apollo 324 robot. For mRNA enrichment and separation from rRNA, the oligo dT-based mRNA isolation kit (Takara

Bio) was employed. The libraries were sequenced on the NovaSeq platform (Illumina) at the Genomics Core Facility at Princeton University with a depth of 20-40 million paired-end strand-specific reads.

Quality assessment of raw and trimmed 61-bp paired reads was performed with FastQC (version 0.12.0). Trimgalore (version 0.6.10) was used to trim the raw RNA-seq reads, removing adapters and primer contamination and poor quality base call (Q < 25). Reads shorter than 30 nt after trimming were discarded. The trimmed RNA-seq reads were then mapped to the KY21 transcriptome using Salmon (v0.42.4, with parameters --libType A, --seqBias, --gcBias, --validateMappings) [143]. Details about alignment quality are given in Table S4. mRNA quantities are presented as transcripts per million (TPM), with a cutoff of 2 TPM as the lower limit for detection across all samples. This cutoff was determined based on the inspection of distribution density plot and corroborated by known markers visualized from *in situ* hybridization chain reaction (HCR) studies [144] at the 16-cell stage, which is when the newly zygotic genes are activated. For each stage, RNA data from biologically independent experiments were pooled to estimate average gene expression.

For *Ciona*, the extended time-series RNA-seq data was obtained from Hu et al. [53] (Table S8). For *Xenopus leavis*, data was sourced from Session et al [115] and Hu et al [53] (Table S9). Gene expression for each species was estimated using Salmon [145], with KY21 annotation for *Ciona* and *Xenbase* X. laevis v10.1 annotation for frog. Gene-level expression was obtained by summing up TPMs from all transcript isoforms per gene using tximport R package [143]. For each stage, RNA data from biologically independent experiments were pooled to estimate average gene expression. A gene was considered expressed if it had a TPM ≥ 2.

To compare gene expression across embryonic stages between the two species, we utilized orthologs, as identified by reciprocal best hits (RBHs). To normalize the data for distinct expression levels and mitigate the impact of highly expressed genes, we applied quantile normalization using the preprocessCore package from Bioconductor [146]. We employed several metrics to estimate gene expression divergence between the two species, including Pearson (r) and Spearman (ρ) correlations, and Cosine similarity.

Gene set enrichment analysis

Gene Ontology (GO) term enrichment analyses were conducted using the gProfiler and topGo functional annotation tools [142,147]. For each cluster, genes were analyzed against a background list comprising all genes expressed across all time points. Enriched GO terms were identified in the categories of ‘molecular function’, ‘cellular component’, and ‘biological process’. A Benjamini-corrected P-value threshold of 0.01 was applied to determine significant enrichment.

Bibliography

1. Hausser, J., Mayo, A., Keren, L., and Alon, U. (2019). Central dogma rates and the trade-off between precision and economy in gene expression. Nat. Commun. 10, 68. 10.1038/s41467-018-07391-8.

2. Li, G.-W., Burkhardt, D., Gross, C., and Weissman, J.S. (2014). Quantifying Absolute Protein Synthesis Rates Reveals Principles Underlying Allocation of Cellular Resources. Cell 157, 624–635. 10.1016/j.cell.2014.02.033.

3. Teixeira, F.K., and Lehmann, R. (2019). Translational Control during Developmental Transitions. Cold Spring Harb. Perspect. Biol. 11, a032987. 10.1101/cshperspect.a032987.

4. Peshkin, L., Wühr, M., Pearl, E., Haas, W., Freeman, R.M., Gerhart, J.C., Klein, A.M., Horb, M., Gygi, S.P., and Kirschner, M.W. (2015). On the Relationship of Protein and mRNA Dynamics in Vertebrate Embryonic Development. Dev. Cell 35, 383–394. 10.1016/j.devcel.2015.10.010.

5. Smits, A.H., Lindeboom, R.G.H., Perino, M., Heeringen, S.J. van, Veenstra, G.J.C., and Vermeulen, M. (2014). Global absolute quantification reveals tight regulation of protein expression in single Xenopus eggs. Nucleic Acids Res. 42, 9880–9891. 10.1093/nar/gku661.

6. Vogel, C., and Marcotte, E.M. (2012). Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nat. Rev. Genet. 13, 227–232. 10.1038/nrg3185.

7. Wegler, C., Ölander, M., Wiśniewski, J.R., Lundquist, P., Zettl, K., Åsberg, A., Hjelmesæth, J., Andersson, T.B., and Artursson, P. (2019). Global variability analysis of mRNA and protein concentrations across and within human tissues. Nar Genom Bioinform 2, lqz010. 10.1093/nargab/lqz010.

8. Raj, A., and Oudenaarden, A. van (2008). Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. Cell 135, 216–226. 10.1016/j.cell.2008.09.050.

9. Cai, L., Friedman, N., and Xie, X.S. (2006). Stochastic protein expression in individual cells at the single molecule level. Nature 440, 358–362. 10.1038/nature04599.

10. Sonneveld, S., Verhagen, B.M.P., and Tanenbaum, M.E. (2020). Heterogeneity in mRNA Translation. Trends Cell Biol. 30, 606–618. 10.1016/j.tcb.2020.04.008.

11. Livingston, N.M., Kwon, J., Valera, O., Saba, J.A., Sinha, N.K., Reddy, P., Nelson, B., Wolfe, C., Ha, T., Green, R., et al. (2023). Bursting translation on single mRNAs in live cells. Mol. Cell 83, 2276-2289.e11. 10.1016/j.molcel.2023.05.019.

12. Aebersold, R., and Mann, M. (2003). Mass spectrometry-based proteomics. Nature 422, 198–207. 10.1038/nature01511.

13. Pappireddi, N., Martin, L., and Wühr, M. (2019). A Review on Quantitative Multiplexed Proteomics. ChemBioChem 20, 1210–1224. 10.1002/cbic.201800650.

14. Johnson, A., Stadlmeier, M., and Wühr, M. (2021). TMTpro Complementary Ion Quantification Increases Plexing and Sensitivity for Accurate Multiplexed Proteomics at

