Transitions in the proteome and phospho-proteome during *Xenopus laevis* development.

Elizabeth Van Itallie[1], Marian Kalocsay[1,2], Martin Wühr[1,3,4], Leonid Peshkin[1,5#], Marc W. Kirschner[1#]

1. Department of Systems Biology, Harvard Medical School, Boston, MA 02115
2. Laboratory of Systems Pharmacology, Harvard Medical School, Boston, MA 02115
3. Department of Molecular Biology, Princeton University, Princeton, NJ 08544
4. The Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ 08544
5. Eugene Bell Center, Marine Biological Laboratory, Woods Hole, MA 02543

[#] Correspondence should be addressed to MWK at marc@hms.harvard.edu or LP at leonid_peshkin@hms.harvard.edu

ABSTRACT

Vertebrate development from an egg to a complex multi-tissue million-cell organism is supported by multiple phases of genome-scale remodeling of the repertoire of proteins and their post-translational modifications, yet so far we know little about these phases. In this paper we present comprehensive characterization of these processes reflected by eleven time points, approximately eleven thousand proteins, and six thousand phospho-forms in two replicates. We find that the most dramatic changes to the proteome occur during the transition to functional organ systems, which occurs as the embryo becomes a tadpole. At that time the absolute amount of non-yolk protein increases two-fold, and there is a shift in the balance of expression from proteins regulating gene expression to receptors, ligands, and proteins involved in cell-cell and cell-environment interactions. Between the early and late tadpole, the median increase of membrane and secreted proteins is substantially higher than that of nuclear proteins. For the first time, we have included quantitative phospho-proteomic data across the same developmental stages. In contrast to the significant protein changes that are concentrated at the end of the time series, the most significant phosphorylation changes are concentrated in the very early stages of development. A clear exception are phosphorylations of proteins involved in gene expression; these increase just after fertilization, with patterns that are highly correlated with the underlying protein changes. To improve our interpretation of this unique data set, we created a pipeline for identifying homologous human phosphorylations from the measured Xenopus phospho-proteome, and we estimated, where possible, the occupancy of phosphorylation sites. Overall, we detected many profound temporal transitions, which suggest concerted changes in developmental strategies in the embryo that are particularly pronounced once early patterning and specification are complete.

INTRODUCTION

How an egg develops into an organism is a fundamental question that was addressed by morphology and comparative anatomy in the 19th century and from a molecular genetic viewpoint in the late 20th century, when many regulatory genes and pathways were identified. More recently, transcriptional profiling has revealed developmental trajectories at the level of gene expression. However, embryos are built of proteins, not genes, so a full description of development requires analysis of changes in the amount, as well as the activity, of all proteins over time. Changes in mRNA have been used as a proxy for changes in protein, but we know that mRNA levels are only partly predictive of protein levels, and mRNA levels give no information about changes in protein activity by post-translational modification.

Mass spectrometry based proteomics with isobaric tags and MS^3 measurement can quantify many proteins across different timepoints [1]. *Xenopus laevis* eggs and embryos are an ideal model for application of this technology to embryogenesis, because they are large (1.2 mm diameter), the yolk can be removed, a single egg has 25 ug of non-yolk protein, and large numbers of staged samples can be collected easily [2–6]. Previously, we reported protein levels across development from the egg through the tailbud stage [6,7] and found the temporal trajectories of most proteins almost flat across this period. However, those data only reported on ~6,000 proteins at six timepoints with no replicates [7]. That study also excluded the majority of organogenesis. Additionally, the measured time period did not extend to the juvenile tadpole. Studies of recruitment of ribosomes into polysomes [8], translation of ribosomal protein mRNA [9], and yolk consumption [10] predict that the greatest amount of protein increase occurs between the late tailbud embryo and juvenile tadpole. Here, we greatly extend the time course and number of proteins quantified to report the first full proteome of a developing embryo.

To properly understand how the proteome directs development, we need to measure not just the amount of all proteins, but also their activity state. No single measurement can do this, but phosphorylation comes closest. Almost half the human proteome is phosphorylated [11], making it by far the most significant post-translational modification for regulating protein activity. Our lab and others previously reported phospho-proteomes during oocyte maturation through the first cleavage, which are dominated by cell cycle changes and events of meiosis [3,12]. However, these studies are not informative for the role of protein phosphorylation in embryonic development. One study included gastrula stage embryos for identification of phosphorylations [13], but no quantitative studies of the phospho-proteome after the cleavage stages have been published. However, measuring and interpreting phospho-proteomics data from a complex embryo has challenges beyond those encountered in protein level proteomics. These include the combined heterogeneity of both protein and phosphorylation level differences across the embryo, the limitations in detection and identification of phosphorylated peptides [14,15], and the dearth of information on how phosphorylation of a specific residue changes protein function or localization [16]. For these reasons, we did not know in advance what we would learn from phospho-proteomics of a developing embryo.

This paper presents the most comprehensive measurement of the proteome and phospho-proteome during embryogenesis to date. It provides an accessible resource for further study and reveals biochemical correlates of developmental processes that have heretofore been characterized principally at the anatomical and transcriptional level.

RESULTS

Data Collection

We measured the proteome and phospho-proteome of *Xenopus laevis* embryogenesis at the stages indicated in *Figure 1A*. We performed the full analysis in duplicate, collecting embryos from two independent clutches. Digested peptides were labeled with tandem mass tag reagents (TMT), and combined for enrichment of phosphorylated peptides using immobilized metal affinity chromatography (*Methods*) [14]. The measurements of the phospho-enriched and phospho-depleted peptides were made with Multi-Notch MS^3 quantification (*Methods*) [1,17].

Data Reliability

After mapping peptides to the *Xenopus laevis* reference XL1.8.3.2 and considering all measurements together when controlling for identification false discovery (2%), we measured a union of 14,545 proteins and 6,735 phospho-forms with 85% and 36%, respectively, measured in both replicates (*Figure 1.B.i,ii*; *Methods*). Phosphorylations were measured on a union of 3,056 proteins. This number is smaller than the number of phospho-forms because 66% of phospho-proteins are measured with more than one phospho-form (*Supplemental Figure 1.A, B.*). For Replicates A and B, 99.8% and 95%, respectively, of phospho-form protein references have a protein measurement as well (*Supplemental Figure 1.C.*). Thus, we measured the proteome with sufficient depth to fully interpret the phospho-proteome measurements.

We next assessed the reproducibility of our replicates. We computed the Pearson Correlation Coefficient (PCC) for each timepoint in Replicate A against all timepoints in Replicate B (*Supplemental Figure 2.A,B.i*). The most similar sample is the other replicate at that time point. We also computed the PCC for all replicate protein and phospho-form trends compared to a scrambled background (*Supplemental Figure 2.A,B. ii*). The higher reproducibility of the protein data is consistent with the fact that we are measuring multiple peptides per protein, whereas the increased noise for the phospho-forms is consistent with measuring only a single peptide. The PCC of the trends in clutch A and B of CBX3 protein, a component of heterochromatin [18], and phosphorylation S-87 on CBX3 are the medians of the distributions of this metric for proteins and phospho-forms (*Figure 1.C.i, ii*).

The most frequent protein trajectories map to the expected progression of developmental processes.

To identify shared trajectories of protein change through development, we clustered all proteins using k-means clustering (*Figure 2.A*; See *Supplemental Figure 3* for the original clusters). The first two clusters are the flattest, and the following six clusters are ordered by the timing of their most prominent temporal event. Fifty-three percent of proteins are in the flat clusters (Clusters 1 and 2), consistent with previous findings from our laboratory [7]. Cluster 3 represents the very small number of proteins that decrease precipitously at oocyte maturation (Stage VI oocyte to egg). Cluster 4 represents the slightly larger number that decrease following fertilization during the period of cleavage. They are low and flat after the

mid-blastula transition (Stage 8). Twenty-five percent of proteins are in cluster 5, which increases gradually following fertilization and remains high until late tailbud, when it slowly declines. Clusters 6, 7, and 8 all increase dramatically no earlier than after the last tailbud stage (Stage 30).

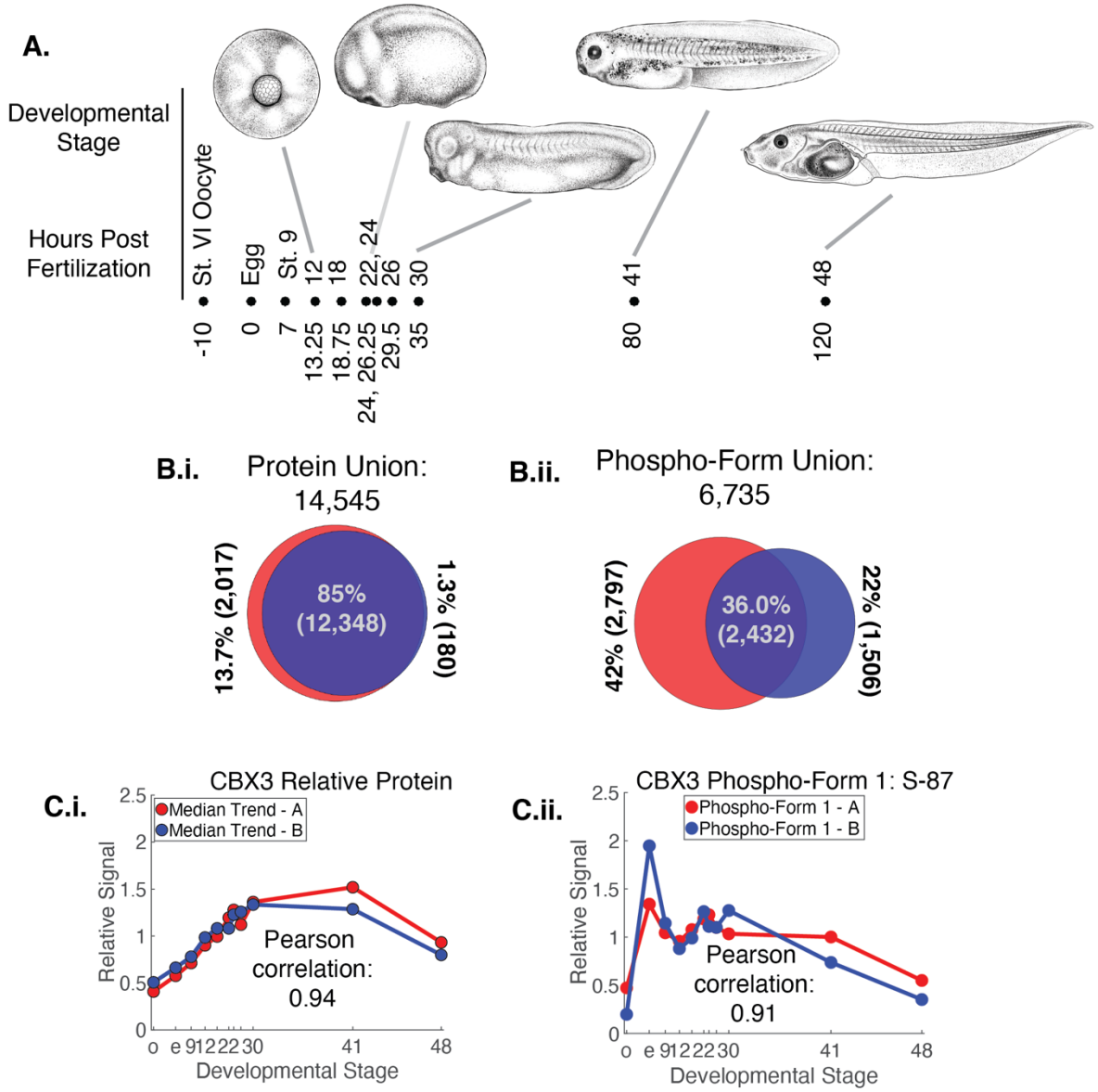


Figure 1. Collection of proteomic and phospho-proteomic data.

A. The eleven *Xenopus laevis* developmental stages collected for proteomic and phospho-proteomic measurement spaced according to Nieuwkoop and Faber times for temperatures between 22-24℃. (The Stage 41 illustration is actually Stage 40 for convenience. See Methods for Illustration information)

B. **(i)** 14,545 proteins were quantified; 85% measured in both replicates.
(ii) 6,735 phospho-forms were quantified; 36% were measured in both replicates.

C. Examples: **(i)** CBX3 proteins and **(ii)** CBX3 phospho-form S-87 trends showing the median Pearson Correlation Coefficient for the distribution of replicate proteins and phospho-forms, respectively.

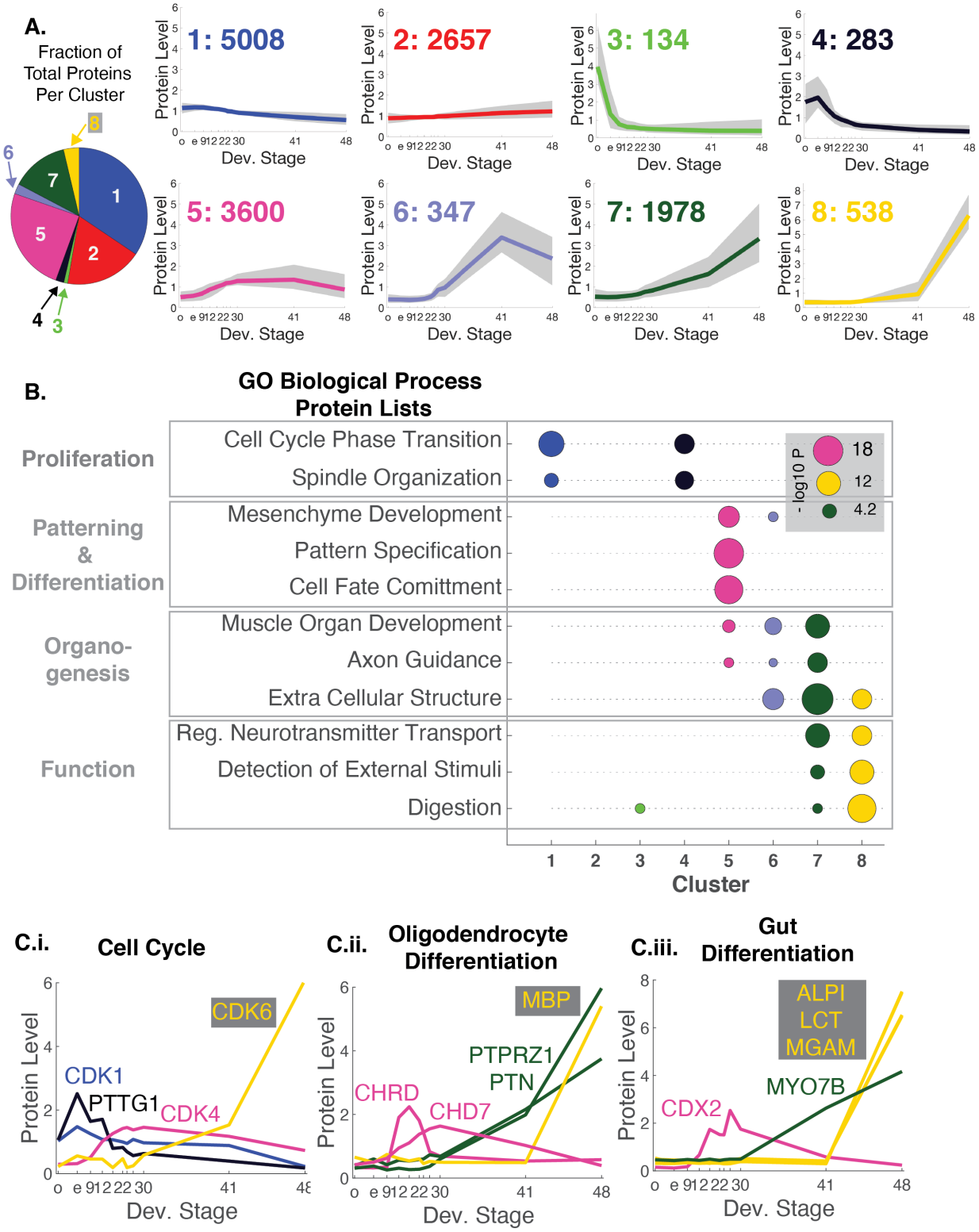


Figure 2. Almost fifty percent of proteins have dynamic relative trends; and their trends are consistent with the progression from proliferation, through specification and organogenesis, to function.

Figure 2 continued

A. Ten k-means clusters of relative protein trends were merged into eight summary trends ordered by the time that the first dynamic event in the trend occurs. The median trend of the cluster is the colored line; the ten to ninety percentile region is shown in light gray. The pie chart shows the fractions of total proteins in each cluster.

B. Over-representation analysis of biological process GO gene sets shows that as development progresses dynamic clusters are associated with proliferation, patterning and differentiation, organogenesis, and tissue function. For each GO protein list shown in black, the p-value for the over-representation of those proteins in the gene set relative to all proteins is shown with a circle, whose area is proportional to the -log10 p-value (hypergeometric) and has the appropriate cluster color. All GO protein lists are overrepresented in at least one cluster with FDR <= 5%. (See *Supplemental Table 1* for the entire set of over-represented sets including those over-represented in clusters 1 and 2).

C. **(i)** Relative protein levels consistent with the transitions in the mitotic cell cycle across development. Securin/PTTG1 is highest in the metaphase arrested egg and decreases following fertilization. CDK4 and CDK6 levels increase following Stage 9 as the embryo enters the period where entry into S phase is under regulation.

(ii, iii) Expression of proteins connected to tissue differentiation and organogenesis: for oligodendrocytes **(ii)** and gut **(iii).** The plots illustrate the temporal progression through these phases, as supported by the GO overrepresentation analysis.

We next analyzed the clusters using Gene Ontology (GO). Selected gene sets that are overrepresented in the clusters that best represent known developmental processes are illustrated in *Figure 2.B* (*Methods; Supplemental Table 1* for all overrepresented GO sets). Cell cycle proteins were notably overrepresented in cluster 4. The precise synchrony of early stages makes it easy to measure these cell cycle dependent changes at the protein level. Several of the cyclin proteins -- Cyclins B1, B2, E1, A1 -- are in cluster 4, where the levels drop as the cell cycle slows down after the mid-blastula transition (MBT) and through gastrulation [19,20]. The same is true for other cell cycle proteins whose levels oscillate during the cell cycle, such as Securin (PTTG1, *Figure 2.C.i.*) [21]. The decrease in mitotic protein levels after the egg in cluster 4 is consistent with both the loss of synchrony and the decrease in the mitotic index. Mitotic proliferation related gene sets are not overrepresented in any of the later clusters; however, proteins involved in regulation of the G1/S transition, such as Cdk4, cyclin D, and RB1, are found in cluster 5, suggesting that regulated cell proliferation continues at a high level throughout development. The gene sets “cell cycle phase transition” and “spindle organization” are also overrepresented in the first flat cluster, including the master cell cycle kinase CDK1, the CDK1 regulator ENSA, the anaphase promoting complex protein CDC16, the aurora kinase regulator BORA, and the centralspindlin kinesin KIF23.

After the mid-blastula transition (MBT), transcription begins in earnest and differentiation commences. “Pattern specification” and “cell fate commitment” gene sets are both overrepresented in cluster 5. The gene sets associated with organogenesis and specific cellular functions (“function”) are overrepresented in more than one cluster. “Muscle organ development” and “axon guidance” are overrepresented in all of the clusters that increase after fertilization except for the cluster that increases after the early tadpole stage. This is consistent with the fact that by Stage 41 the tadpole musculature has been established and the development of axons and nerve fibers is extensive [22]. The gene sets related to organ function, such as “regulation of neurotransmitter transport,” “detection of external stimuli,” and “digestion,” are all overrepresented in the last two clusters, which increase in level from the tadpole at Stage 41 to the true juvenile at Stage 48. This is the period when we expect specific cellular functions to emerge in organs. Examples of the temporal patterns of specific proteins of oligodendrocyte differentiation (*Figure 2.C.ii*) and gut differentiation (*Figure 2.C.iii*) exemplify progression of differentiation leading to organ function. For example, the neural lineage progenitors including glial cells are physically located in the dorsal part of the embryo. Chordin (CHRD) is a secreted protein in the Mangold-Spemann organizer that patterns the dorsal ventral axis during gastrulation and neurulation, particularly the brain and spinal cord [23]. CHD7 is a bromodomain ATP dependent remodeling protein that stabilizes gene expression and is involved in specifying the oligodendrocyte lineage [24]. PTPRZ1 is a receptor protein tyrosine phosphatase that inhibits embryonic oligodendrocyte differentiation in mice, but its inhibitory function is blocked by the ligand pleiotrophin (PTN) [25]. Myelin basic protein (MBP) is one of the most abundant proteins in myelinated sheaths, which are the terminal differentiated product of the oligodendrocyte lineage. Protein levels of MBP increase only after Stage 41. This is consistent with mRNA levels of the brain isoforms of MBP not being detectable at Stage 40, and MBP protein being detectable in the Stage 47 spinal cord [26]. In contrast to neural progenitors, the gut is a ventral organ. CDX2 is involved in transcriptional regulation of the posterior gut epithelium [27,28]. MYO7B is an unconventional myosin that is expressed in microvilli of enterocytes

in the posterior gut in mice (and also in the kidney) [29]. Intestinal type alkaline phosphatase (ALPI), lactase-phlorizin hydrolase (LCT), and intestinal maltase-glucoamylase (MGAM) are all digestion enzymes found in the small intestine. Food is first found in the gut at Stage 46 [22]. Thus, the measured protein dynamics are consistent with the progression from patterning and fate specification (CHRD, CHD7, CDX2) at the transcriptional level to regulated commitment and emergence of structure (PTPRZ1, PTN, MYO7B), to juvenile function (MBP, ALPI, LCT, MGAM). We expect that specific proteins that follow a similar trajectory in other lineages could be found in the data and may be identified by a combination of single cell transcriptional analysis and proteomics.

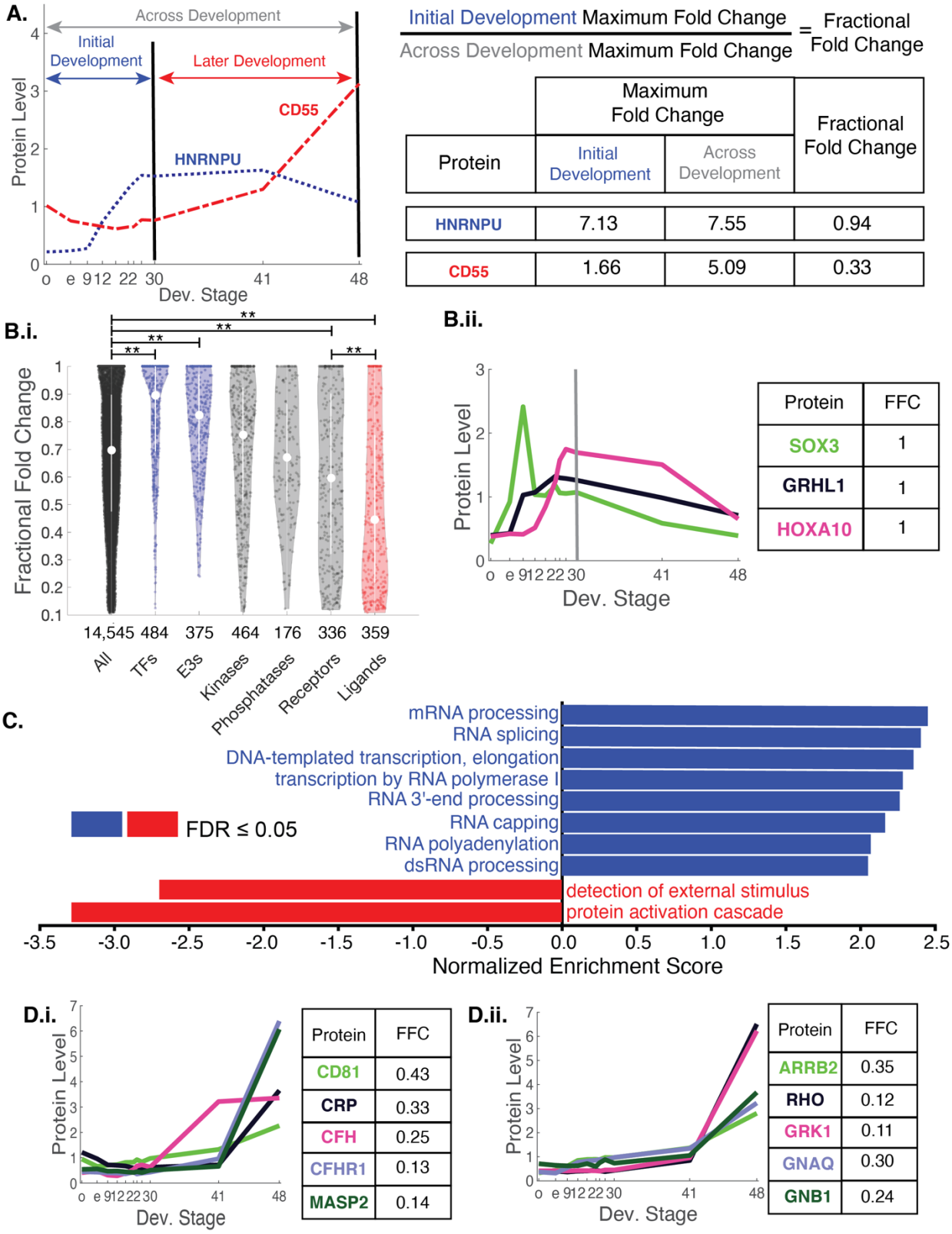
Major proteome changes during organogenesis

There is a natural subdivision of the proteomic dataset into two periods: proteins involved in proliferation, patterning, and specification in the first period, and those acting in organogenesis in the latter period. Based on the median trends of the clusters in Figure 2, we chose Stage 30 as the breakpoint between the two periods. For each protein in the database we can ask what fraction of the overall magnitude of protein level change happens before Stage 30 as compared to the overall change from the oocyte to the feeding tadpole, Stage 48 (*Figure 3.A*; *Supplemental Figure 4.A*). The ratio of these two values, which we call the fractional fold change (FFC), will be greater than or less than 0.5 if the majority of the change is achieved before or after Stage 30. *Figure 3.A* illustrates these metrics for the nuclear RNA binding protein HNRNPU and the complement pathway receptor CD55. The FFC value of HNRNPU is very close to one, whereas that of CD55 is 0.33. Thus, for HNRNPU more change occurs before Stage 30, while the opposite is true for CD55.

We calculated the FFC for different classes of regulatory proteins: transcription factors (TFs), ubiquitin ligases, kinases, phosphatases, receptors, and ligands [30] (*Figure 3.B.i*). The median FFC for TFs, ubiquitin ligases, receptors, and ligands are all significantly different from the median of all proteins. The median FFCs of TFs and ubiquitin ligases are larger than the median of all proteins (0.70) and higher than 0.5 (0.90 and 0.82, respectively). This is consistent with TFs and ubiquitin ligases regulating gene expression and patterning events that occur before Stage 30, and with ubiquitin ligases also regulating progression through the cell cycle, which occurs before Stage 30 [31–33] [34,35]. As examples, the transcription factors SOX3, GRHL1, and HOXA10 used for early neural specification, epithelial development, and anterior-posterior organization/myeloid differentiation, respectively, all have FFCs equal to one (*Figure 3.B.ii*) (Specific examples for ubiquitin ligases are in *Supplemental Figure 4.B.*) In striking contrast to TFs and E3s, receptors and ligands have median FFCs that are smaller than that of all proteins (0.60 and 0.45, respectively, as compared to 0.7).

To determine which other biological processes are associated with the periods before, as opposed to after, Stage 30, we performed Gene Set Enrichment Analysis using the FFC as our quantitative metric. Multiple GO biological process gene sets related to transcription, mRNA processing and metabolism, and ribosomal RNA transcription ("transcription by RNA polymerase I") are statistically significantly enriched for FFCs closer to one compared to the set of all measured proteins (*Figure 3.C.*). This is consistent with the distribution of transcription factor FFCs and the obvious importance of regulation of gene expression for patterning and differentiation early in development. It is also known that transcription of ribosomal rRNA begins during gastrulation (before Stage 30) [36].

The GO gene sets that have FFCs closer to zero (changes late in development) include "protein activation cascade" and "detection of external stimulus." The "protein activation cascade" set includes many secreted proteins (e.g. CRP, CFH, CFHR1, and MASP), and both gene sets include receptors (e.g. CD81, RHO, GRK1) (*Figure 3.D*). The GO set "detection of external stimulus" includes G-protein coupled receptor signaling proteins (e.g. ARRB2, GNAQ, GNB1). At the cellular level these proteins have to do with intracellular/extracellular communication, and, on the organismal scale, with interaction of the embryo with its environment. This is consistent with the basic difference between the Stage 30 tailbud embryo and Stage 48 tadpole -- a difference not in patterning or specification, but in the capacity to interact with the external environment (sensory organs) and particularly to interact with the internal environment, such as required for a functional digestive system, nervous system, and immune system. By Stage 30 physiological function has replaced establishment of organization as the preoccupation of the organism.



Protein	Maximum Fold Change: Initial Development	Maximum Fold Change: Across Development	Fractional Fold Change
HNRNPU	7.13	7.55	0.94
CD55	1.66	5.09	0.33

Protein	FFC
SOX3	1
GRHL1	1
HOXA10	1

Protein	FFC
CD81	0.43
CRP	0.33
CFH	0.25
CFHR1	0.13
MASP2	0.14

Protein	FFC
ARRB2	0.35
RHO	0.12
GRK1	0.11
GNAQ	0.30
GNB1	0.24

Figure 3. Gene expression proteins change in level before Stage 30, and ligands and proteins involved in cell-cell and cell-environment communication change mainly after Stage 30.