the MS2 Level. J. Proteome Res. 20, 3043–3052. 10.1021/acs.jproteome.0c00813.
15. Thompson, A., Schäfer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., Johnstone, R., Mohammed, A.K.A., and Hamon, C. (2003). Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS. Anal. Chem. *75*, 1895–1*90*4. 10.1021/ac026*2560*.
16. Demichev, V., Messner, C.B., Vernardis, S.I., Lilley, K.S., and Ralser, M. (2020). DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. Nat. Methods *17*, 41–44. 10.1038/s41592-019-0638-x.
17. Ammar, C., Schessner, J.P., Willems, S., Michaelis, A.C., and Mann, M. (2023). Accurate label-free quantification by directLFQ to compare unlimited numbers of proteomes. Mol. Cell. Proteom., 100581. 10.1016/j.mcpro.2023.100581.
18. McAlister, G.C., Nusinow, D.P., Jedrychowski, M.P., Wühr, M., Huttlin, E.L., Erickson, B.K., Rad, R., Haas, W., and Gygi, S.P. (2014). MultiNotch MS3 Enables Accurate, Sensitive, and Multiplexed Detection of Differential Expression across Cancer Cell Line Proteomes. Anal. Chem. *86*, 7150–7158. 10.1021/ac502040v.
19. Lucitt, M.B., Price, T.S., Pizarro, A., Wu, W., Yocum, A.K., Seiler, C., Pack, M.A., Blair, I.A., FitzGerald, G.A., and Grosser, T. (2008). Analysis of the Zebrafish Proteome during Embryonic Development*. Mol. Cell. Proteom. 7, 981–994. 10.1074/mcp.m700382-mcp200.
20. Gao, Y., Liu, X., Tang, B., Li, C., Kou, Z., Li, L., Liu, W., Wu, Y., Kou, X., Li, J., et al. (2017). Protein Expression Landscape of Mouse Embryos during Pre-implantation Development. Cell Reports 21, 3957–3969. 10.1016/j.celrep.2017.11.111.
21. Purushothaman, K., Das, P.P., Presslauer, C., Lim, T.K., Johansen, S.D., Lin, Q., and Babiak, I. (2019). Proteomics Analysis of Early Developmental Stages of Zebrafish Embryos. Int J Mol Sci 20, 6359. 10.3390/ijms20246359.
22. Sun, L., Bertke, M.M., Champion, M.M., Zhu, G., Huber, P.W., and Dovichi, N.J. (2014). Quantitative proteomics of Xenopus laevis embryos: expression kinetics of nearly 4000 proteins during early development. Sci. Rep. 4, 4365. 10.1038/srep04365.
23. Abdulghani, M., Song, G., Kaur, H., Walley, J.W., and Tuteja, G. (2019). Comparative Analysis of the Transcriptome and Proteome during Mouse Placental Development. J Proteome Res 18, 2088–2099. 10.1021/acs.jproteome.8b00970.
24. Delsuc, F., Brinkmann, H., Chourrout, D., and Philippe, H. (2006). Tunicates and not cephalochordates are the closest living relatives of vertebrates. Nature *439*, 965–968. 10.1038/nature04336.
25. Dehal, P., Satou, Y., Campbell, R.K., Chapman, J., Degnan, B., Tomaso, A.D., Davidson, B., Gregorio, A.D., Gelpke, M., Goodstein, D.M., et al. (2002). The Draft Genome of Ciona intestinalis: Insights into Chordate and Vertebrate Origins. Science 298, *215*7–2167. 10.1126/science.*108*0049.
26. Olinski, R.P., Lundin, L.-G., and Hallböök, F. (2006). Conserved Synteny Between the Ciona Genome and Human Paralogons Identifies Large Duplication Events in the

945 Molecular Evolution of the Insulin-Relaxin Gene Family. Mol. Biol. Evol. 23, 10–22.
946 10.1093/molbev/msj002.
947 27. Simakov, O., Bredeson, J., Berkoff, K., Marletaz, F., Mitros, T., Schultz, D.T.,
948 O'Connell, B.L., Dear, P., Martinez, D.E., Steele, R.E., et al. (2022). Deeply conserved
949 synteny and the evolution of metazoan chromosomes. Sci. Adv. 8, eabi5884.
950 10.1126/sciadv.abi5884.
951 28. Kikuta, H., Laplante, M., Navratilova, P., Komisarczuk, A.Z., Engström, P.G.,
952 Fredman, D., Akalin, A., Caccamo, M., Sealy, I., Howe, K., et al. (2007). Genomic
953 regulatory blocks encompass multiple neighboring genes and maintain conserved
954 synteny in vertebrates. Genome Res. 17, 545–555. 10.1101/gr.6086307.
955 29. Doglio, L., Goode, D.K., Pelleri, M.C., Pauls, S., Frabetti, F., Shimeld, S.M., Vavouri,
956 T., and Elgar, G. (2013). Parallel Evolution of Chordate Cis-Regulatory Code for
957 Development. PLoS Genet. 9, e1003904. 10.*137*1/journal.pgen.1003904.
958 30. Sanges, R., Hadzhiev, Y., Gueroult-Bellone, M., Roure, A., Ferg, M., Meola, N.,
959 Amore, G., Basu, S., Brown, E.R., Simone, M.D., et al. (2013). Highly conserved elements
960 discovered in vertebrates are present in non-syntenic loci of tunicates, act as enhancers
961 and can be transcribed during development. Nucleic Acids Res. 41, 3600–3618.
962 10.1093/nar/gkt030.
963 31. Prummel, K.D., Hess, C., Nieuwenhuize, S., Parker, H.J., Rogers, K.W., Kozmikova,
964 I., Racioppi, C., Brombacher, E.C., Czarkwiani, A., Knapp, D., et al. (2019). A conserved
965 regulatory program initiates lateral plate mesoderm emergence across chordates. Nat.
966 Commun. 10, 3857. 10.1038/s41467-019-*115*61-7.
967 32. Abitua, P.B., Wagner, E., Navarrete, I.A., and Levine, M. (2012). Identification of a
968 rudimentary neural crest in a non-vertebrate chordate. Nature *492*, 104–107.
969 10.1038/nature11589.
970 33. Abitua, P.B., Gainous, T.B., Kaczmarczyk, A.N., Winchell, C.J., Hudson, C., Kamata,
971 K., Nakagawa, M., Tsuda, M., Kusakabe, T.G., and Levine, M. (2015). The pre-vertebrate
972 origins of neurogenic placodes. Nature *524*, 462–465. 10.1038/nature14657.
973 *34*. Stolfi, A., Gainous, T.B., Young, J.J., Mori, A., Levine, M., and Christiaen, L. (2010).
974 Early Chordate Origins of the Vertebrate Second Heart Field. Science 329, 565–568.
975 10.1126/science.1190*181*.
976 35. Stolfi, A., Ryan, K., Meinertzhagen, I.A., and Christiaen, L. (2015). Migratory neuronal
977 progenitors arise from the neural plate borders in tunicates. Nature *527*, 371–374.
978 10.1038/nature15758.
979 36. Horie, R., Hazbun, A., Chen, K., Cao, C., Levine, M., and Horie, T. (2018). Shared
980 evolutionary origin of vertebrate neural crest and cranial placodes. Nature 560, 228–232.
981 10.1038/s41586-018-0385-7.
982 37. Lemaire, L.A., Cao, C., Yoon, P.H., Long, J., and Levine, M. (2021). The
983 hypothalamus predates the origin of vertebrates. Sci. Adv. 7, eabf7452.
984 10.1126/sciadv.abf7452.

38. Long, J., Mariossi, A., Cao, C., Mo, Z., Thompson, J.W., Levine, M.S., and Lemaire, L.A. (2023). Cereblon influences the timing of muscle differentiation in Ciona tadpoles. Proc. Natl. Acad. Sci. *120*, e2309989120. 10.1073/pnas.2309989120.
39. Reeves, W.M., Wu, Y., Harder, M.J., and Veeman, M.T. (2017). Functional and evolutionary insights from the Ciona notochord transcriptome. Development *144*, 3375–*3387*. 10.1242/dev.156174.
40. Papadogiannis, V., Pennati, A., Parker, H.J., Rothbächer, U., Patthey, C., Bronner, M.E., and Shimeld, S.M. (2022). Hmx gene conservation identifies the origin of vertebrate cranial ganglia. Nature *605*, 701–705. 10.1038/s41586-022-04742-w.
41. Cao, C., Lemaire, L.A., Wang, W., Yoon, P.H., Choi, Y.A., Parsons, L.R., Matese, J.C., Wang, W., Levine, M., and Chen, K. (2019). Comprehensive single-cell transcriptome lineages of a proto-vertebrate. Nature *571*, 349–354. 10.1038/s41586-019-*138*5-y.
42. Keller, T.E., Han, P., and Yi, S.V. (2016). Evolutionary Transition of Promoter and Gene Body DNA Methylation across Invertebrate–Vertebrate Boundary. Mol Biol Evol 33, 1019–1028. 10.1093/molbev/msv345.
43. Zhang, T., Xu, Y., Imai, K., Fei, T., Wang, G., Dong, B., Yu, T., Satou, Y., Shi, W., and Bao, Z. (2020). A single-cell analysis of the molecular lineage of chordate embryogenesis. Sci Adv 6, eabc4*773*. 10.1126/sciadv.abc4773.
44. Sladitschek, H.L., Fiuza, U.-M., Pavlinic, D., Benes, V., Hufnagel, L., and Neveu, P.A. (2020). MorphoSeq: Full Single-Cell Transcriptome Dynamics Up to Gastrulation in a Chordate. Cell 181, 922-935.e21. 10.1016/j.cell.2020.03.055.
45. Suzuki, M.M., Mori, T., and Satoh, N. (2016). The Ciona intestinalis cleavage clock is independent of DNA methylation. Genomics 108, 168–176. 10.1016/j.ygeno.2016.10.001.
46. Madgwick, A., Magri, M.S., Dantec, C., Gailly, D., Fiuza, U.-M., Guignard, L., Hettinger, S., Gomez-Skarmeta, J.L., and Lemaire, P. (2019). Evolution of embryonic cis-regulatory landscapes between divergent Phallusia and Ciona ascidians. Dev Biol *448*, 71–87. 10.1016/j.ydbio.2019.01.003.
47. Kubo, A., Suzuki, N., Yuan, X., Nakai, K., Satoh, N., Imai, K.S., and Satou, Y. (2010). Genomic cis-regulatory networks in the early Ciona intestinalis embryo. Development 137, 1613–1623. 10.1242/dev.046*78*9.
48. Irie, N., and Kuratani, S. (2011). Comparative transcriptome analysis reveals vertebrate phylotypic period during organogenesis. Nat. Commun. 2, 248. 10.1038/ncomms1248.
49. Yanai, I., Peshkin, L., Jorgensen, P., and Kirschner, M.W. (2011). Mapping Gene Expression in Two Xenopus Species: Evolutionary Constraints and Developmental Flexibility. Dev. Cell 20, 483–496. 10.1016/j.devcel.2011.03.015.
50. Gerstein, M.B., Rozowsky, J., Yan, K.-K., Wang, D., Cheng, C., Brown, J.B., Davis, C.A., Hillier, L., Sisu, C., Li, J.J., et al. (2014). Comparative analysis of the transcriptome

1025 across distant species. Nature *512*, 445–448. 10.1038/nature13424.
1026 51. Chan, M.E., Bhamidipati, P.S., Goldsby, H.J., Hintze, A., Hofmann, H.A., and Young,
1027 R.L. (2021). Comparative transcriptomics reveals distinct patterns of gene expression
1028 conservation through vertebrate embryogenesis. Genome Biol. Evol. 13, evab160-.
1029 10.1093/gbe/evab160.
1030 52. Uesaka, M., Kuratani, S., and Irie, N. (2022). The developmental hourglass model
1031 and recapitulation: An attempt to integrate the two models. J. Exp. Zoöl. Part B: Mol. Dev.
1032 Evol. 338, 76–86. 10.1002/jez.b.23027.
1033 *53*. Hu, H., Uesaka, M., Guo, S., Shimai, K., Lu, T.-M., Li, F., Fujimoto, S., Ishikawa, M.,
1034 Liu, S., Sasagawa, Y., et al. (2017). Constrained vertebrate evolution by pleiotropic
1035 genes. Nat. Ecol. Evol. 1, 1722–1730. 10.1038/s41559-017-0318-0.
1036 54. Marlétaz, F., Firbas, P.N., Maeso, I., Tena, J.J., Bogdanovic, O., Perry, M., Wyatt,
1037 C.D.R., Calle-Mustienes, E. de la, Bertrand, S., Burguera, D., et al. (2018). Amphioxus
1038 functional genomics and the origins of vertebrate gene regulation. Nature *564*, 64–70.
1039 10.1038/s41586-018-0734-6.
1040 55. Levin, M., Anavy, L., Cole, A.G., Winter, E., Mostov, N., Khair, S., Senderovich, N.,
1041 Kovalev, E., Silver, D.H., Feder, M., et al. (2016). The mid-developmental transition and
1042 the evolution of animal body plans. Nature *531*, 637–641. 10.1038/nature16994.
1043 56. Wu, L., Ferger, K.E., and Lambert, J.D. (2019). Gene Expression Does Not Support
1044 the Developmental Hourglass Model in Three Animals with Spiralian Development. Mol.
1045 Biol. Evol. 36, 1373–1383. 10.1093/molbev/msz065.
1046 57. P., B.-E.O.R., E., J.J., and Richardson, M.K. (2003). Inverting the hourglass:
1047 quantitative evidence against the phylotypic stage in vertebrate development. Proc. R.
1048 Soc. Lond. Ser. B: Biol. Sci. *270*, 341–346. 10.1098/rspb.2002.2242.
1049 58. Dunn, C.W., Zapata, F., Munro, C., Siebert, S., and Hejnol, A. (2018). Pairwise
1050 comparisons across species are problematic when analyzing functional genomic data.
1051 Proc. Natl. Acad. Sci. 115, E409–E417. 10.1073/pnas.1707515115.
1052 59. Yanai, I. (2018). Development and Evolution through the Lens of Global Gene
1053 Regulation. Trends Genet. 34, 11–20. 10.1016/j.tig.2017.09.011.
1054 60. Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen,
1055 W., and Selbach, M. (2011). Global quantification of mammalian gene expression control.
1056 Nature *473*, 337–342. 10.1038/nature10098.
1057 61. Laurent, J.M., Vogel, C., Kwon, T., Craig, S.A., Boutz, D.R., Huse, H.K., Nozue, K.,
1058 Walia, H., Whiteley, M., Ronald, P.C., et al. (2010). Protein abundances are more
1059 conserved than mRNA abundances across diverse taxa. PROTEOMICS 10, 4209–4212.
1060 10.1002/pmic.201000327.
1061 62. Goldberger, R.F. (1980). Biological Regulation and Development, Molecular
1062 Organization and Cell Function (Springer) 10.1007/978-1-4684-9933-9.
1063 63. Baxi, A.B., Lombard-Banek, C., Moody, S.A., and Nemes, P. (2018). Proteomic
1064 Characterization of the Neural Ectoderm Fated Cell Clones in the Xenopus laevis Embryo