Figure 3 continued

A. Introducing the fraction fold change metric (FFC). The measured time period is split at Stage 30 into "initial development" (blue) and "later development" (red). In order to determine how much of the maximum fold change across the entirety of development occurs during "initial development," we divide to maximum fold change during initial development by the maximum fold change across development. We illustrate the metric with the two examples from the data – HNRNPU and CD55 – that have very different fraction fold changes.

B. **(i)** Violin plots of the FFCs for the six different classes and all proteins (white circle = median; white line = interquartile region). The medians of transcription factors (TFs), E3 ubiquitin ligases (E3s), receptors and ligands are statistically significantly different than that of all proteins (** = $p < 0.001$; Kruskal Wallis test, Tukey-Kramer critical value for multiple comparisons).

(ii) Examples of transcription factor relative trends with FFCs equal to one.

C. Gene Set Enrichment Analysis (GSEA) results for the FFC metric using Biological Process GO sets shows that processes related to gene expression are statistically significantly enriched for FFCs closer to 1 with the background of all measured proteins. In contrast, processes related to cell-cell and cell-environment interactions are statistically significantly depleted for FFCs closer to 1. All gene sets shown are statistically significantly enriched or depleted with FDR <= 5%. See *Supplemental Table 2* for complete list of enriched and depleted GO sets.

D. Protein trends for some well-known members of the **(i)** protein activation cascade and **(ii)** detection of external stimulus gene sets.

Embryonic growth: quantitative changes in non-yolk protein between the late tailbud and juvenile tadpole

We have up to this point considered the relative, not the absolute, changes of proteins across development. However, by combining the relative protein changes with the changes in mass of total non-yolk protein we can calculate the absolute protein abundances across development. The total amount of protein per embryo (average of 10 embryos) after removal of yolk for all of the timepoints in our proteomics dataset except for the oocyte is shown in *Figure 4.A* (*Methods*). Yolk, having been endocytosed from protein produced in the mother's liver, is held until neurulation before it begins to break down [10]. In the egg, 85% of protein is yolk (*Supplemental Figure 5*). By Stage 48 there is no detectable yolk protein. In early embryogenesis before Stage 30, there is no measurable increase in non-yolk protein amount in the embryo. Later the average amount of non-yolk protein increases 1.3-fold between Stage 30 and Stage 41, and 1.5-fold between Stage 41 and Stage 48, for an overall two-fold change from Stage 30 to 48. If all the yolk protein were converted into amino acids and if all the amino acids were converted into new protein, we would expect protein content to increase about 7.5-fold, not 2-fold. One explanation for this discrepancy could be that the amino acids derived from yolk are mostly metabolized as an energy source. This would be consistent with the finding in a different amphibian that excretion of nitrogen increases 10-fold at the time of embryo hatching (Bialaszewicz and Mincowna, 1921) [37].

To convert the relative changes in protein to the molar amount of protein change for the proteins in our dataset, we cross-referenced our data with the reference of absolute protein concentrations in the egg to determine the relationship between mass spectrometry signal and protein concentration [2] (*Methods; Supplemental Figure 6*). Using these measurements, we are able to determine the number of moles for each protein in our dataset and the change in the total amount of moles of protein across our time series for the measured portion of the proteome (*Figure 4.B.*; *Methods*). This method of calculation shows a change of 2.3-fold, slightly greater than the mass measurement.

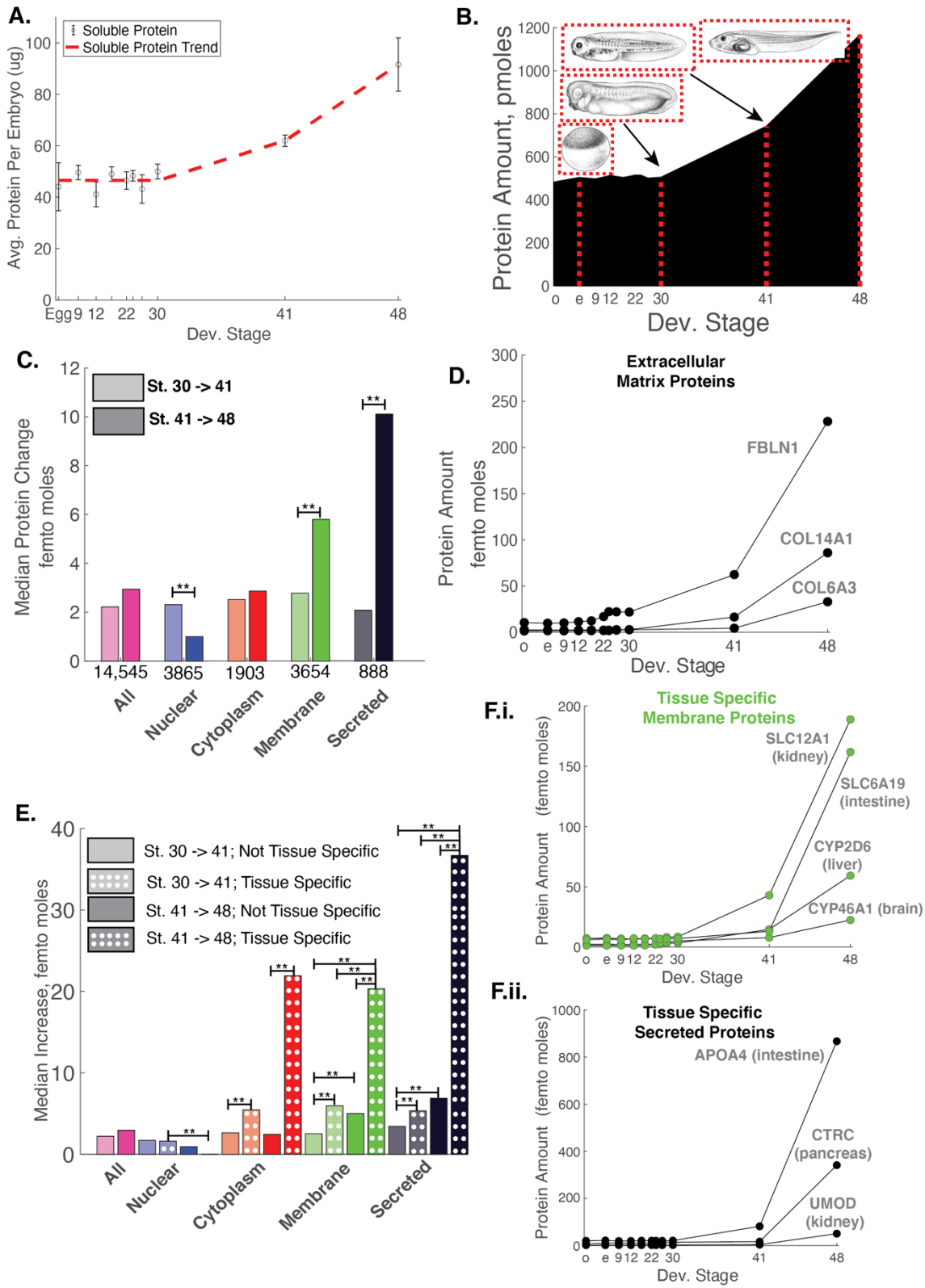


Figure 4. The proteome increases two-fold in mass between the late tailbud and tadpole stages. Tissue specific proteins contribute disproportionately to this increase, especially between the early and late tadpole.

Figure 4 continued

A. Amount of protein per egg/embryo (average of ten) at ten developmental stages. The mean of three technical replicates is shown with 95% confidence intervals.

B. The total amount of protein, in picomoles, at the measured developmental stages. Red lines show the developmental stages of the embryo drawings.

C. Median absolute change of proteins by annotated localization for Stage 30 to 41 (lighter shade) and Stages 41 to 48 (darker shade). Nuclear, membrane, and secreted proteins have different median changes between the two time periods. (** = $p < 0.001$; Kruskal Wallis test, Tukey-Kramer critical value for multiple comparisons).

D. Absolute protein amount trends for three different extracellular matrix proteins.

E. Median absolute change in femtomoles for the localization classifications additionally stratified by tissue specificity (not tissue specific: no white line, tissue specific: white line) for Stages 30 to 41 (lighter shade) and Stages 41 to 48 (darker shade). Tissue specific cytoplasmic, membrane, and secreted proteins have larger median changes than non-specific proteins for the same localization and time period.

F. The absolute amount of protein across development for three tissue specific **(i)** membrane and **(ii)** secreted proteins.

Changes between early and late organogenesis

The amount of protein increase between Stages 30 and 41 is not very different from the amount of protein increase between Stages 41 and 48. We wanted to know whether the composition of protein changes during these periods is also similar. Our analysis of regulatory protein classes revealed that receptors and ligands are more dynamic after Stage 30 than the set of all measured proteins. Inspired by this finding, and by the desire to include larger numbers of proteins in our analysis, we decided to investigate classes of proteins defined by localization. We classified 71% of proteins as either nuclear, cytoplasmic, membrane, or secreted (*Methods*). Between Stages 41 and 48, the median change of membrane and secreted proteins is larger than that for all proteins, and the median change between Stages 41 and 48 is larger than that between Stages 30 and 41 (*Figure 4.C*). Examples of secreted proteins that show large molar increases between Stages 41 and 48 are extracellular matrix proteins fibulin-1 (FBLN1), collagen alpha-1(XIV) chain, and collagen alpha-3(VI) chain (*Figure 4.D*). In contrast to membrane and secreted proteins, the median change of nuclear proteins is much smaller between Stages 41 and 48 than between Stages 30 and 41. The median increase for cytoplasmic proteins is very similar. Thus, despite the similarity in change of total protein, the composition of change is different for the two periods. Between Stages 30 and 41, membrane and secreted protein constitute 42% of the protein increase; between Stages 41 and 48, they constitute 59%.

The late changes observed in our data likely reflect multiple organs and their constituent cell types initiating their physiological functions between Stages 41 and 48. We examined how the increase in protein levels relates to tissue specificity. In this broad characterization, we classified proteins as non-tissue-specific if they were either present in the egg or *not tissue enriched* in adult organs. We classified proteins as tissue-specific if they were both not present in the egg and *tissue enriched* in adult organs [38] (*Methods*). By these criteria, 6% of proteins in the dataset are tissue specific (*Supplemental Table 3*). Despite this small overall number, tissue-specific proteins constitute 19% of the total amount of molar increase between Stages 41 and 48 (*Supplemental Table 4*). Membrane and tissue specific proteins increase more than non-specific ones both between Stages 41 and 48 and between Stages 30 and 41. However, the increase is statistically significantly greater between Stages 41 and 48 than between Stages 30 and 41 (*Figure 4.E*). Two common types of tissue-specific membrane proteins that increase after Stage 30 are cytochrome P450 monooxygenase enzymes (e.g. CYP2D6 in the liver and CYP46A1 in the brain) and solute transporters (e.g. SLC12A1 in the kidney and SLC6A19 in the intestine) (*Figure 4.F.i*). Examples of tissue-specific secreted proteins that increase after Stage 30, but show particularly dramatic increase after Stage 41, include apolipoprotein A-IV (APOA4) in the intestine, chymotrypsin-C (CTRC) in the pancreas, and uromodulin (UMOD) in the kidney (*Figure 4.F.ii*). Cytoplasmic tissue-specific proteins also increase more than non-tissue-specific cytoplasmic proteins during both of these periods. Many of the tissue-specific cytoplasmic proteins are expressed in muscle tissues (e.g. MYH1, MYOM2, MYOZ2) (*Supplemental Figure 7*). The large molar increase of tissue-specific proteins is consistent with the overall increase in proteins from the egg stage, where these proteins are not present at all (or at very low levels) to the juvenile tadpole, where they are at a sufficient concentration to support physiological functions. Though the larger median increase of tissue specific proteins than other

proteins is a feature of the entire period after Stage 30, their contribution to the increasing proteome is much larger after Stage 41.

Taken together, these observations suggest that a dramatic change in the biochemistry of protein production occurs after Stage 30, resulting in increase of the soluble molar proteome, and that a second transition event after Stage 41 results in a change in the composition of the types of proteins that are made during the crucial period when organs initiate their physiological function.

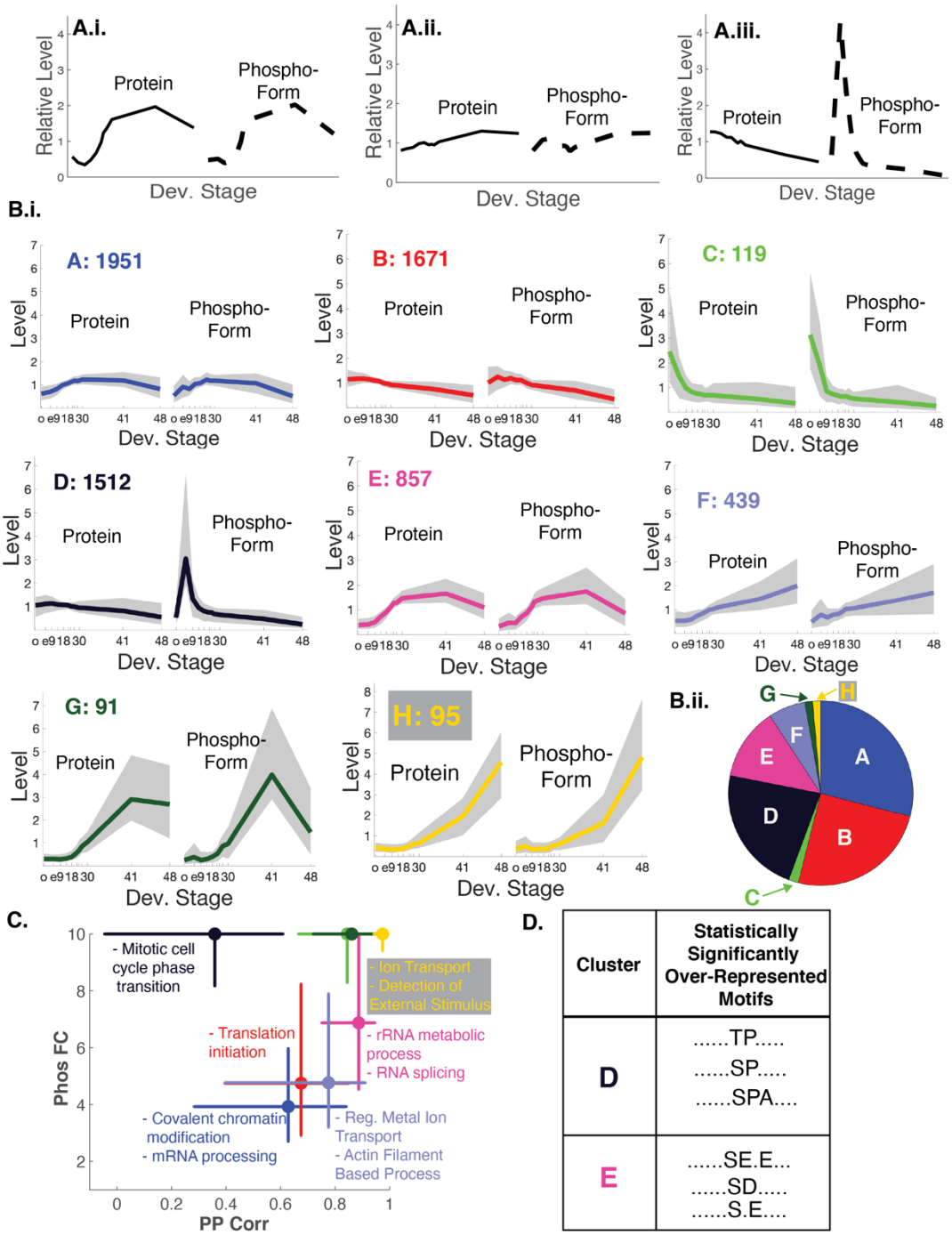


Figure 5. Mitotic cell cycle and gene expression related phosphorylation dominate the phospho-proteome.