by High-Resolution Mass Spectrometry. ACS Chem. Neurosci. 9, 2064–2073. 10.1021/acschemneuro.7b00525.
64. Wühr, M., Freeman, R.M., Presler, M., Horb, M.E., Peshkin, L., Gygi, S.P., and Kirschner, M.W. (2014). Deep Proteomics of the Xenopus laevis Egg using an mRNA-Derived Reference Database. Curr. Biol. 24, 1467–1475. 10.1016/j.cub.2014.05.044.
65. Gupta, M., Sonnett, M., Ryazanova, L., Presler, M., and Wühr, M. (2018). Quantitative Proteomics of Xenopus Embryos I, Sample Preparation. Methods Mol Biology Clifton N J *1865*, 175–194. 10.1007/978-1-4939-8784-9_13.
66. Evans, V.C., Barker, G., Heesom, K.J., Fan, J., Bessant, C., and Matthews, D.A. (2012). De novo derivation of proteomes from transcriptomes for transcript and protein identification. Nat. Methods 9, 1207–1211. 10.1038/nmeth.2227.
67. Kaplan, N.A., Wang, W., and Christiaen, L. (2019). Initial characterization of Wnt-Tcf functions during Ciona heart development. Dev. Biology 448, 199–209. 10.1016/j.ydbio.2018.12.018.
68. Sharma, S., Wang, W., and Stolfi, A. (2019). Single-cell transcriptome profiling of the Ciona larval brain. Dev. Biology 448, 226–*236*. 10.1016/j.ydbio.2018.09.023.
69. Wang, W., Niu, X., Stuart, T., Jullian, E., Mauck, W.M., Kelly, R.G., Satija, R., and Christiaen, L. (2019). A single-cell transcriptional roadmap for cardiopharyngeal fate diversification. Nat. Cell Biology 21, 674–686. 10.1038/s41556-019-0336-z.
70. Consortium, T.U., Bateman, A., Martin, M.-J., Orchard, S., Magrane, M., Ahmad, S., Alpi, E., Bowler-Barnett, E.H., Britto, R., Bye-A-Jee, H., et al. (2022). UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Res. 51, D523–D531. 10.1093/nar/gkac1052.
71. Satou, Y., Nakamura, R., Yu, D., Yoshida, R., Hamada, M., Fujie, M., Hisata, K., Takeda, H., and Satoh, N. (2019). A Nearly Complete Genome of Ciona intestinalis Type A (C. robusta) Reveals the Contribution of Inversion to Chromosomal Evolution in the Genus Ciona. Genome Biol. Evol. 11, 3144–3157. 10.1093/gbe/evz228.
72. Satou, Y., Tokuoka, M., Oda-Ishii, I., Tokuhiro, S., Ishida, T., Liu, B., and Iwamura, Y. (2021). A Manually Curated Gene Model Set for an Ascidian, Ciona robusta (Ciona intestinalis Type A). Zoöl. Sci. 39, 253–260. 10.2108/zs210102.
73. Tsagkogeorga, G., Cahais, V., and Galtier, N. (2012). The Population Genomics of a Fast Evolver: High Levels of Diversity, Functional Constraint, and Molecular Adaptation in the Tunicate Ciona intestinalis. Genome Biol. Evol. 4, 852–861. 10.1093/gbe/evs054.
74. Santesmasses, D., Mariotti, M., and Guigó, R. (2017). Selenoproteins, Methods and Protocols. Methods Mol Biology *1661*, 17–28. 10.1007/978-1-4939-7258-6_2.
75. Satou, Y., Hamaguchi, M., Takeuchi, K., Hastings, K.E.M., and Satoh, N. (2006). Genomic overview of mRNA 5′-leader trans-splicing in the ascidian Ciona intestinalis. Nucleic Acids Res. 34, 3378–3388. 10.1093/nar/gkl418.
76. Messerschmidt, D.M., Vries, W. de, Ito, M., Solter, D., Ferguson-Smith, A., and Knowles, B.B. (2012). Trim28 Is Required for Epigenetic Stability During Mouse Oocyte

to Embryo Transition. Science *335*, *1499*–1502. 10.1126/science.1216154.
77. Bultman, S.J., Gebuhr, T.C., Pan, H., Svoboda, P., Schultz, R.M., and Magnuson, T. (2006). Maternal BRG1 regulates zygotic genome activation in the mouse. Genes Dev. 20, 1744–1754. 10.1101/gad.1435*106*.
78. Toralova, T., Kinterova, V., Chmelikova, E., and Kanka, J. (2020). The neglected part of early embryonic development: maternal protein degradation. Cell. Mol. Life Sci. 77, 3177–3194. 10.1007/s00018-020-03482-2.
79. Zhang, H., Ji, S., Zhang, K., Chen, Y., Ming, J., Kong, F., Wang, L., Wang, S., Zou, Z., Xiong, Z., et al. (2023). Stable maternal proteins underlie distinct transcriptome, translatome, and proteome reprogramming during mouse oocyte-to-embryo transition. Genome Biol. 24, 166. 10.1186/s13059-023-02997-8.
80. Yamada, L., Saito, T., Taniguchi, H., Sawada, H., and Harada, Y. (*2009*). Comprehensive Egg Coat Proteome of the Ascidian Ciona intestinalis Reveals Gamete Recognition Molecules Involved in Self-sterility*. J. Biol. Chem. *284*, 9402–9410. 10.1074/jbc.m809672200.
81. Nomura, M., Nakajima, A., and Inaba, K. (2009). Proteomic profiles of embryonic development in the ascidian Ciona intestinalis. Dev. Biol. *325*, 468–481. 10.1016/j.ydbio.2008.10.038.
82. Gillespie, M.A., Palii, C.G., Sanchez-Taltavull, D., Shannon, P., Longabaugh, W.J.R., Downes, D.J., Sivaraman, K., Espinoza, H.M., Hughes, J.R., Price, N.D., et al. (2020). Absolute Quantification of Transcription Factors Reveals Principles of Gene Regulation in Erythropoiesis. Mol. Cell 78, 960-974.e11. 10.1016/j.molcel.2020.03.031.
83. Imai, K.S., Hino, K., Yagi, K., Satoh, N., and Satou, Y. (2004). Gene expression profiles of transcription factors and signaling molecules in the ascidian embryo: towards a comprehensive understanding of gene networks. Development *131*, 4047–4058. 10.1242/dev.01270.
84. Ahn, H.R., and Kim, G.J. (2012). The Ascidian Numb Gene Involves in the Formation of Neural Tissues. Dev. Reprod. 16, 371–378. 10.12717/dr.2012.16.4.371.
85. Walton, K.D., Croce, J.C., Glenn, T.D., Wu, S.-Y., and McClay, D.R. (2006). Genomics and expression profiles of the Hedgehog and Notch signaling pathways in sea urchin development. Dev. Biol. *300*, 153–164. 10.1016/j.ydbio.2006.08.064.
86. Oda-Ishii, I., Kubo, A., Kari, W., Suzuki, N., Rothbächer, U., and Satou, Y. (2016). A Maternal System Initiating the Zygotic Developmental Program through Combinatorial Repression in the Ascidian Embryo. PLoS Genet. 12, e1006045. 10.1371/journal.pgen.1006045.
87. Yamada, L. (2006). Embryonic expression profiles and conserved localization mechanisms of pem/postplasmic mRNAs of two species of ascidian, Ciona intestinalis and Ciona savignyi. Dev. Biol. *296*, 524–536. 10.1016/j.ydbio.2006.05.018.
88. Shirae-Kurabayashi, M., Matsuda, K., and Nakamura, A. (2011). Ci-Pem-1 localizes to the nucleus and represses somatic gene transcription in the germline of Ciona