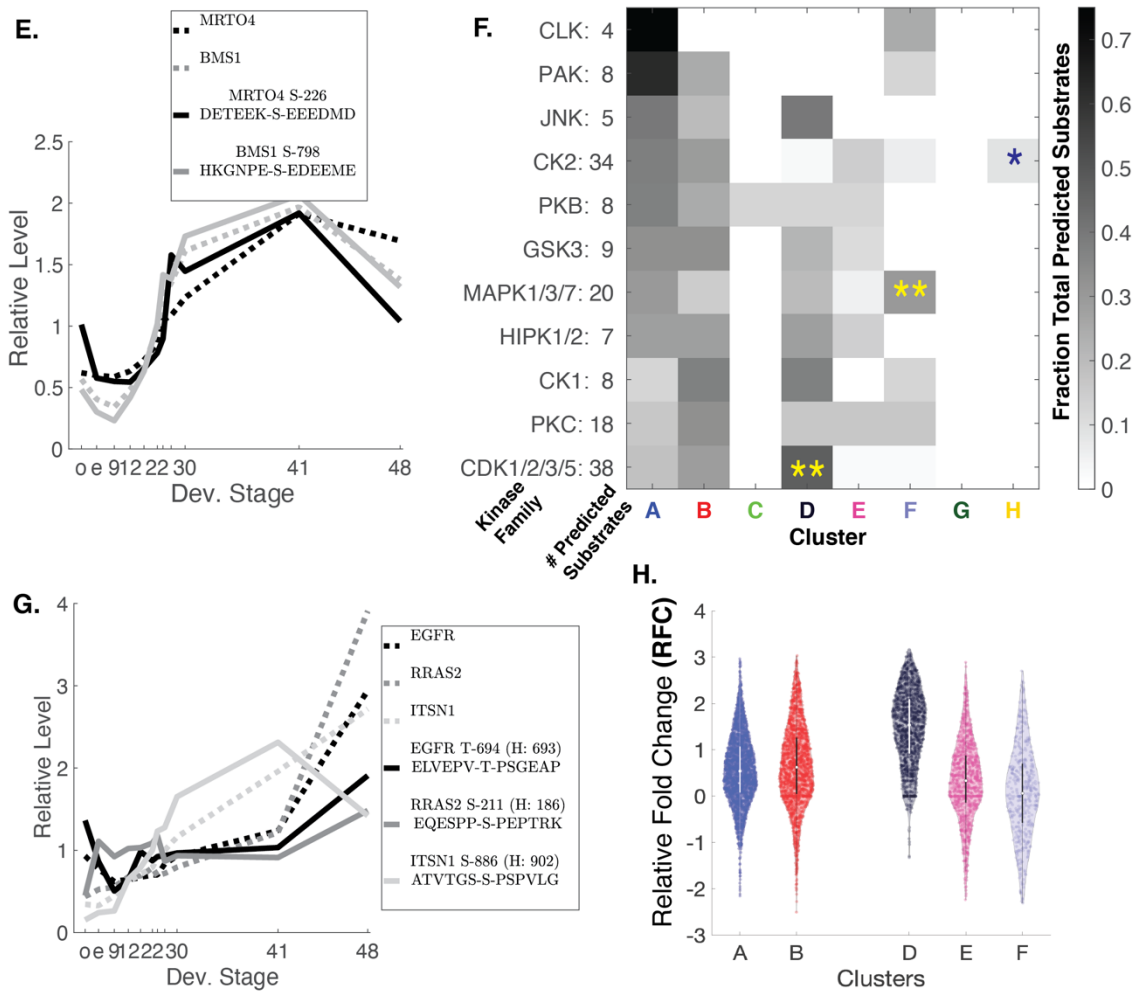


Figure 5 continued

A. Phospho-form dynamics are investigated together with their underlying protein dynamics and without normalization to the protein trend in order to distinguish **(i)** phospho-forms that are similar to their protein trend but have significant changes in magnitude from **(ii)** those that are also similar but do not change very much during development. **(iii)** Other phospho-forms have very different trends than the protein they are measured on.

B. Over 50% of phospho-forms are relatively flat and the majority of dynamic phospho-forms show the majority of their increase before the tadpole stages. **(i)** The protein and phospho-form trends were clustered together as a vector of twenty-two timepoints. Ten k-means clusters were merged into eight summary clusters. Cluster medians (colored lines) and ten to ninety percentile regions (light gray area) are plotted separately for proteins and phospho-forms to aid interpretation. **(ii)** The fraction of the total phospho-forms that are in each cluster is represented by a pie chart where the colors are the same as the cluster medians.

C. Cluster median Pearson correlation between phospho-form and protein and phospho-form maximum fold change across development are shown for all clusters with error bars representing the 25 to 75% interquartile range using the same cluster colors as (B). The maximum allowable phospho-form max. fold change across development is 10. GO Biological Process protein sets that are overrepresented for a given cluster with FDR <= 5% are shown in the same cluster color. See *Supplemental Table 4* for the complete set of overrepresented sets.

D. Cluster D phospho-forms are statistically significantly overrepresented for multiple proline directed motifs, and cluster E phospho-forms are statistically significantly overrepresented for multiple acidic motifs.

Figure 5 continued

E. Representative examples of acidic phosphorylations in Cluster E on proteins involved in ribosome biogenesis in the nucleolus.

F. The fraction of the total predicted kinase family to substrate pairs that are in each of the eight different clusters. The substrates of three families of kinases are overrepresented in specific clusters (yellow or blue star(s); ** = $p < 0.001$; * = $p < 0.05$).

G. The distributions of Dynamic Disparities (DD) for the clusters that comprise more than 5% of the data visualized with violin plots. Cluster D (black) has the largest median, and cluster F has the smallest (periwinkle). Cluster F has the largest portion (45%) of members with DDs less than 0. (white circle = median; white or black line = interquartile region)

H. Predicted substrates of (i) MAPK1/3/7 family kinases in Cluster F (a subset), and (ii) Casein Kinase 2 (CK2) in Cluster H (all). Protein trends are the dashed lines and phospho-form trends are the solid lines. The numbers in the parentheses are the homologous human residues and the motifs are the +/- six amino acids around the phosphorylated *Xenopus laevis* residue.

Changes in protein phosphorylation are related to proliferation and gene expression.

To begin to understand the patterns of protein phosphorylation in development, we analyzed together the phospho-form trend and the relative trend of the corresponding protein for each of the phospho-forms. If the protein and phospho-form trends are highly correlated, then we would not want to infer that the amount of protein that is phosphorylated is changing; it may simply be that the phosphorylation is constitutive and necessary for protein function (*Figure 5.A.i,ii*). If the protein and phospho-form trends are not correlated, then the relative amount of protein that is phosphorylated must be changing, and it is much more likely that phosphorylation has a regulatory function (*Figure 5.A.iii*). We capture this relationship using the Pearson Correlation Coefficient of the phospho-form and protein trends, which we term “PP Corr” (*Supplemental Figure 8.A*). We also want to consider the magnitude of phosphorylation change. In order to compare the amount of change in phosphorylation levels between phospho-forms, we determine the maximum fold change of the phosphorylation trend across the time series, which we term “Phos FC” (*Supplemental Figure 8.B*). We used these metrics to compare k-means co-clusters of the protein and phosphorylation trends (*Figure 5.B.; Supplemental Figure 9* for the original ten clusters). (These co-clusters are labelled A-G to distinguish them from the protein-only clusters in *Figure 2.*) The median values of PP Corr and Phos FC for each cluster (with bars covering the interquartile – 25th to 75th percentile -- region) are shown in *Figure 5.C.* along with some of the biological process GO sets that are overrepresented (See *Supplemental Table 5* for complete lists).

Together, Clusters A and B comprise 54% of the phospho-form data. Both have moderate Phos FCs and moderate positive PP Corrs. Clusters D and E are also sizeable clusters, but they have high Phos FCs. GO sets related to gene expression are overrepresented in both clusters A and E. In both cases the phosphorylation levels decrease between Stages 41 and 48, which is in agreement with the dynamics of gene expression proteins across development and suggests that the majority of phosphorylation related to gene expression is constitutive. Cluster D has the smallest PP Corr. The gene set “mitotic cell cycle phase transition” is overrepresented in Cluster D, which is consistent with the fact that the phosphorylation dynamics match the varying role of the cell cycle during development. The small PP Corr is consistent with the importance of phosphorylation for specific regulation of cell cycle progression [39], and the role of phosphatases in removing mitotic phosphorylation, which must result in divergence of protein and phosphorylation dynamics [40,41]. Cluster F is the only cluster with more than 100 phospho-forms where the median phosphorylation abundance increases after Stage 30. Clusters C, G, and H together comprise just under 5% of the data but represent some of the most dramatic changes.

Changes in protein phosphorylation are the result of kinase and/or phosphatase activity (with the caveat of protein degradation). In an effort to connect the protein substrate phosphorylations to the kinases responsible, we used two approaches. We first determined the phosphorylation motifs that are overrepresented in the clusters [42–44] (*Figure 5.D.; Methods; Supplemental Table 6* for complete lists). The overrepresentation of proline-directed ([S/T]-P) motifs in Cluster D is consistent with high activity of master-cell cycle kinase CDK1 in the egg. (We measure the inhibitory doubly phosphorylated (T-14 and Y-15 in human) phospho-form of CDK1 in both replicates, and as expected it is at its lowest

level (least inhibited) in the egg; *Supplemental Figure 10.*) Cluster D is also the only cluster where a threonine phospho-acceptor motif is overrepresented (see below). Multiple acidic motifs are overrepresented in Cluster E. Casein Kinase 2 (CK2), which phosphorylates residues proximal to acidic residues and is known to phosphorylate proteins involved in ribosomal RNA processing and ribosome biogenesis, is a plausible candidate kinase for many phosphorylations in this cluster [45,46]. Examples of nucleolar proteins with phosphorylations in acidic motifs are mRNA protein turnover protein 4 homolog (MTRO4) and ribosome biogenesis protein BMS1 homolog (BMS1), which function in pre-60S subunit assembly and pre-40S subunit assembly, respectively (*Figure 5.E*) [47,48].

We also used NetworKIN to predict sequence-to-kinase mappings [49–51] (*Figure 5.F.; Methods*). (We developed a custom pipeline that uses global and local sequence homology to identify human phosphorylations that are homologous to our measured Xenopus ones. See *Supplemental Figure 11* and *Methods*.) The statistically significant overrepresentation of CDK family substrates in Cluster D can be considered a positive control. Predicted substrates of the MAPK1/3/7 and NLK family are overrepresented in Cluster F. *Figure 5.G.* shows three examples of MAPK1/3/7 predicted substrates in Cluster F. Epidermal growth factor receptor (EGFR) is a receptor tyrosine kinase that activates many downstream signaling pathways including MAPK1/MAPK2. RRAS2 is a GTPase that modulates MAPK pathway signaling, and Intersectin-1 (ITSN1) is an adaptor protein that connects the endocytic pathways and regulation of the actin cytoskeleton [52,53]. Predicted substrates of CSNK2 (CK2) are statistically significantly overrepresented in Cluster H and are shown in *Supplemental Figure 12.* Cluster H has the highest PP Corr and is similar in this regard to Cluster E, which has the second highest PP Corr and in which acidic motifs are overrepresented. When the entire dataset is considered together, phosphorylations in acidic regions have the highest PP Corr compared to other broad motif classifications (see below); this is consistent with known properties of CK2 and evidence for low phosphatase selectivity for phosphorylations proximal to acidic residues [54,55].

So far, we have considered only the relationship between phospho-form and protein dynamics. We can also ask about the relationship between the magnitude of phospho-form change as compared to protein change. If Phos FC is the maximum fold change of the phospho-form across development, then Prot FC is the maximum fold change of the associated protein across development. We term the log2 value of the ratio of Phos FC to Prot FC the Relative FC or RFC (*Supplemental Figure 13*). If the phospho-form changes more than the protein, the RFC will be greater than 0; if the protein changes more, the RFC will be less than 0. Cluster F has the lowest median RFC (0.06; *Figure 5.H*; *Supplemental Figure 14 for replicates*). Why does a cluster that increases after Stage 41 have the lowest median RFC? One explanation could be that phosphorylation is becoming more tissue specific even on broadly expressed proteins, and thus the increase in phospho-form level is less than the increase in protein level. This possibility is supported by a study of juvenile mouse tissues that found that 50% of phosphorylations were measured in a single tissue, but only 85% of the proteins on which these phosphorylations were measured were themselves tissue-specific [56].

Threonine phosphorylations show weaker correlation between protein and phospho-form dynamics.

A surprising finding from our clustering approach is that the only cluster with a PP Corr less than 0.5 is the mitotic phosphorylation cluster. However, the large interquartile regions of many of the clusters shows that the data also include phospho-forms with trends divergent from that of their proteins that do not peak at the egg. We decided to classify the phospho-form data instead by the identities/properties of the phosphorylated amino acid and the surrounding (+/- six) amino acids. We used a published motif classification scheme that classifies motifs with the following hierarchy: 1) proline-directed, from 2) acidic, or finally 3) basic [57]. Phospho-forms with motifs not meeting any of these classifications were grouped as “other” (*Figure 6.A; 6.B.i*). We also grouped our phospho-forms by phospho-acceptor residue (serine vs. threonine) (*Figure 6.B.ii*). The preponderance of serine phospho-acceptors over threonine is consistent with previous studies in mouse tissues and human cell lines which reported 11-15.5% of phosphorylation on threonine [56–58]. The threonine phospho-forms have the lowest median PP Corr of all classes (*Figure 6.C*). This is not just the case for a certain motif class; threonine phospho-acceptor proline, basic, and “other” phospho-forms all have statistically significantly smaller median PP Corrs than the set of all phospho-forms (*Supplemental Figure 15*). There is evidence from multiple studies that this lower PP Corr may be attributable not to the kinase regulation but to phosphatase regulation [55,59,60]. An alternative possibility could be that threonine phospho-acceptor phospho-forms appear less correlated to protein levels because of noise. We compared serine and threonine phospho-forms measured in both replicates with a background of scrambled serine and threonine replicated phospho-forms, respectively, and found no evidence that threonine phospho-form trends are less reproducible than serine phospho-form trends (*Supplemental Figure 16*). Many of the threonine phospho-acceptor phospho-forms are proline directed (61%). This is consistent with the overrepresentation of the TP

motif in the mitotic phosphorylation cluster that has the lowest PP Corr and is also consistent with studies of phosphatase regulation of mitotic exit [40,55,59,60].

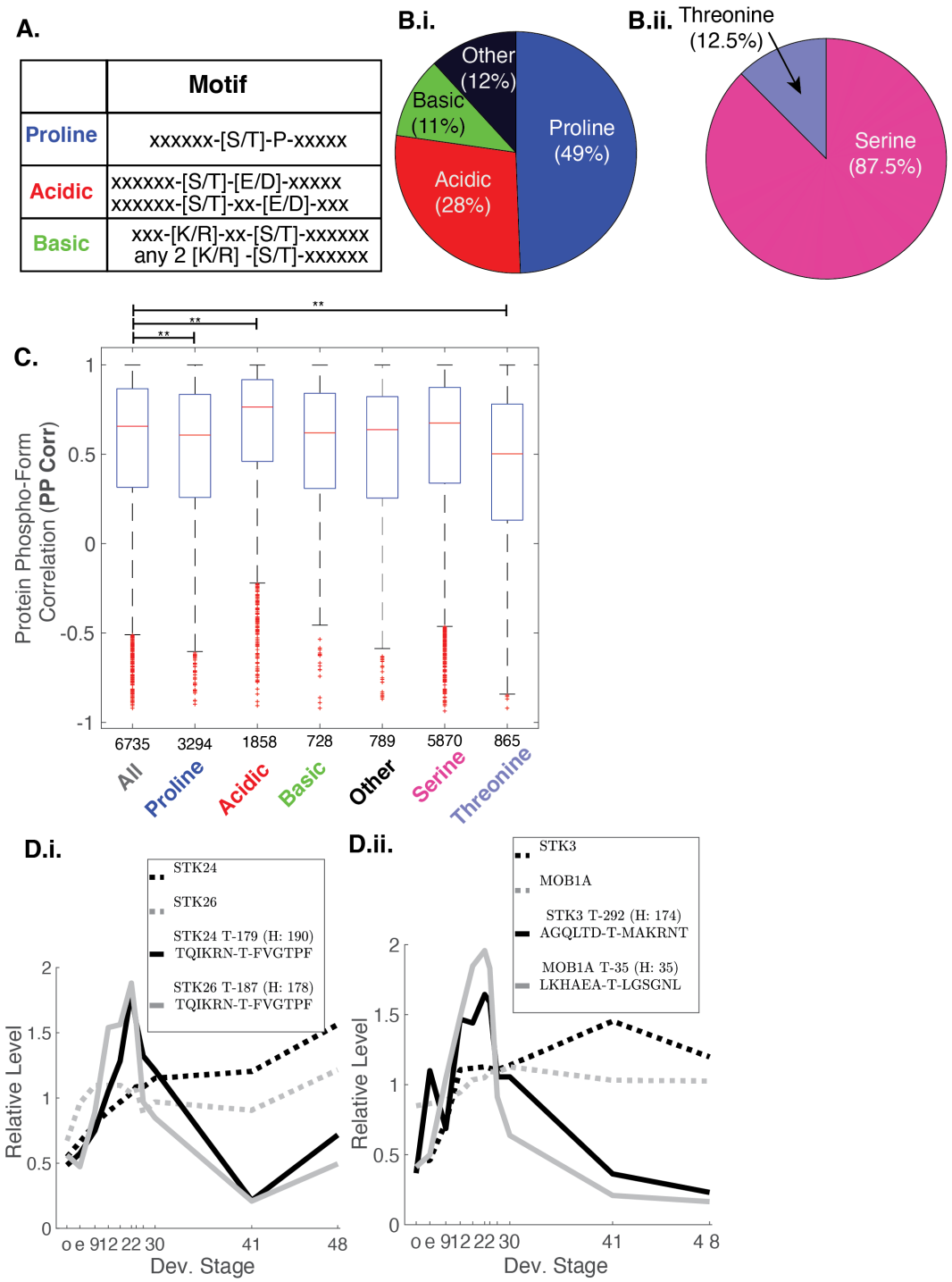


Figure 6. Acidic motif phosphorylations are highly correlated with their associated protein trend; and threonine phosphorylations are less correlated with their associated protein trend.

Figure 6 continued

A. Definitions of the motif classifications.

B. Fraction of phospho-forms by **(i)** motif classification and **(ii)** phospho-acceptor residue. Almost 50% of phospho-forms have proline-directed motifs. The vast majority of phospho-forms have serine phospho-acceptor residues.

C. Acidic motifs are more correlated than the set of all phospho-forms, and threonine phospho-acceptor phospho-forms are the least correlated. (** = $p < 0.001$; Kruskal Wallis test, Tukey-Kramer critical value for multiple comparisons)

D. **(i, ii)** Examples of threonine phosphorylations that do not show strong correlation between the protein and phospho-form trends and increase in phosphorylation level after the MBT.

There are threonine phospho-acceptor phospho-forms with lower PP Corrs that increase in phosphorylation level after the MBT, in contrast to the pattern of high mitotic phosphorylation in the egg. *Figure 6.D.i.* shows examples of threonine phospho-forms on the kinases STE24/MST3 and STE26/MST4. For MST3 the phospho-form is at the active site and the phosphorylation induces activity [61]. STE24 and STK26 regulate the cytoskeleton, cell polarity, and Golgi Apparatus positioning [62–64]. *Figure 6.D.ii.* shows threonine phosphorylations on the Hippo signaling pathway proteins STK3/MST2 and MOB1A. MOB1A is the mediator between the upstream kinases STK3/MST2 and MST1, and the downstream kinases which include LATS1 [65]. The T-35 phosphorylation on MOB1A is a substrate phosphorylation of STK3 [66]. The striking accumulations of these phosphorylations on the Hippo machinery during the neurula and early tailbud period, and their decrease thereafter, suggests a very specific developmental role for YAP signaling at this stage of embryonic development. This may represent a concerted transition in the whole embryo, like the MBT, or it could just be a very strong signal confined to a small set of cells.

Phospho-occupancy: estimation with confidence intervals from endogenous dynamics across development

A fundamental limitation of relative measurement of phospho-proteomics is that it only records the relative change in phosphorylation but does not reveal what fraction of protein is actually phosphorylated at the specific phospho-form residue(s). Without that information, (termed “phospho-occupancy”), five-fold increases from 1% to 5% cannot be distinguished from five-fold increases from 20% to 100%. We previously developed a method that estimates phospho-occupancy along with confidence intervals using endogenous measurements and phosphatase treatment from multiplexed measurement [3]. In this study, we prioritized measuring more time points over phosphatase treatment, and thus are limited to using only endogenous changes. For a limited number of phospho-forms, the endogenous dynamics are sufficient to estimate occupancy with accuracy (*Supplemental Figure 17*; Methods). *Figure 7* shows four examples of phospho-occupancy with confidence intervals for phospho-forms that have homologous human sites. Nuclear-receptor binding protein (NRBP1) is a RAC interacting protein associated with proliferation and membrane trafficking [67,68]. The effect of phosphorylation on serine 415 of NRBP1 has not previously been characterized. The pattern of low phospho-occupancy in the oocyte, high occupancy in the egg, and decreasing occupancy after the egg and cleavage stages suggests that its phosphorylation is cell cycle dependent. Talin-1 (TLN1) is a very large cytoskeletal protein that is present at cell-cell and cell substrate adhesions. Serine 497 (Human Serine 425) was characterized as a target of Cdk5 which, when phosphorylated, results in less turnover of focal adhesions in a cell culture model [69]. The dramatic decrease in the phosphorylation of TLN1 after Stage 9 reminds us that the MBT not only regulates the cell cycle and transcription but massively turns on cell motility [19]. The other two examples also have reasonable interpretations. PAK2 is a highly conserved protein kinase that regulates signaling and apoptosis with an autophosphorylation site at serine 141 [70][71,72]. Phosphorylation after Stage 9 is consistent with the loss of inhibition of apoptosis around the beginning of gastrulation and the dramatic increase in cell movement during gastrulation [73,74]. MOB1A is a component of the Hippo Signaling pathway; the same site was discussed above in *Figure 6.D.ii.* It is important to remember that even with knowledge of occupancy and information from homologous residues in other systems, the functional impacts of phosphorylation on proteins during development cannot be inferred without studies using targeted perturbations in the model of interest. Future phospho-proteomics studies in development should be designed with phosphatase treatment to enable occupancy estimation for a much larger number of phospho-forms.

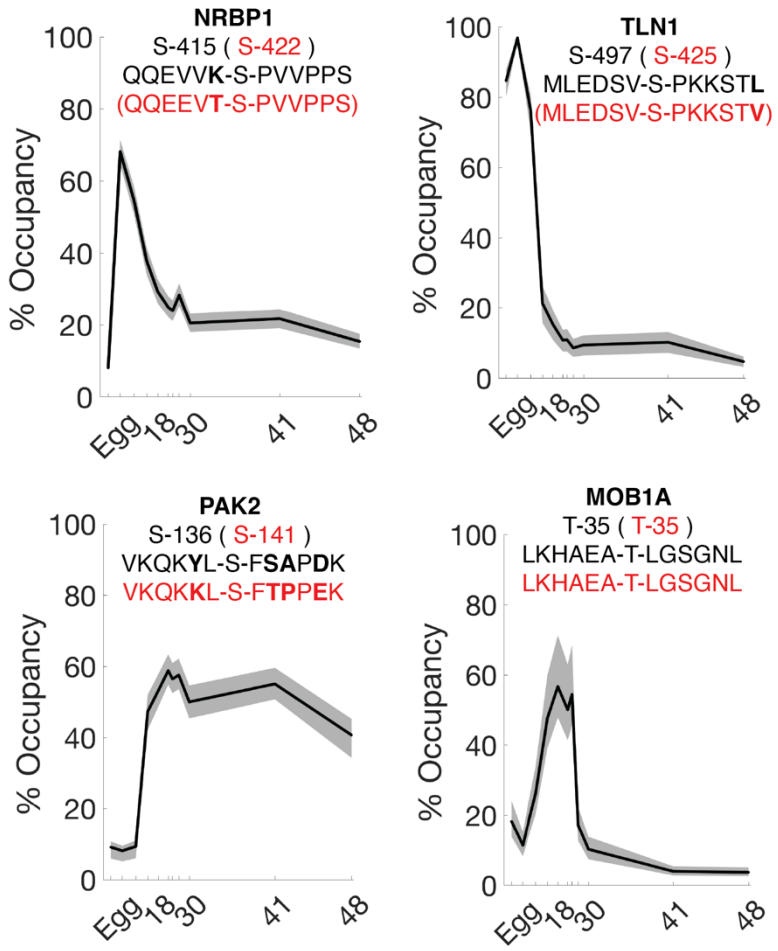


Figure 7. Examples of phospho-occupancy estimation with confidence intervals using endogenous dynamics.

The best estimate of the occupancy of the phosphorylation with 90% confidence interval in gray. The homologous human residue and motif are shown in red. All of the examples are Replicate A data. See *Supplemental Figure 17* for same phospho-forms from Replicate B.

DISCUSSION

We have produced extensive datasets that provide researchers information about protein and phosphorylation changes of interest during embryonic development in Xenopus and have interpreted this data at the proteome and phospho-proteome wide level. Xenopus may be the optimal common developmental model for proteomic and phospho-proteomic analysis, principally because the embryos are large enough and can be easily synchronized. We have been able to report, for the first time, quantitative protein phosphorylation levels in any embryo after the cleavage stages. Due to the present limitations of mass spectrometry methods, we have still only recorded a limited depth of the expected total phospho-proteome. Likely tens of thousands of phosphorylated residues remain to be measured [11]. The improved depth and temporal resolution during early development did not challenge our previous finding that abundance changes of the majority of proteins are very small across early and mid-stages of embryogenesis. However, by extending the time period of our measurement to the juvenile, we observe multiple biochemical transitions. We report a two-fold increase in total (non-yolk) protein between Stages 30 and 48. The composition of the proteome shifts after Stage 41 because membrane and secreted proteins increase much more than nuclear proteins. Thus, it is not until organ systems become functional that dramatic changes in the proteome occur.

Despite the inclusion of many timepoints that have a low mitotic index, the most prominent trend of large changes in phosphorylation level are the phospho-forms that are highest in the metaphase-arrested egg. That so many of these phosphorylations are detected, even though the majority decrease dramatically in level by the next time point, is likely explained by their high occupancy in the mitotically arrested egg. This strongly expected high occupancy at the metaphase arrested egg is due to both the fact that metaphase phosphorylations are present at high occupancy [58,75], and the fact that the metaphase arrest egg constitutes a perfectly synchronous and homogeneous state. Though the goal of this study was to measure phosphorylations after the midblastula transition, the prominence of cell cycle related phosphorylation in this data demonstrates again the not yet fully tapped power of the Xenopus system for studying cell cycle phosphorylation at the phospho-proteome wide level.

The most prominent protein and phosphorylation changes after fertilization but before the embryo becomes a tadpole are connected with the regulation of gene expression. Usually, gene expression during development is considered in terms of transcription factors with temporally and spatially discrete expression patterns. Phosphorylation related to gene expression is usually considered only for specific well characterized phosphorylations that regulate transcription factor localization and stability. However, the protein and phosphorylation changes we detect are on proteins expected to be used ubiquitously, suggesting that there is also an embryo-wide level at which the prioritization of regulated gene expression is occurring. Many of these phosphorylations related to gene expression occur in acidic motifs. Phospho-proteome wide studies of phospho-occupancy in yeast and a cell culture model found that the set of phosphorylations in acidic motifs have higher occupancy than the set of all phosphorylations [76,77]. It would be interesting to determine if these phosphorylations are necessary for protein function or a byproduct of the combination of non-specific phosphorylation by kinases and low phosphatase activity.

The changes in the absolute levels of proteins and the increase of the total non-yolk protein amount after Stage 30 together provide strong evidence that after Stage 30 there are embryo-wide biochemical events necessary for the development of functional organs. We measure large increases in relative protein levels, the total amount of protein in the embryo increases, and the composition of the proteome shifts to contain a larger fraction of membrane and secreted proteins. Tissue-specific proteins increase dramatically, particularly after Stage 41. In contrast to the change of the proteome, we measure few large changes in the phospho-proteome. The phosphorylations that do increase during this period are all on proteins that are also increasing. Thus, dramatic change in phosphorylation levels of proteins whose relative levels do not change is not a signature of the transition from specified to functional. It will be particularly interesting to have more occupancy information about phosphorylation during this period. It is possible that, because of the much lower mitotic index in these juvenile tadpole tissues and the decreased level of proteins involved in gene expression, the absolute amount of phosphorylation in juvenile and adult tissues may actually be lower than in earlier embryos.