intestinalis embryos. Development 138, 2871–2881. 10.1242/dev.058131.
89. Tsitsiridis, G., Steinkamp, R., Giurgiu, M., Brauner, B., Fobo, G., Frishman, G., Montrone, C., and Ruepp, A. (2022). CORUM: the comprehensive resource of mammalian protein complexes–2022. Nucleic Acids Res 51, D539–D545. 10.1093/nar/gkac1015.
90. Bochman, M.L., and Schwacha, A. (2009). The Mcm Complex: Unwinding the Mechanism of a Replicative Helicase. Microbiol. Mol. Biol. Rev. 73, 652–683. 10.1*128*/mmbr.00019-09.
91. Kubota, H. (2002). Function and regulation of cytosolic molecular chaperone CCT. Vitam. Horm. 65, 313–331. 10.1016/s0083-6729(02)65069-1.
92. Uehara, R., Nozawa, R., Tomioka, A., Petry, S., Vale, R.D., Obuse, C., and Goshima, G. (2009). The augmin complex plays a critical role in spindle microtubule generation for mitotic progression and cytokinesis in human cells. Proc. Natl. Acad. Sci. 106, 6998–7003. 10.1073/pnas.0901587106.
93. Liang, J., Xia, L., Oyang, L., Lin, J., Tan, S., Yi, P., Han, Y., Luo, X., Wang, H., Tang, L., et al. (2020). The functions and mechanisms of prefoldin complex and prefoldin-subunits. Cell Biosci. 10, 87. 10.1186/s13578-020-00446-8.
94. Elias, J.E., and Gygi, S.P. (2007). Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nat. Methods 4, 207–214. 10.1038/nmeth1019.
95. Cao, W.X., Kabelitz, S., Gupta, M., Yeung, E., Lin, S., Rammelt, C., Ihling, C., Pekovic, F., Low, T.C.H., Siddiqui, N.U., et al. (2020). Precise Temporal Regulation of Post-transcriptional Repressors Is Required for an Orderly Drosophila Maternal-to-Zygotic Transition. Cell Reports 31, 107783. 10.1016/j.celrep.2020.107783.
96. White, R.J., Collins, J.E., Sealy, I.M., Wali, N., Dooley, C.M., Digby, Z., Stemple, D.L., Murphy, D.N., Billis, K., Hourlier, T., et al. (2017). A high-resolution mRNA expression time course of embryonic development in zebrafish. eLife 6, e30860. 10.7554/elife.30860.
97. Hebenstreit, D., Fang, M., Gu, M., Charoensawan, V., Oudenaarden, A. van, and Teichmann, S.A. (2011). RNA sequencing reveals two major classes of gene expression levels in metazoan cells. Mol. Syst. Biol. 7, 497–497. 10.1038/msb.2011.28.
98. Imai, K.S., Levine, M., Satoh, N., and Satou, Y. (2006). Regulatory Blueprint for a Chordate Embryo. Science *312*, 1183–1187. 10.1126/science.1123404.
99. Kobayashi, K., Maeda, K., Tokuoka, M., Mochizuki, A., and Satou, Y. (2018). Controlling Cell Fate Specification System by Key Genes Determined from Network Structure. iScience 4, 281–293. 10.1016/j.isci.2018.05.004.
100. Mochizuki, Y., Satou, Y., and Satoh, N. (2003). Large-scale characterization of genes specific to the larval nervous system in the ascidian Ciona intestinalis. genesis 36, 62–71. 10.1002/gene.10199.
101. Satou, Y., Takatori, N., Yamada, L., Mochizuki, Y., Hamaguchi, M., Ishikawa, H., Chiba, S., Imai, K., Kano, S., Murakami, S.D., et al. (2001). Gene expression profiles in

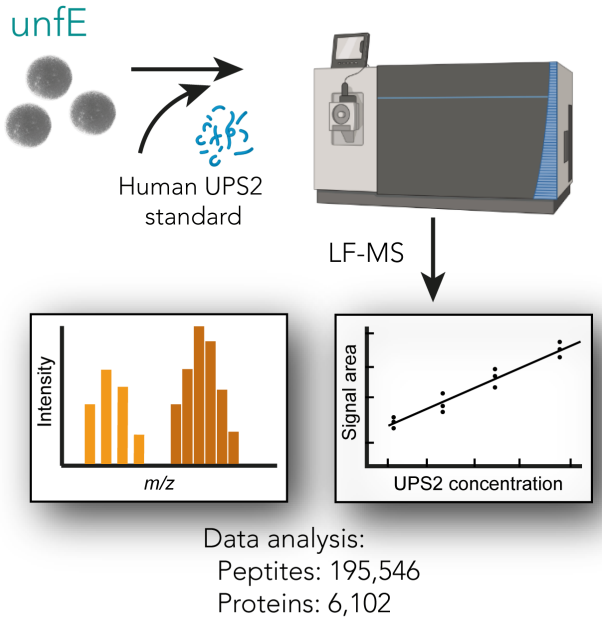
Ciona intestinalis tailbud embryos. Development 128, 2893–2904. 10.1242/dev.128.15.2893.
102. Kawai, N., Ogura, Y., Ikuta, T., Saiga, H., Hamada, M., Sakuma, T., Yamamoto, T., Satoh, N., and Sasakura, Y. (2015). Hox10-regulated endodermal cell migration is essential for development of the ascidian intestine. Dev. Biol. *403*, 43–56. 10.1016/j.ydbio.2015.03.018.
103. Buccitelli, C., and Selbach, M. (2020). mRNAs, proteins and the emerging principles of gene expression control. Nat. Rev. Genet. 21, 630–644. 10.1038/s41576-020-0258-4.
104. Wang, D., Eraslan, B., Wieland, T., Hallström, B., Hopf, T., Zolg, D.P., Zecha, J., Asplund, A., Li, L., Meng, C., et al. (2019). A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. Mol. Syst. Biology 15, e8503. 10.15252/msb.20188503.
105. Mergner, J., Frejno, M., List, M., Papacek, M., Chen, X., Chaudhary, A., Samaras, P., Richter, S., Shikata, H., Messerer, M., et al. (2020). Mass-spectrometry-based draft of the Arabidopsis proteome. Nature *579*, 409–414. 10.1038/s41586-020-2094-2.
106. Sonnett, M., Gupta, M., Nguyen, T., and Wühr, M. (2018). Quantitative Proteomics for Xenopus Embryos II, Data Analysis. Methods Mol Biology Clifton N J 1865, 195–215. 10.1007/978-1-4939-8784-9_14.
107. Itallie, E.V., Kalocsay, M., Wühr, M., Peshkin, L., and Kirschner, M.W. (2021). Transitions in the proteome and phospho-proteome during Xenopus laevis development. bioRxiv, 2021.08.05.455309. 10.1101/2021.08.05.455309.
108. Delsuc, F., Philippe, H., Tsagkogeorga, G., Simion, P., Tilak, M.-K., Turon, X., López-Legentil, S., Piette, J., Lemaire, P., and Douzery, E.J.P. (2018). A phylogenomic framework and timescale for comparative studies of tunicates. BMC Biol. 16, 39. 10.1186/s12915-018-0499-2.
109. Izzi, S.A., Colantuono, B.J., Sullivan, K., Khare, P., and Meedel, T.H. (2013). Functional studies of the Ciona intestinalis myogenic regulatory factor reveal conserved features of chordate myogenesis. Dev. Biol. *376*, 213–223. 10.1016/j.ydbio.2013.01.033.
110. Veeman, M.T., Nakatani, Y., Hendrickson, C., Ericson, V., Lin, C., and Smith, W.C. (2007). chongmague reveals an essential role for laminin-mediated boundary formation in chordate convergence and extension movements. Development 135, 33–41. 10.1242/dev.010892.
111. Hanel, M.L., Wuebbles, R.D., and Jones, P.L. (2009). Muscular dystrophy candidate gene FRG1 is critical for muscle development. Dev. Dyn. *238*, 1502–1512. 10.1002/dvdy.21830.
112. Wuebbles, R.D., Hanel, M.L., and Jones, P.L. (2009). FSHD region gene 1 (FRG1) is crucial for angiogenesis linking FRG1 to facioscapulohumeral muscular dystrophy-associated vasculopathy. Dis. Model. Mech. 2, 267–274. 10.1242/dmm.002261.
113. Lyabin, D.N., Eliseeva, I.A., and Ovchinnikov, L.P. (2014). YB-1 protein: functions and regulation. Wiley Interdiscip. Rev.: RNA 5, 95–110. 10.1002/wrna.1200.