The coincidental change of the proteome with organ function emergence demands that we think more about the transition between embryonic and adult, rather than study them separately. Studies using embryonic stem cells to investigate how the proteome changes with differentiation have observed increase in translation levels in differentiated embryoid bodies [78], as well as increased translational efficiency of membrane proteins [79]. However, embryoid bodies leave much to be desired from the perspective of cell-type specific emergence of function. This work has highlighted that, despite the spatial heterogeneity of developing vertebrates, which is particularly pronounced in the tadpole, temporal synchronicity is still observable, showing that *Xenopus laevis* provide a useful model to study these events. It would be particularly interesting to have higher temporal resolution data at all levels of the gene expression trajectory, in order to understand the contribution of RNA levels, translation levels, and protein stability to the transition from early specification to functional tissues. Single-cell transcriptomics data during this period could also provide information about whether the mRNA for non-tissue specific proteins increases in different tissues at the same time, or whether there is temporal heterogeneity in expression that correlates with the functionality of different organs. The data presented in this paper show that embryogenesis involves two different periods, the first being a period of generation of organization and difference, with minimal requirement for new protein. Only once this organized heterogeneity has been established do we see the second period, a period of biochemical differentiation where substantial change to the proteome does occur. The first period has been studied extensively by developmental biologists. The second period – the hints of which are present in work done half a century ago – is ripe for renewed attention.

ACKNOWLEDGEMENTS

The authors enthusiastically thank R. Michael Gage for preparing the sample for mass spectrometry measurement, Rachael Jonas-Closs for excellent frog husbandry, Alexander Lukyanov for writing the initial motif scoring code, Marc S. Presler for writing the initial phospho-occupancy code, and Kathleen M. Keating and Alex Truong for implementing the parallelization of the phospho-occupancy code. This work was supported by NIH Grants HD091846 and HD073104 (to MWK and LP).

CONTRIBUTIONS

Conceptualization (EV, MHW, LP, MWK), Data curation (EV, MK), Formal Analysis (EV), Funding acquisition (LP, MWK), Investigation (EV, MHW, MK, LP), Software (EV), Supervision (MK, MHW, LP, MWK), Visualization (EV), Writing – original draft (EV), Writing – reviewing & editing (EV, MK, LP, MWK)

METHODS

Illustrations

Figure 1.A. and *Figure 4.B.* illustrations © 2021 Natalya Zahn, CC BY-NC 4.0 ."
The source of the illustrations is Xenbase (RRID:SCR_003280).

Sample Preparation, Mass Spectrometry Measurement, and Preparing Data for Analysis

Samples: All handling of *Xenopus laevis* adults was carried out in accordance with HMS IACUC protocol IS00001365. *Xenopus laevis* J-line embryos were collected at Stages oocyte VI, egg, 9, 12, 18, 22, 24, 26, 30, 41, and 48. Embryos were de-jellied in 2% cysteine, pH 7.8, and flash frozen for later preparation. Stage VI oocytes were obtained by surgery from the same females from which eggs were collected for fertilization. Oocytes were harvested after the females had replenished their oocytes following the stimulation that resulted in collection of eggs.

Lysis: Frozen embryos/eggs at each stage were thawed and 5-6 uL of lysis buffer (1% NP-40, 250 mM sucrose (EMD cat# 8550), 10 mM EDTA (VWR cat# VW1474-01), 1 tablet Roche Complete mini protease inhibitor (cat# 11836153001) and 1 tablet Roche Phos STOP tablet (cat# 04906837001) per 10 mL, 25 mM HEPES (Sigma cat# H3375-500G) pH 7.2, 10 uM Combrestatin 4A (Santa Cruz cat# sc-204697), 10 uM Cytochalasin D (Santa Cruz cat# sc201442), 1 mM TCEP (ThermoFisher cat# 20490)) was added per embryo/egg on ice. The embryos/eggs were lysed by pipetting followed by thorough vortexing well. Preparation of the lysate from the Stage 48 tadpole required more extensive maceration of the sample with a pipette tip than did less mature samples.

Yolk removal: The yolk was removed from the samples by spinning at 4,000 xg for 4 minutes in a microcentrifuge. The lipids in the sample were then resuspended by lightly flicking the tube, being careful not to resuspend any of the yolk. The supernatant was transferred to another tube, and 2% SDS (Amresco cat# M112-500ML) was added. HEPES (Sigma cat# H3375-500G) (pH 7.2) was added to a final concentration of 100 mM. Additionally, during the yolk removal step material pelleted in addition to the expected yolk and pigment.

Alkylation and cysteine protection: DTT (5 mM, pH ~8.0, Sigma cat# 43819-1G) was added to each sample and the sample was incubated at 60 C for 20 minutes, then cooled to RT. Next, NEM (15 mM, Sigma cat# E3876-25G) was added to each sample and the samples were incubated for 20 minutes at room temperature. Finally, DTT (5 mM, pH~8.0, Sigma 43819-1G) was added again and the samples were incubated at room temperature for 10 minutes.

Enzymatic Digestions: The proteins in the sample were precipitated using methanol/chloroform and were resuspended in 6 M Guanidine HCl (Sigma cat# G3272-1KG), buffered using 50 mM EPPS (Alfa Aesar cat# A13714), pH 9.0) to an estimated protein concentration of 5 mg/mL. The samples were heated to 60 C for 5 minutes and then allowed to cool to room temperature. The protein concentrations of each sample were determined by BCA assay (ThermoFisher cat# 23225). Next, each sample was diluted to 2 M Guanidine HCl with

5 mM EPPS (pH 9.0, Alfa Aesar cat# A13714), ensuring that the pH was at least 8.5. Lys-C (Wako Chemicals cat# 129-02541) was added at the higher of 1:100 w/w or 20 ng/uL and incubated for 12 hours at room temperature. The samples were diluted to 0.5 M Guanidine HCl with 5 mM EPPS (pH 9.0, Alfa Aesar cat# A13714), ensuring that the pH was at least 8.5. Lys-C (Wako Chemicals cat# 129-02541) was added again, as above, and allowed to digest for at least 15 minutes at room temperature. Additionally, Trypsin (Promega cat# V5113) was added at the higher of 1:50 w/w or 10 ng/uL and allowed to digest at 37 C for 8 hours. Samples were speed-vacuumed to dryness.

TMT labeling: Each sample (stage) was labeled using a distinct channel of TMT label (ThermoFisher cat# 90111 & A34807). Samples were resuspended in 500 mM EPPS (pH 8.0, Alfa Aesar cat# A13714), checking that the pH was close to 8.0, and incubated at 65 C for 5 minutes. After the samples returned to room temperature, fifteen uL of TMT reagent (20 mg/mL, ThermoFisher cat# 90111 & A34807) were added per each 100 ug of protein, and the reactions were incubated for 2 hours at room temperature. A small subset (~1 ug/condition) was quenched and tested for TMT labeling efficiency, missed cleavage rate, and normalized total peptide count. Reactions were quenched by first heating the samples to 60 C for 5 minutes and then cooling to room temperature, followed by incubation with 0.5% hydroxylamine (Sigma cat# H9876) for 15 minutes at room temperature. The samples were then combined in another tube containing phosphoric acid (JT Baker cat# B34P0200) at 5% of the total combined volume. The combined sample was then speed-vacuumed to dryness.

Peptide purification, phosphopeptide enrichment and peptide fractionation: After TMT labeling, peptides were re-suspended in 1% formic acid and 0.1% trifluoroacetic acid (TFA) and purified by reversed phase chromatography using C18 Sep-Pak cartridges (Waters #WAT054945). After preconditioning the matrix with methanol and two full column washes with 1% formic acid with 0.1% TFA, peptides were bound by gravity flow. Columns were washed with 3 column volumes of 1% formic acid, peptides eluted with 95% acetonitrile in 1% formic acid and speed-vacuumed to dryness. Phosphopeptides were enriched by immobilized metal affinity chromatography (IMAC) using an Fe-NTA matrix (High-Select™ Fe-NTA phosphopeptide enrichment kit, Thermo Scientific #A32992) according to manufacturer's instructions. In brief, peptides were resuspended in loading buffer (0.1% TFA, 80% acetonitrile), allowed to bind to Fe-NTA matrix for 30 minutes, washed 2 times with loading buffer and once with water. Flow throughs of Fe-NTA columns were collected, pooled and peptides speed-vacuumed to dryness. Fe-NTA bound, enriched phosphopeptides were eluted with alkaline ammonia solution in water twice and dried down in parallel. Enrichment of phospho-peptides after TMT labelling avoids variability in the enrichment between time points and allows the protein level normalization scheme to be applied to the phospho-form data. The ability to use the same normalization scaling for each sample was especially important, since we expected large differences in bulk phosphorylation during our early time points due to the very high mitotic phosphorylation state of the egg [41,58,75].

Phospho-peptides were fractionated using alkaline reversed phase chromatography (Pierce™ High pH Reversed Phase Peptide Fractionation Kit, Thermo Scientific #84868) into 12 fractions using increasing concentrations of acetonitrile from 10%-80% in basic binding buffer. Fractions 1 and 7, 2 and 8, 3 and 9, 4 and 10, 5 and 11, and 6 and 12 were combined. For fractionation of phospho de-enriched TMT labeled peptides, HPLC fractionation by alkaline reversed phase chromatography was carried out using an Agilent 1200 Series HPLC system with a flow rate of 600 ul/min over a 65 minute gradient with increasing concentrations of acetonitrile. 96 fractions of a 65-minute gradient of 13-44% buffer B (90% acetonitrile, 10mM ammonium bicarbonate, pH 8) in buffer A (5% acetonitrile, 10mM ammonium bicarbonate, pH 8) were collected and combined into 24 pooled fractions for MS analysis. Prior to injection, peptide samples were purified via Stage tips using punched out cookies of C18 3M™ Empore™ extraction disc material (Fisher Scientific #14-386-2) as matrix. Phosphopeptides were re-suspended in 1% formic acid prior to MS injection, non-phospho peptides in 1% formic acid with 5% acetonitrile.

Mass spectrometry measurement: Data were collected using a Multi Notch MS^3 TMT method [17] using Orbitrap Fusion Lumos mass spectrometers (Thermo Fisher Scientific) coupled to a Proxeon EASY-nLC 1000 liquid chromatography (LC) system (Thermo Fisher Scientific). The 100 uM inner diameter nanospray capillary columns used were packed with C18 resin (2.6 µm, 150 Å, Thermo Fisher Scientific). Peptides of each fraction were separated over 4 and 5-hour acidic acetonitrile gradients for non-phospho peptides and 2.5 hours for phosphopeptides by LC prior to mass spectrometry (MS) injection. The scan sequence started with recording of an MS1 spectrum (Orbitrap analysis; resolution 120,000; mass range 400−1400 Th). MS2 analysis followed

collision-induced dissociation (CID, CE=35) with a maximum ion injection time of 100-200 ms. For phospho-peptide measurement, a multi-stage activation method was used with a neutral loss of 97.9763/z. For TMT quantification, MS3 precursors were fragmented by high-energy collision induced dissociation (HCD) and analyzed in the Orbitrap at a resolution of 50,000 at 200 Th. Further details on the LC and MS parameters and settings used here were described previously [80].

Mapping and normalization: For mapping of protein and phospho data (peptide-spectra matches for MS), we used, as a main database of reference, sequences from the *X. laevis* genome assembly (DoE JGI REF; v9r1 of assembly v1.8.3.2: a total of 45,099 sequences) downloaded from Xenbase [81] (ftp://ftp.xenbase.org/pub/Genomics/JGI/Xenla9.1/1.8.3.2/). Peptides were searched with a SEQUEST-based in-house software with a target decoy database strategy and a false discovery rate (FDR) of 2% set for peptide-spectrum matches by linear discriminant analysis (LDA). The peptide and phospho-peptide identifications from both replicates were then combined for a final collapsed protein-level FDR of 2%. Phospho-peptide localization confidence was quantified using the AScore method. All analyzed phospho-forms have AScore values >= 13. Quantitative information on peptides and phospho-peptides was derived from MS3 scans. For phospho-forms and parent-forms, only peptides with a sum of TMT s/n >= 200 were used. Details of the TMT intensity quantification method and further search parameters applied were described recently [80]. The channels were further normalized assuming that the same amount of protein per time point was combined after multiplexing, using total time point signal as a proxy for protein amount. Peptides were summarized per protein using the BACIQ-P pipeline [82]. All analysis for relative proteins and phospho-form trends was done using the mean trend, and after analysis of reproducibility the relative trends from replicates were averaged. Human gene symbols were assigned to the protein references using a custom HMMR-based algorithm (*See below*). Due to a partial genome duplication in the pseudo-tetraploid *Xenopus laevis*, many proteins have homoeologous genes (sometimes termed alloalleles), which have separate gene references [83]. These alloalleles are analyzed separately.

Phospho-forms are distinct peptides with distinct phosphorylation events that could have a single phosphorylation or multiple phosphorylations. Almost 90% of measured phospho-forms in this dataset have only one phosphorylation (*Supplemental Figure 1.A.*). We identified all measured peptides that had the same residues phosphorylated. In the cases where multiple versions exist (because of internal lysines and arginines), we used only the most fully cleaved version of the peptide. Additionally, we retain only phospho-peptides that uniquely align to the protein reference to which they have been assigned.

Assigning Human Gene Symbols to Xenopus laevis references: The *X. laevis* 9.1 reference genome and respective gene models lack gene symbols or have generic alias for over 11k genes even though in many cases an unambiguous match can be identified by sequence homology to human genes. Since most of the functional analysis in biology depends on gene sets and thus proper nomenclature, we fill the gaps in the gene symbol assignments. Specifically, we assigned gene symbols to each *X. laevis* 9.1 transcript using our own pipeline, based on a reciprocal best match by HMMER [84] to a curated target reference set of 20,137 human protein sequences. The original set was obtained from UniProt. We used a BLOSUM80 weight matrix appropriate for an estimated 350 million year time of divergence between human and amphibians. The ties among matches with identical E-value were resolved based on bit-score. Sequences which did not get bidirectional best match were assigned a symbol based on the unidirectional best HMMER match. Only matches with an E-value better than 1e-5 were considered. In total, our pipeline matched a total of 34,051 *Xenopus laevis* genes to one of 17,710 unique gene symbols, which provides 11,341 more *Xenopus* genes with a gene symbol than available in the original annotation supplied with the genome assembly release, significantly enhancing data usability.

Data Accessibility: Mass spectrometry spectra files are uploaded to proteome exchange. The mapped and normalized relative protein, relative phospho, absolute protein, and phospho-occupancy datasets are included as spread sheets including the data that integrates both replicates used to generate Figure 2 through 6 and the complete replicates. MATLAB code for generating the main text figures is on GitHub at.

Over Representation Analysis (ORA): We employed a two-step approach. First we used the WebGESTALT web server (www.webgestalt.org) to perform ORA with the “Biological Process nonRedundant" database [85]. For each cluster, we used the assigned human gene symbols for the proteins in that cluster as the “gene list” and used all of the human gene symbols assigned to proteins quantified as the “background”. The set of human gene symbols for the measured proteins is not unique, both because of allo-alleales and cases where multiple Xenopus proteins map to the same human protein,

and not all replicate gene symbols are in the same “gene lists" that are tested for overrepresentation. The complete set of overrepresented GO gene sets for all clusters with FDR <= 5% is reported in *Supplemental Table 1*.

For the subset of gene sets shown in Figure 2B, we identified all of the clusters for which the gene set was over-represented. Then we determined the p-value of the hypergeometric test for each cluster in which the gene set is over-represented. If the p-value is less than or equal to 0.05, it is visualized with a circle of area proportional to the -log10 of the p-value.

Gene Set Enrichment Analysis (GSEA): We used the WebGESTALT web server (www.webgestalt.org) to perform GSEA for the fraction of maximum fold change before the tadpole metric. The metric was calculated for each protein quantified in the dataset, and assigned human gene symbols were used. The full list of positively and negatively enriched genes is shown in *Supplemental Table 2*.

Measuring average total protein amount per embryo: Sets of ten embryos from the same clutch were flash frozen at the same developmental stages as the proteomics dataset, except that Stage VI oocytes were not collected. The egg through Stage 41 samples were prepared exactly as described above under **Sample Preparation and Mass Spectrometry Measurement**. An additional egg sample and the Stage 48 sample were lysed in the same lysis buffer but with SDS already added to 2%, and there was no yolk spin out. After cysteine protection, and before the methanol chloroform precipitation, the concentration of a 1:10 dilution of the lysate (1:50 for the egg with yolk) was measured with a reducing agent compatible BCA assay (ThermoFisher 23252).

Estimating Absolute Protein Amounts for Individual Proteins: BLASTP was performed reciprocally for references from v8.3.2.1 and the PHROG database with e-value cutoff of 1e-20. Pairs of references where both members are unique matches and both members are measured in the respective datasets were used to estimate the linear relationship between the log of the fraction of the MS1 signal attributed to the egg and the log of estimated protein concentration in the egg from the PHROG database reference [2]. The MATLAB function fit_2D_data.m was used to perform the orthogonal regression because there is error in both dimensions [86]. The resulting fit and the assumption that the volume of non-yolk cytoplasm in the egg is 0.33 uL were used to estimate the number of moles of protein in the egg for all measured proteins. These values were then scaled by the relative protein change data and the relative change in absolute amount of total protein in the embryo, to determine the number of moles at the other stages. It was assumed that the total amount of protein in the oocyte and egg is the same.

Protein Localization Classification: Localization annotations for nucleus, membrane, secreted, and cytoplasm were downloaded from the manually annotated and reviewed SwissProt database via UniProt. We considered proteins as nuclear if they were annotated as only nuclear or as nuclear and cytoplasmic; cytoplasmic if they were only annotated as cytoplasmic; and membrane if they were annotated as only membrane or as membrane and cytoplasmic. We considered proteins as secreted if they were annotated as secreted or both membrane and secreted.

Tissue Specificity: Proteins were classified as tissue specific if they were both (1) not measured in the egg proteome database, and (2) classified as “Tissue enriched” or “Group Enriched” in the Tissue Atlas Database. A protein was considered as not having been measured in the egg only if all matching protein references had not been measured. For the small fraction of references that do not have a match in the egg database, we classified them as not being in the egg if their estimated abundance in the egg was less than two standard deviations from the mean value for proteins measured in the egg database.

Identifying homologous phospho-acceptors between Xenopus and human using global and local sequence homology: First we identified the human sequence in the reference set of human protein sequences that is the best match for each of the Xenopus protein references on which we have measured a phospho-site. We used the human-phosphosite-fastas.fasta from Phosphosite.org after filtering to remove all isoform references. We chose the BLASTP result with the smallest e-value, and only used alignments if the e-value is less than 1e-20. For all the phosphorylated residues on each Xenopus reference, we used the alignment to determine if there is a corresponding potentially phosphorylated residue that aligns on the matched human reference. At this step there are five possible outcomes from evaluating the alignments at each queried residue position: (1) aligned to an identical human residue *(match)*; (2) S->T or T->S mismatch *(match)*; (3) aligned to a different human residue *(mismatch)*; (4) aligned to a gap in the human reference

(mismatch); and (5) found on a part of the *Xenopus laevis* reference that does not align to the human reference *(mismatch)*. We considered serine and threonine residues that align to the opposite residue as matches, for two reasons: (1) mutations between S and T are relatively common over evolutionary time; and (2) there are many kinases that phosphorylate both serine and threonines. In the cases where the Xenopus residue aligns to a human residue, we determine the six amino acids that flank the aligned residue N- and C-terminally (the "motif"). Then we score the local sequence homology for all these aligned motifs. The motif score is based on the Blosum 90 substitution penalty matrix. The alignment score is calculated from the flanking amino-acids only, not the aligned amino-acids. Since the alignment score depends not only on the number of matches, but also on the identities of the constituent amino acids, we normalize for sequence effects of the score by finding the "best score" (largest of human or xenopus alignment score to itself) and "worst score" (smallest of human or xenopus alignment score to its flipped self). The "motif score" is the Xenopus against human alignment score minus the "worst score" divided by the "best score" minus the worst score. (*Motif Score = (Xenopus: Human alignment - "worst score") / ("best score" - "worst score")*). Then, using the alignments to non-phosphorylatable residues as our "false discoveries," we determined the false discovery rate as a function of motif score. We retain matches that have motif scores greater than or equal to 0.75, which means that we have an FDR of 7.9%. For python code and detailed README see: https://github.com/e-vanitallie/Homologous_Phos

Motif Enrichment: Motif enrichment was performed using the R motif-x implementation [42]. For each of the eight k-means clusters, the +/- 6 amino acid motifs for all phospho-forms in the clusters were the foreground set and the +/- 6 amino acid motifs for all of the phospho-forms in the dataset were used as the background. The enrichment was performed with the central residue as Serine and Threonine for all clusters; thus, sixteen total tests were performed. The min.seqs parameter was set to 20 and the pval.cutoff parameter was set to 2.6e-4 to account for multiple hypothesis testing, as recommended [43]. Where, as a consequence, multiply phosphorylated peptides had multiple motifs per phospho-form, we considered each motif separately.

Predicting Kinase Family to Substrate Matches with NetworKIN: For all phospho-forms where there is a homologous human match (see above for homology determination method), the homologous human information was used as input into NetworKIN. Kinase to motif predictions with NetworkIN scores greater than 0.2 were retained (http://networkin.info/index_ht.shtml).

Occupancy Estimation: Since we used the relationship between the relative change of phospho-forms and measured not-phosphorylated peptides from the corresponding tryptic region ("parent peptides") to estimate phospho-occupancy (20), the first step in the process was to identify these parent peptides. We used the sequence of the phospho-form peptides that are assigned to a given protein reference to identify the tryptic region that would result in a parent peptide. Then we found all peptides that overlapped that region. Multiple peptides can come from the region because of incomplete cleavage events. If a peptide overlaps the entire region and is the only measured peptide that overlaps the region, we used it as a parent form. Parent peptides must map to only one location on the protein reference that they are assigned to. We estimated phospho-occupancy for replicate A and B phospho-forms only using parent forms identified in the replicate A and replicate B unmodified data, respectively.

Phospho-occupancies were estimated with confidence intervals using essentially the same method as that described in Presler et al. [3], but with two changes. We normalized the phospho-form and parent-peptide data by the protein trend before estimating occupancy, and instead of using orthogonal regression to solve the system of linear equations, we used singular value decomposition.

References

1. Ting L, Rad R, Gygi SP, Haas W. MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. Nat Methods. 2011 Nov;8(11):937–40.

2. Wühr M, Freeman RM, Presler M, Horb ME, Peshkin L, Gygi SP, Kirschner MW. Deep Proteomics of the Xenopus laevis Egg using an mRNA-Derived Reference Database. Curr Biol. 2014 Jul;24(13):1467–75.

3. Presler M, Van Itallie E, Klein AM, Kunz R, Coughlin ML, Peshkin L, Gygi SP, Wühr M, Kirschner MW. Proteomics of phosphorylation and protein dynamics during fertilization and meiotic exit in the *Xenopus* egg. Proc Natl Acad Sci. 2017 Dec 12;114(50):E10838–47.

4. Wühr M, Güttler T, Peshkin L, McAlister GC, Sonnett M, Ishihara K, Groen AC, Presler M, Erickson BK, Mitchison TJ, Kirschner MW, Gygi SP. The Nuclear Proteome of a Vertebrate. Curr Biol. 2015 Oct;25(20):2663–71.

5. Gupta M, Sonnett M, Ryazanova L, Presler M, Wühr M. Quantitative Proteomics of Xenopus Embryos I, Sample Preparation. In: Vleminckx K, editor. Xenopus [Internet]. New York, NY: Springer New York; 2018 [cited 2020 Dec 15]. p. 175–94. (Methods in Molecular Biology; vol. 1865). Available from: http://link.springer.com/10.1007/978-1-4939-8784-9_13

6. Sonnett M, Gupta M, Nguyen T, Wühr M. Quantitative Proteomics for Xenopus Embryos II, Data Analysis. In: Vleminckx K, editor. Xenopus [Internet]. New York, NY: Springer New York; 2018 [cited 2020 Dec 15]. p. 195–215. (Methods in Molecular Biology; vol. 1865). Available from: http://link.springer.com/10.1007/978-1-4939-8784-9_14

7. Peshkin L, Wühr M, Pearl E, Haas W, Freeman RM, Gerhart JC, Klein AM, Horb M, Gygi SP, Kirschner MW. On the Relationship of Protein and mRNA Dynamics in Vertebrate Embryonic Development. Dev Cell. 2015 Nov;35(3):383–94.

8. Woodland HR. Changes in the polysome content of developing Xenopus laevis embryos. Dev Biol. 1974 Sep;40(1):90–101.

9. Baum EZ, Wormington WM. Coordinate expression of ribosomal protein genes during Xenopus development. Dev Biol. 1985 Oct;111(2):488–98.

10. Jorgensen P, Steen JAJ, Steen H, Kirschner MW. The mechanism and pattern of yolk consumption provide insight into embryonic nutrition in Xenopus. Development. 2009 May 1;136(9):1539–48.

11. Vlastaridis P, Kyriakidou P, Chaliotis A, Van de Peer Y, Oliver SG, Amoutzias GD. Estimating the total number of phosphoproteins and phosphorylation sites in eukaryotic proteomes. GigaScience [Internet]. 2017 Feb 1 [cited 2020 Jul 6];6(2). Available from: https://academic.oup.com/gigascience/article/doi/10.1093/gigascience/giw015/2865214

12. Peuchen EH, Cox OF, Sun L, Hebert AS, Coon JJ, Champion MM, Dovichi NJ, Huber PW. Phosphorylation Dynamics Dominate the Regulated Proteome during Early Xenopus Development. Sci Rep. 2017 Dec;7(1):15647.

13. McGivern JV, Swaney DL, Coon JJ, Sheets MD. Toward defining the phosphoproteome of *Xenopus laevis* embryos. Dev Dyn. 2009 Jun;238(6):1433–43.

14. Fíla J, Honys D. Enrichment techniques employed in phosphoproteomics. Amino Acids. 2012 Sep;43(3):1025–47.

15. Fermin D, Walmsley SJ, Gingras A-C, Choi H, Nesvizhskii AI. LuciPHOr: Algorithm for Phosphorylation Site Localization with False Localization Rate Estimation Using Modified Target-Decoy Approach. Mol Cell Proteomics. 2013 Nov;12(11):3409–19.

16. Hornbeck PV, Zhang B, Murray B, Kornhauser JM, Latham V, Skrzypek E. PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. Nucleic Acids Res. 2015 Jan 28;43(D1):D512–20.

17. McAlister GC, Nusinow DP, Jedrychowski MP, Wühr M, Huttlin EL, Erickson BK, Rad R, Haas W, Gygi SP. MultiNotch MS3 Enables Accurate, Sensitive, and Multiplexed Detection of Differential Expression across Cancer Cell Line Proteomes. Anal Chem. 2014 Jul 15;86(14):7150–8.

18. Nielsen AL, Oulad-Abdelghani M, Ortiz JA, Remboutsika E, Chambon P. Heterochromatin Formation in Mammalian Cells: Interaction between Histones and HP1 Proteins. Mol Cell. :11.

19. Newport J, Kirschner M. A major developmental transition in early xenopus embryos: I. characterization and timing of cellular changes at the midblastula stage. Cell. 1982 Oct;30(3):675–86.

20. Saka Y, Smith JC. Spatial and Temporal Patterns of Cell Division during Early Xenopus Embryogenesis. Dev Biol. 2001 Jan;229(2):307–18.

21. Identification of a Vertebrate Sister-Chromatid Separation Inhibitor Involved in Transformation and Tumorigenesis. 2021;6.

22. Nieuwkoop PD, Faber J. Normal Table of Xenopus Laevis (Daudin). 2nd ed. New York, NY: Routledge; 1994.

23. Sasai Y. Xenopus chordin: A novel dorsalizing factor activated by organizer-specific homeobox genes. Cell. 1994 Dec;79(5):779–90.

24. Zhao C. Dual Requirement of CHD8 for Chromatin Landscape Establishment and Histone Methyltransferase Recruitment to Promote CNS Myelination and Repair. :25.

25. Kuboyama K, Fujikawa A, Suzuki R, Tanga N, Noda M. Role of Chondroitin Sulfate (CS) Modification in the Regulation of Protein-tyrosine Phosphatase Receptor Type Z (PTPRZ) Activity. J Biol Chem. 2016 Aug;291(35):18117–28.

26. Nanba R, Fujita N, Nagata S. Structure and expression of myelin basic protein gene products in Xenopus laevis. Gene. 2010 Jul;459(1–2):32–8.

27. Gao N, White P, Kaestner KH. Establishment of Intestinal Identity and Epithelial-Mesenchymal Signaling by Cdx2. Dev Cell. 2009 Apr;16(4):588–99.

28. Chalmers AD, Slack JMW, Beck CW. Regional gene expression in the epithelia of the Xenopus tadpole gut. Mech Dev. 2000 Aug;96(1):125–8.

29. Chen Z-Y, Hasson T, Zhang D-S, Schwender BJ, Derfler BH, Mooseker MS, Corey DP. Myosin-VIIb, a Novel Unconventional Myosin, Is a Constituent of Microvilli in Transporting Epithelia. Genomics. 2001 Mar;72(3):285–96.

30. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci. 2005 Oct 25;102(43):15545–50.

31. Bentovim L, Harden TT, DePace AH. Transcriptional precision and accuracy in development: from measurements to models and mechanisms. Development. 2017 Nov 1;144(21):3855–66.

32. Monsoro-Burq AH. PAX transcription factors in neural crest development. Semin Cell Dev Biol. 2015 Aug;44:87–96.

33. Koide T, Hayata T, Cho KWY. Xenopus as a model system to study transcriptional regulatory networks. Proc Natl Acad Sci. 2005 Apr 5;102(14):4943–8.