114. Kumari, P., Gilligan, P.C., Lim, S., Tran, L.D., Winkler, S., Philp, R., and Sampath, K. (2013). An essential role for maternal control of Nodal signaling. eLife 2, e00683. 10.7554/elife.00683.
115. Session, A.M., Uno, Y., Kwon, T., Chapman, J.A., Toyoda, A., Takahashi, S., Fukui, A., Hikosaka, A., Suzuki, A., Kondo, M., et al. (2016). Genome evolution in the allotetraploid frog Xenopus laevis. Nature *538*, 336–343. 10.1038/nature19840.
116. Winkley, K.M., Kourakis, M.J., DeTomaso, A.W., Veeman, M.T., and Smith, W.C. (2019). Tunicate gastrulation. Curr Top Dev Biol *136*, 219–242. 10.1016/bs.ctdb.2019.09.001.
117. Schrimpf, S.P., Weiss, M., Reiter, L., Ahrens, C.H., Jovanovic, M., Malmström, J., Brunner, E., Mohanty, S., Lercher, M.J., Hunziker, P.E., et al. (2009). Comparative Functional Analysis of the Caenorhabditis elegans and Drosophila melanogaster Proteomes. PLoS Biol. 7, e1000048. 10.1371/journal.pbio.1000048.
118. Liu, J., and Robinson-Rechavi, M. (2018). Adaptive Evolution of Animal Proteins over Development: Support for the Darwin Selection Opportunity Hypothesis of Evo-Devo. Mol. Biol. Evol. 35, 2862–2872. 10.1093/molbev/msy175.
119. Roux, J., and Robinson-Rechavi, M. (2008). Developmental Constraints on Vertebrate Genome Evolution. PLoS Genet. 4, e1000311. 10.1371/journal.pgen.1000311.
120. Klont, F., Bras, L., Wolters, J.C., Ongay, S., Bischoff, R., Halmos, G.B., and Horvatovich, P. (2018). Assessment of Sample Preparation Bias in Mass Spectrometry-Based Proteomics. Anal. Chem. 90, 5405–5413. 10.1021/acs.analchem.8b00600.
121. Brunetti, R., Gissi, C., Pennati, R., Caicci, F., Gasparini, F., and Manni, L. (2015). Morphological evidence that the molecularly determined Ciona intestinalis type A and type B are different species: Ciona robusta and Ciona intestinalis. J. Zoöl. Syst. Evol. Res. 53, 186–193. 10.1111/jzs.12101.
122. Christiaen, L., Wagner, E., Shi, W., and Levine, M. (2009). Isolation of Sea Squirt (Ciona) Gametes, Fertilization, Dechorionation, and Development: Figure 1. Cold Spring Harb. Protoc. 2009, pdb.prot5344. 10.1101/pdb.prot5344.
123. Hotta, K., Mitsuhara, K., Takahashi, H., Inaba, K., Oka, K., Gojobori, T., and Ikeo, K. (2007). A web-based interactive developmental table for the ascidian Ciona intestinalis, including 3D real-image embryo reconstructions: I. From fertilized egg to hatching larva. Dev. Dyn. 236, 1790–1805. 10.1002/dvdy.21188.
124. Abdul-Wajid, S., Veeman, M.T., Chiba, S., Turner, T.L., and Smith, W.C. (2014). Exploiting the Extraordinary Genetic Polymorphism of Ciona for Developmental Genetics with Whole Genome Sequencing. Genetics *197*, 49–59. 10.1534/genetics.114.161778.
125. Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29, 644–652. 10.1038/nbt.1883.

126. Brozovic, M., Dantec, C., Dardaillon, J., Dauga, D., Faure, E., Gineste, M., Louis, A., Naville, M., Nitta, K.R., Piette, J., et al. (2017). ANISEED 2017: extending the integrated ascidian database to the exploration and evolutionary comparison of genome-scale datasets. Nucleic Acids Res. 46, gkx1108-. 10.1093/nar/gkx1108.
127. Smit, Hubley, and Green, R.& (2015). RepeatMasker Open-4.0.
128. Pertea, G., Huang, X., Liang, F., Antonescu, V., Sultana, R., Karamycheva, S., Lee, Y., White, J., Cheung, F., Parvizi, B., et al. (2003). TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. Bioinformatics 19, 651–652. 10.1093/bioinformatics/btg034.
129. Huang, X., and Madan, A. (1999). CAP3: A DNA Sequence Assembly Program. Genome Res. 9, 868–877. 10.1101/gr.9.9.868.
130. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. 215, 403–410. 10.1016/s0022-2836(05)80360-2.
131. Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28, 3150–3152. 10.1093/bioinformatics/bts565.
132. Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22, 1658–1659. 10.1093/bioinformatics/btl158.
133. Wessel, D., and Flügge, U.I. (1984). A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. Anal. Biochem. 138, 141–143. 10.1016/0003-2697(84)90782-6.
134. Nguyen, T., Costa, E.J., Deibert, T., Reyes, J., Keber, F.C., Tomschik, M., Stadlmeier, M., Gupta, M., Kumar, C.K., Cruz, E.R., et al. (2022). Differential nuclear import sets the timing of protein access to the embryonic genome. Nat. Commun. 13, 5887. 10.1038/s41467-022-33429-z.
135. Edwards, A., and Haas, W. (2015). Multiplexed Quantitative Proteomics for High-Throughput Comprehensive Proteome Comparisons of Human Cell Lines. Methods Mol. Biol. *1394*, 1–13. 10.1007/978-1-4939-3341-9_1.
136. Rappsilber, J., Mann, M., and Ishihama, Y. (2007). Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat. Protoc. 2, 1896–1906. 10.1038/nprot.2007.261.
137. Zahn, N., James-Zorn, C., Ponferrada, V.G., Adams, D.S., Grzymkowski, J., Buchholz, D.R., Nascone-Yoder, N.M., Horb, M., Moody, S.A., Vize, P.D., et al. (2022). Normal Table of Xenopus development: a new graphical resource. Development 149, dev200356. 10.1242/dev.200356.
138. Eng, J.K., McCormack, A.L., and Yates, J.R. (1994). An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J. Am. Soc. Mass Spectrom. 5, 976–989. 10.1016/1044-0305(94)80016-2.

139. Savitski, M.M., Wilhelm, M., Hahne, H., Kuster, B., and Bantscheff, M. (2015). A Scalable Approach for Protein False Discovery Rate Estimation in Large Proteomic Data Sets[S]. Mol. Cell. Proteom. 14, 2394–2404. 10.1074/mcp.m114.046995.
140. Satou, Y., Kawashima, T., Shoguchi, E., Nakayama, A., and Satoh, N. (2005). An Integrated Database of the Ascidian, Ciona intestinalis: Towards Functional Genomics. Zoöl. Sci. 22, 837–843. 10.2108/zsj.22.837.
141. Nitta, K.R., Vincentelli, R., Jacox, E., Cimino, A., Ohtsuka, Y., Sobral, D., Satou, Y., Cambillau, C., and Lemaire, P. (2019). High-Throughput Protein Production and Purification, Methods and Protocols. Methods Mol. Biol. *2025*, 487–517. 10.1007/978-1-4939-9624-7_23.
142. Kolberg, L., Raudvere, U., Kuzmin, I., Adler, P., Vilo, J., and Peterson, H. (2023). g:Profiler—interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). Nucleic Acids Res. 51, W207–W212. 10.1093/nar/gkad347.
143. Soneson, C., Love, M.I., and Robinson, M.D. (2016). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Research 4, 1521. 10.12688/f1000research.7563.2.
144. Treen, N., Chavarria, E., Weaver, C.J., Brangwynne, C.P., and Levine, M. (2023). An FGF timer for zygotic genome activation. Genes Dev. 37, 80–85. 10.1101/gad.350164.122.
145. Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods 14, 417–419. 10.1038/nmeth.4197.
146. Bolstad, B. (2023). preprocessCore: A collection of pre-processing functions. https://bioconductor.org/packages/preprocessCore.
147. Alexa, A., and Rahnenfuhrer, J. (2023). topGO: Enrichment Analysis for Gene Ontology. R package version 2.54.0. 10.18129/b9.bioc.topgo.

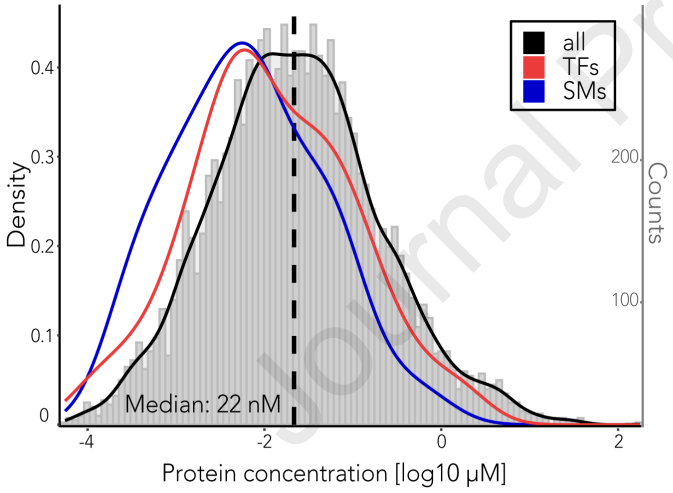
A



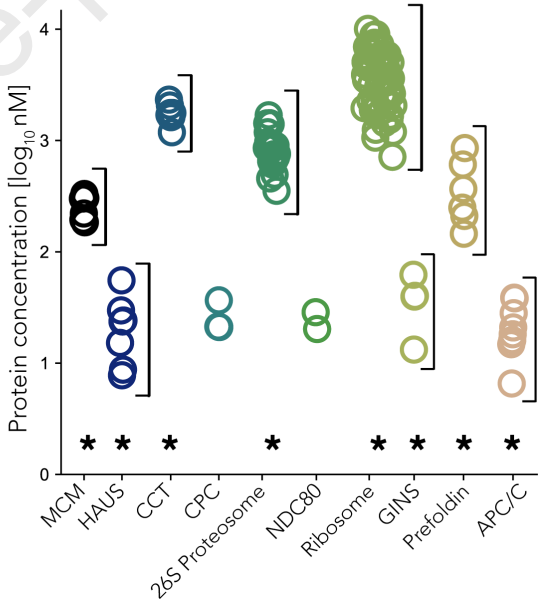
B

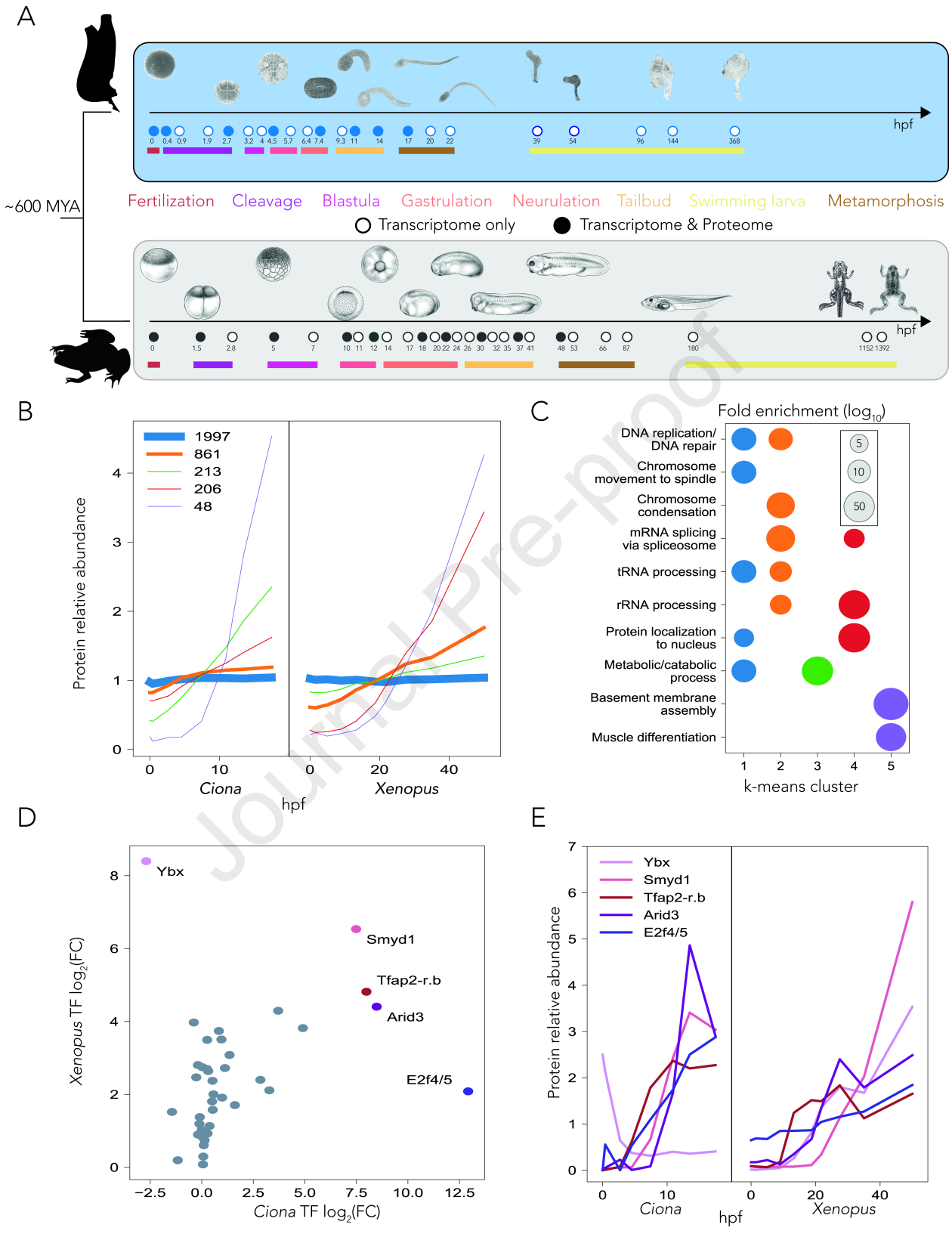
Rank	Ciona gene description	Concentration
1	Vitellogenin	180 µM
2	ATP synthase (alpha)	37 µM
3	Actin	36 µM
4	ATP synthase (beta)	36 µM
5	Ribosomal stalk subunit P2	35 µM
4071	Paired box 2/5/8.b	8.5 nM
4493	GATA binding protein a	3.6 nM
6083	Forkhead box p	~1 nM