34. King RW, Deshaies RJ, Peters J-M, Kirschner MW. How Proteolysis Drives the Cell Cycle. Science. 1996 Dec 6;274(5293):1652–9.

35. Mark KG, Rape M. Ubiquitin-dependent regulation of transcription in development and disease. EMBO Rep [Internet]. 2021 Mar 28 [cited 2021 Apr 3]; Available from: https://onlinelibrary.wiley.com/doi/10.15252/embr.202051078

36. Brown DD, Littna E. RNA synthesis during the development of Xenopus laevis, the South African clawed toad. J Mol Biol. 1964 Jan;8(5):669–87.

37. Gregg JR, Ballentine R. Nitrogen metabolism of Rana pipiens during embryonic development. J Exp Zool. 1946 Oct;103(1):143–68.

38. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson A, Kampf C, Sjostedt E, Asplund A, Olsson I, Edlund K, Lundberg E, Navani S, Szigyarto CA-K, Odeberg J, Djureinovic D, Takanen JO, Hober S, Alm T, Edqvist P-H, Berling H, Tegel H, Mulder J, Rockberg J, Nilsson P, Schwenk JM, Hamsten M, von Feilitzen K, Forsberg M, Persson L, Johansson F, Zwahlen M, von Heijne G, Nielsen J, Ponten F. Tissue-based map of the human proteome. Science. 2015 Jan 23;347(6220):1260419–1260419.

39. Lahav-Baratz S, Sudakin V, Ruderman JV, Hershko A. Reversible phosphorylation controls the activity of cyclosome-associated cyclin-ubiquitin ligase. Proc Natl Acad Sci. 1995 Sep 26;92(20):9303–7.

40. Cundell MJ, Hutter LH, Nunes Bastos R, Poser E, Holder J, Mohammed S, Novak B, Barr FA. A PP2A-B55 recognition signal controls substrate dephosphorylation kinetics during mitotic exit. J Cell Biol. 2016 Aug 29;214(5):539–54.

41. Karsenti E, Bravo R, Kirschner M. Phosphorylation changes associated with the early cell cycle in Xenopus eggs. Dev Biol. 1987 Feb;119(2):442–53.

42. Wagih O, Sugiyama N, Ishihama Y, Beltrao P. Uncovering Phosphorylation-Based Specificities through Functional Interaction Networks. Mol Cell Proteomics. 2016 Jan;15(1):236–45.

43. Chou MF, Schwartz D. Biological Sequence Motif Discovery Using *motif-x*. In: Baxevanis AD, Petsko GA, Stein LD, Stormo GD, editors. Current Protocols in Bioinformatics [Internet]. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2011 [cited 2021 Apr 9]. p. bi1315s35. Available from: http://doi.wiley.com/10.1002/0471250953.bi1315s35

44. Schwartz D, Gygi SP. An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. Nat Biotechnol. 2005 Nov;23(11):1391–8.

45. Meggio F, Pinna LA. One-thousand-and-one substrates of protein kinase CK2? FASEB J. 2003 Mar;17(3):349–68.

46. Pfaff M, Anderer FA. Casein kinase II accumulation in the nucleolus and its role in nucleolar phosphorylation. Biochim Biophys Acta BBA - Mol Cell Res. 1988 Apr;969(1):100–9.

47. Wegierski T, Billy E, Nasr F, Filipowicz W. Bms1p, a G-domain-containing protein, associates with Rcl1p and is required for 18S rRNA biogenesis in yeast. RNA. 2001 Sep;7(9):1254–67.

48. Michalec B, Krokowski D, Grela P, Wawiórka L, Sawa-Makarska J, Grankowski N, Tchórzewski M. Subcellular localization of ribosomal P0-like protein MRT4 is determined by its N-terminal domain. Int J Biochem Cell Biol. 2010 May;42(5):736–48.

49. Horn H, Schoof EM, Kim J, Robin X, Miller ML, Diella F, Palma A, Cesareni G, Jensen LJ, Linding R. KinomeXplorer: an integrated platform for kinome biology studies. Nat Methods. 2014 Jun;11(6):603–4.

50. Miller ML, Jensen LJ, Diella F, Jorgensen C, Tinti M, Li L, Hsiung M, Parker SA, Bordeaux J, Sicheritz-Ponten T, Olhovsky M, Pasculescu A, Alexander J, Knapp S, Blom N, Bork P, Li S, Cesareni G, Pawson T, Turk BE, Yaffe MB, Brunak S, Linding R. Linear Motif Atlas for Phosphorylation-Dependent Signaling. Sci Signal. 2008 Sep 2;1(35):ra2–ra2.

51. Linding R, Jensen LJ, Ostheimer GJ, Miron IM, Diella F, Colwill K, Taylor L, Elder K, Metalnikov P, Nguyen V, Pasculescu A, Jin J, Park JG, Samson LD, Woodgett JR, Russell RB, Bork P, Yaffe MB, Pawson T. Systematic Discovery of In Vivo Phosphorylation Networks. :12.

52. Yamabhai M, Hoffman NG, Hardison NL, McPherson PS, Castagnoli L, Cesareni G, Kay BK. Intersectin, a Novel Adaptor Protein with Two Eps15 Homology and Five Src Homology 3 Domains. J Biol Chem. 1998 Nov;273(47):31401–7.

53. Rosario M. Activation of the Raf/MAP kinase cascade by the Ras-related protein TC21 is required for the TC21-mediated transformation of NIH 3T3 cells. EMBO J. 1999 Mar 1;18(5):1270–9.

54. Borgo C, D'Amore C, Cesaro L, Sarno S, Pinna LA, Ruzzene M, Salvi M. How can a traffic light properly work if it is always green? The paradox of CK2 signaling. Crit Rev Biochem Mol Biol. 2021 Apr 11;1–39.

55. Hoermann B, Kokot T, Helm D, Heinzlmeir S, Chojnacki JE, Schubert T, Ludwig C, Berteotti A, Kurzawa N, Kuster B, Savitski MM, Köhn M. Dissecting the sequence determinants for dephosphorylation by the catalytic subunits of phosphatases PP1 and PP2A. Nat Commun. 2020 Dec;11(1):3583.

56. Huttlin EL, Jedrychowski MP, Elias JE, Goswami T, Rad R, Beausoleil SA, Villén J, Haas W, Sowa ME, Gygi SP. A Tissue-Specific Atlas of Mouse Protein Phosphorylation and Expression. Cell. 2010 Dec;143(7):1174–89.

57. Villen J, Beausoleil SA, Gerber SA, Gygi SP. Large-scale phosphorylation analysis of mouse liver. Proc Natl Acad Sci. 2007 Jan 30;104(5):1488–93.

58. Sharma K, D'Souza RCJ, Tyanova S, Schaab C, Wiśniewski JR, Cox J, Mann M. Ultradeep Human Phosphoproteome Reveals a Distinct Regulatory Nature of Tyr and Ser/Thr-Based Signaling. Cell Rep. 2014 Sep;8(5):1583–94.

59. McCloy RA, Parker BL, Rogers S, Chaudhuri R, Gayevskiy V, Hoffman NJ, Ali N, Watkins DN, Daly RJ, James DE, Lorca T, Castro A, Burgess A. Global Phosphoproteomic Mapping of Early Mitotic Exit in Human Cells Identifies Novel Substrate Dephosphorylation Motifs□S. :19.

60. Malik R, Lenobel R, Santamaria A, Ries A, Nigg EA, Körner R. Quantitative Analysis of the Human Spindle Phosphoproteome at Distinct Mitotic Stages. J Proteome Res. 2009 Oct 2;8(10):4553–63.

61. Lu T-J, Lai W-Y, Huang C-YF, Hsieh W-J, Yu J-S, Hsieh Y-J, Chang W-T, Leu T-H, Chang W-C, Chuang W-J, Tang M-J, Chen T-Y, Lu T-L, Lai M-D. Inhibition of Cell Migration by Autophosphorylated Mammalian Sterile 20-Like Kinase 3 (MST3) Involves Paxillin and Protein-tyrosine Phosphatase-PEST. J Biol Chem. 2006 Dec 15;281(50):38405–17.

62. Thompson BJ, Sahai E. MST kinases in development and disease. J Cell Biol. 2015 Sep 14;210(6):871–82.

63. Mardakheh FK, Self A, Marshall CJ. RHO binding to FAM65A regulates Golgi reorientation during cell migration. J Cell Sci. 2016 Dec 15;129(24):4466–79.

64. ten Klooster JP, Jansen M, Yuan J, Oorschot V, Begthel H, Di Giacomo V, Colland F, de Koning J, Maurice MM, Hornbeck P, Clevers H. Mst4 and Ezrin Induce Brush Borders Downstream of the Lkb1/Strad/Mo25 Polarization Complex. Dev Cell. 2009 Apr;16(4):551–62.

65. Lai Z-C, Wei X, Shimizu T, Ramos E, Rohrbaugh M, Nikolaidis N, Ho L-L, Li Y. Control of Cell Proliferation and Apoptosis by Mob as Tumor Suppressor, Mats. Cell. 2005 Mar;120(5):675–85.

66. Ni L, Zheng Y, Hara M, Pan D, Luo X. Structural basis for Mob1-dependent activation of the core Mst–Lats kinase cascade in Hippo signaling. Genes Dev. 2015 Jul 1;29(13):1416–31.

67. Wu Q, Zhou X, Li P, Wang W, Wang J, Tan M, Tao L, Qiu J. High NRBP1 expression promotes proliferation and correlates with poor prognosis in bladder cancer. J Cancer. 2019;10(18):4270–7.

68. De Langhe S, Haataja L, Senadheera D, Groffen J, Heisterkamp N. Interaction of the small GTPase Rac3 with NRBP, a protein with a kinase-homology domain. Int J Mol Med [Internet]. 2002 May 1 [cited 2021 May 16]; Available from: http://www.spandidos-publications.com/10.3892/ijmm.9.5.451

69. Huang C, Rajfur Z, Yousefi N, Chen Z, Jacobson K, Ginsberg MH. Talin phosphorylation by Cdk5 regulates Smurf1-mediated talin head ubiquitylation and cell migration. Nat Cell Biol. 2009 May;11(5):624–30.

70. Zhao Z, Manser E. PAK family kinases: Physiological roles and regulation. Cell Logist. 2012 Apr;2(2):59–68.

71. Gatti A, Huang Z, Tuazon PT, Traugh JA. Multisite Autophosphorylation of p21-activated Protein Kinase γ-PAK as a Function of Activation. J Biol Chem. 1999 Mar;274(12):8022–8.

72. Jung J-H, Traugh JA. Regulation of the Interaction of Pak2 with Cdc42 via Autophosphorylation of Serine 141. J Biol Chem. 2005 Dec 2;280(48):40025–31.

73. Keller R, Shook D. Gastrulation in Amphibians. In: Gastrulation: From Cells to Embryo. p. 171–203.

74. Stack JH, Newport JW. Developmentally regulated activation of apoptosis early in Xenopus gastrulation results in cyclin A degradation during interphase of the cell cycle. Development. 1997 Aug 15;124(16):3185–95.

75. Olsen JV, Vermeulen M, Santamaria A, Kumar C, Miller ML, Jensen LJ, Gnad F, Cox J, Jensen TS, Nigg EA, Brunak S, Mann M. Quantitative Phosphoproteomics Reveals Widespread Full Phosphorylation Site Occupancy During Mitosis. Sci Signal. 2010 Jan 12;3(104):ra3–ra3.

76. Lim MY, O'Brien J, Paulo JA, Gygi SP. Improved Method for Determining Absolute Phosphorylation Stoichiometry Using Bayesian Statistics and Isobaric Labeling. J Proteome Res. 2017 Nov 3;16(11):4217–26.

77. Wu R, Haas W, Dephoure N, Huttlin EL, Zhai B, Sowa ME, Gygi SP. A large-scale method to measure absolute protein phosphorylation stoichiometries. Nat Methods. 2011 Aug;8(8):677–83.

78. Gabut M, Bourdelais F, Durand S. Ribosome and Translational Control in Stem Cells. Cells. 2020 Feb 21;9(2):497.

79. Ingolia NT, Lareau LF, Weissman JS. Ribosome Profiling of Mouse Embryonic Stem Cells Reveals the Complexity and Dynamics of Mammalian Proteomes. Cell. 2011 Nov;147(4):789–802.

80. Paulo JA, O'Connell JD, Gygi SP. A Triple Knockout (TKO) Proteomics Standard for Diagnosing Ion Interference in Isobaric Labeling Experiments. J Am Soc Mass Spectrom. 2016 Oct;27(10):1620–5.

81. Bowes JB, Snyder KA, Segerdell E, Jarabek CJ, Azam K, Zorn AM, Vize PD. Xenbase: gene expression and improved integration. Nucleic Acids Res. 2010 Jan;38(suppl_1):D607–12.

82. Peshkin L, Gupta M, Ryazanova L, Wühr M. Bayesian Confidence Intervals for Multiplexed Proteomics Integrate Ion-statistics with Peptide Quantification Concordance. Mol Cell Proteomics. 2019 Oct 1;18(10):2108–20.

83. Session AM, Uno Y, Kwon T, Chapman JA, Toyoda A, Takahashi S, Fukui A, Hikosaka A, Suzuki A, Kondo M, van Heeringen SJ, Quigley I, Heinz S, Ogino H, Ochi H, Hellsten U, Lyons JB, Simakov O, Putnam N, Stites J, Kuroki Y, Tanaka T, Michiue T, Watanabe M, Bogdanovic O, Lister R, Georgiou G, Paranjpe SS, van Kruijsbergen I, Shu S, Carlson J, Kinoshita T, Ohta Y, Mawaribuchi S, Jenkins J, Grimwood J, Schmutz J, Mitros T, Mozaffari SV, Suzuki Y, Haramoto Y, Yamamoto TS, Takagi C, Heald R, Miller K, Haudenschild C, Kitzman J, Nakayama T, Izutsu Y, Robert J, Fortriede J, Burns K, Lotay V, Karimi K, Yasuoka Y, Dichmann DS, Flajnik MF, Houston DW, Shendure J, DuPasquier L, Vize PD, Zorn AM, Ito M, Marcotte EM, Wallingford JB, Ito Y, Asashima M, Ueno N, Matsuda Y, Veenstra GJC, Fujiyama A, Harland RM, Taira M, Rokhsar DS. Genome evolution in the allotetraploid frog Xenopus laevis. Nature. 2016 Oct;538(7625):336–43.

84. Savova V, Pearl EJ, Boke E, Nag A, Adzhubei I, Horb ME, Peshkin L. Transcriptomic insights into genetic diversity of protein-coding genes in X. laevis. Dev Biol. 2017 Apr;424(2):181–8.

85. Liao Y, Wang J, Jaehnig EJ, Shi Z, Zhang B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. Nucleic Acids Res. 2019 Jul 2;47(W1):W199–205.

86. Ivo Petras. Total Least Squares Method [Internet]. MATLAB Central File Exchange; [cited 2021 May 23]. Available from: https://www.mathworks.com/matlabcentral/fileexchange/31109-total-least-squares-method