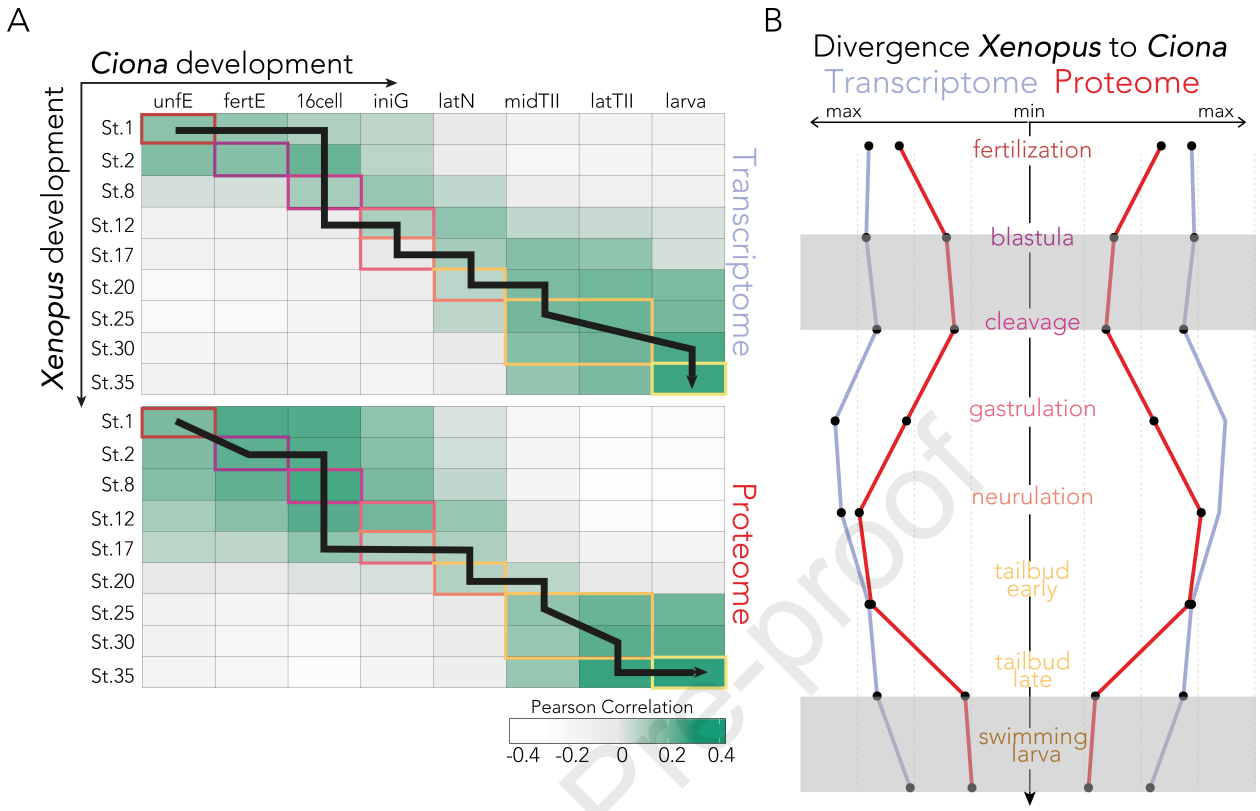
C



D







Highlights

- Resource of absolute concentration for ~6,000 proteins in the *Ciona* egg.
- Comprehensive quantitative analysis of ~7,000 proteins during *Ciona* development.
- Embryonic protein dynamics are evolutionarily more conserved than those of mRNA.
- Cross-species protein dynamic comparison supports an inverse hourglass model.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Ciona robusta formerly *Ciona intestinalis* type A	San Diego, USA	N/A
Chemicals, peptides, and recombinant proteins		
Pierce Protease Inhibitor Mini Tablets, EDTA Free	Thermo Scientific	Cat#PI88666
Lysyl Endopeptidase, MS Grade (Lys-C)	Wako Pure Chemical	Cat#125-05061
Sequencing Grade Modified Trypsin	Promega	Cat#V5111
RNase A, DNase and protease-free	Thermo Scientific	Cat#EN0531
Trypsin Protease, MS Grade	Thermo Scientific	Cat#90305
TRI Reagent	Sigma-Aldrich	Cat#93289
TMTsixplex Isobaric Label Reagent Set	Thermo Scientific	Cat#90062
Sep-Pak C18 1 cc Vac Cartridge	Waters	Cat#WAT054955
Pierce C18 Spin Tips & Columns	Thermo Scientific	Cat#84850
TURBO DNase	Invitrogen	Cat#AM2238
Critical commercial assays		
Quick Start Bradford Protein Assay Kit 1	Bio-Rad	Cat#5000201
Proteomics Dynamic Range Standard Set	Sigma-Aldrich	Cat#232-650-8
RNA Clean & Concentrator Kit	Zymo	Cat#R1017
PrepX RNA-Seq for Illumina Library Kit	Takara Bio	Cat#640097
Pierce BCA Protein Assay Kits	Thermo Scientific	Cat#23225
Deposited data		
Raw and analyzed RNA-seq data	This paper	GEO: GSE237005
Raw proteomics data	This paper	ProteomeXchange: PXD043619
Ciona bulk RNA-seq	Reeves et al.[39]; Kaplan et al.[67]; Sharma et al.[68]; Wang et al.[69]	NCBI SRA PRJNA376667,PRJNA508201,PRJNA498494,PRJNA529900
KH *Ciona* Transcriptome	ANISEED[126]	https://aniseed.fr/
Homo sapiens proteome	Uniprot[70]	Proteome ID: UP000005640
Gallus gallus proteome	Uniprot[70]	Proteome ID: UP000000539
Xenopus tropicalis proteome	Uniprot[70]	Proteome ID: UP000008143
Danio rerio proteome	Uniprot[70]	Proteome ID: UP000000437
Branchiostoma floridae proteome	Uniprot[70]	Proteome ID: UP000001554
Strongylocentrotus purpuratus proteome	Uniprot[70]	Proteome ID: UP000007110

Ciona robusta proteome	Uniprot[70]	Proteome ID: UP000008144
KY21 *Ciona* proteome	Satou et al.[72]	http://ghost.zool.kyoto-u.ac.jp/download_ht.html
Xenopus laevis v10.1 proteome	NCBI	NCBI RefSeq assembly GCF_017654675.1
Ciona time-series RNA-seq data	Hu et al.[53]	NCBI SRA PRJDB3785
Xenopus laevis time-series RNA-seq data	Hu et al.[53] ;Session et al.[115]	NCBI SRA PRJDB3785, PRJNA296953
UPS2 proteomics standards FASTA file	Sigma-Aldrich	https://www.sigmaaldrich.com/deepweb/assets/sigmaaldrich/marketing/global/fasta-files/ups1-ups2-sequences.fasta
Software and algorithms		
Mass Spec Protein Reference Tool	Wühr et al.[64]	https://kirschner.med.harvard.edu/tools/mz_ref_db.html
Python	Python Software Foundation	https://www.python.org
BLAST (version 2.10.1)	Altschul et al.[130]	https://blast.ncbi.nlm.nih.gov/doc/blast-help/downloadblastdata.html; RRID:SCR_001653; RRID:SCR_001010
Trinity (version 2.11)	Grabherr et al.[125]	https://github.com/trinityrnaseq/trinityrnaseq/releases; RRID:SCR_013048
SeqClean	Dana-Farber Cancer Institute	https://sourceforge.net/projects/seqclean/files/
RepeatMasker (version 4.1)	Smit et al.[127]	https://www.repeatmasker.org/RepeatMasker/; RRID:SCR_012954
TGICL (version 2.1)	Pertea et al.[128]	https://sourceforge.net/projects/tgicl/files/tgicl%20v2.1/

CAP3	Huang et al.[129]	https://faculty.sites.iastate.edu/xqhuang/cap3-assembly-program; RRID:SCR_007250
CD-HIT (version 4.8.1)	Fu et al.[131]; Li et al.[132]	https://github.com/weizhongli/cdhit; RRID:SCR_007105
R (gProfiler, topGo)	Kolberg et al.[142]; Alexa et al.[147]	RRID:SCR_006809; RRID:SCR_014798
FastQC (version 0.12.0)	Babraham Bioinformatics	https://github.com/s-andrews/FastQC; RRID:SCR_014583
Trimgalore (version 0.6.10)	Babraham Institute	https://github.com/FelixKrueger/TrimGalore; RRID:SCR_011847
Salmon	Patro et al.[145]	https://github.com/COMBINE-lab/salmon; RRID:SCR_017036
Other		
Genome annotation files, transcription factor and signaling molecules databases used for RNA-seq and proteomics analyses, alignment files used in orthology assignment and other additional files	This paper	https://github.com/andreamariossi/proteome_ciona
Adapted *Ciona* schematics	Hotta et al.[123]	https://chordate.bpni.bio.keio.ac.jp/chordate/faba/1.4/top.html
Xenopus illustrations	Xenbase, Zahn et al.[137]	https://www.xenbase.org/xenbase/zahn.do