Supplementary Information Appendix

Proteomics of phosphorylation and protein dynamics during fertilization and meiotic exit in the *Xenopus* egg

Marc Presler[1], Elizabeth Van Itallie[1], Allon M. Klein[1], Ryan Kunz[2], Margaret L. Coughlin[1], Leonid Peshkin[1], Steven P. Gygi[2], Martin Wühr[1-4], Marc W. Kirschner[1,4]

Affiliations:
[1] Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA
[2] Department of Cell Biology, Harvard Medical School, Boston, MA 02115, USA
[3] Department of Molecular Biology, The Lewis-Sigler Institute, Princeton University, Princeton, NJ 08544
[4] Correspondence: wuhr@princeton.edu, marc@hms.harvard.edu

Table of Contents:

Supplementary Figures

Supplementary Methods:

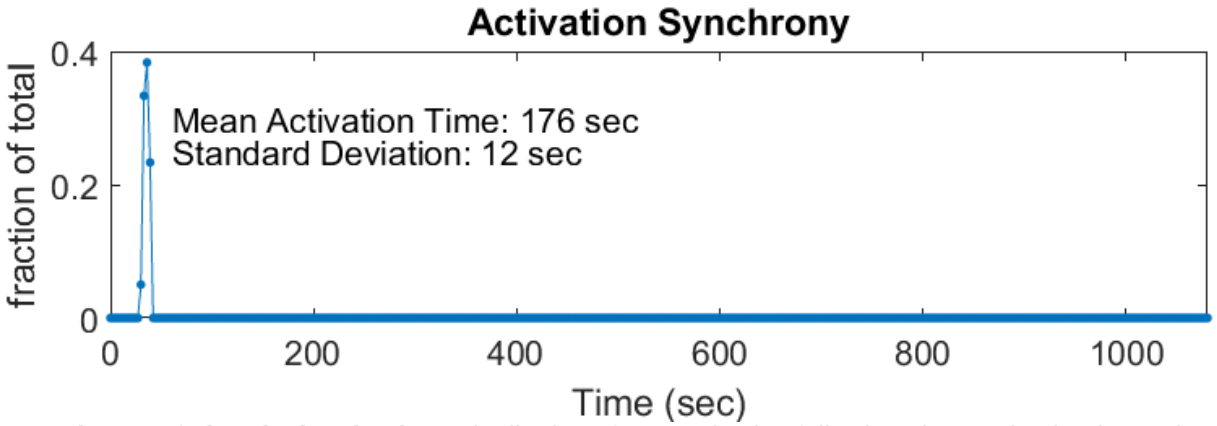

Figure S1: **Synchrony of electrical activation**. Distribution of egg activation following electro-shock. The scale of the X-axis is 18 minutes (displayed in seconds). Of the 400 eggs used in each experiment, ~70 were filmed per experiment in parallel with the collection of the time course. Images taken every 5 seconds (see Movies S1, S2, S3 and S4 filmed for each replicate; 14 frames/sec). The movies show de-jellied *Xenopus* eggs mostly oriented with the pigmented animal pole facing upward. The calcium waves and cortical contractions are visible. Of the ~250 eggs filmed across all replicates, only two failed to activate. The distribution of egg activation is shown above, assayed by the calcium wave (surface contraction wave) crossing the midpoint of the animal pole. Given the resolution of the movie, we conservatively estimate that all eggs which activate do so with a standard deviation of less than 15 seconds. Eggs that are inverted so the vegetal (i.e., white) side is facing upward are not quantified.

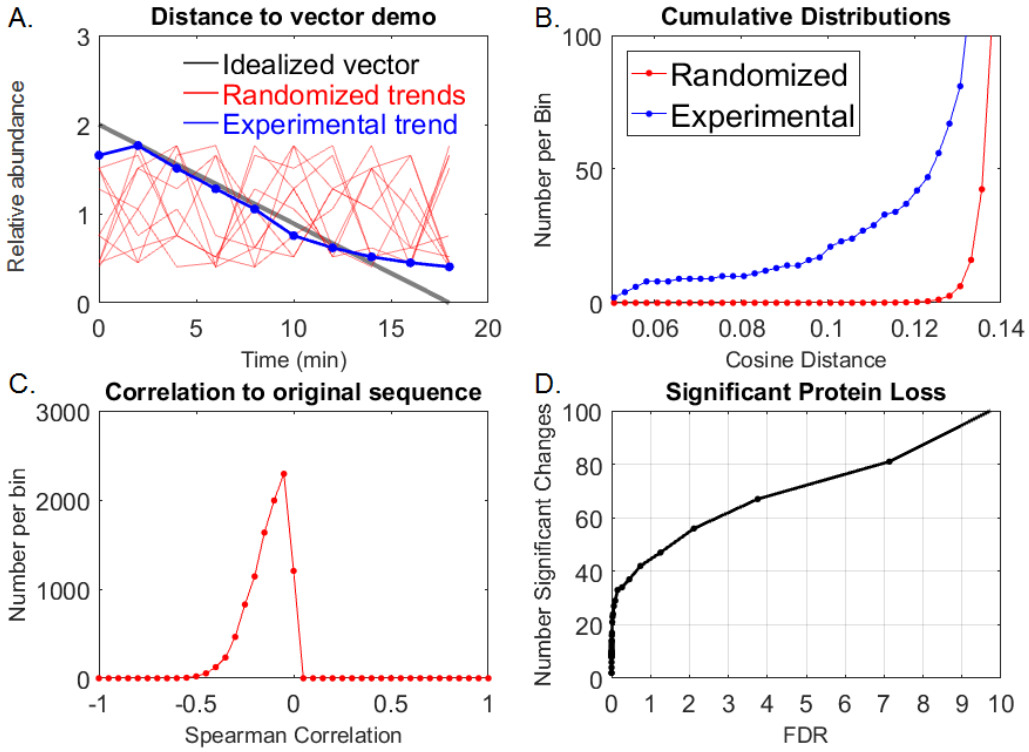

Figure S2: **Estimating a false discovery rate for protein loss**. Principles and visualization of the FDR calculation. See Supplementary Methods for further explanation of the implementation to the full dataset with replicates.

A) Plots to demonstrate the geometric distance measurements used in calculating false discovery rate. Here, an idealized 10-element degradation vector (gray, mean normalized to match the experimental data) is shown. The Cosine distance of the experimental trends (blue, in this case, Cyclin B1) was measured to the idealized vector. The time sequence of the experimental trend was then iteratively randomized (red) and Cosine distances re-calculated each time.

B) The cumulative distributions of distances for all proteins (experimental) and pooled randomized trends from the 10,000 iterations. These distributions are used to calculate the FDR at a given cosine distance to the idealized vector (Randomized/(Experimental + Randomized) *100). Since the randomized values are pooled, we apply one cutoff to the entire dataset. Multiple hypothesis correction is therefore not required.

C) Distribution of the correlation of the randomized values to the original time sequence. The original time sequence is “0, 2, 4,..18min”, which is scrambled during the randomizations. We excluded any randomized sequence that was positively correlated with the original sequence (e.g., “0, 4*, 2*,...18min”, which is essentially identical to the original order). This was done to avoid overestimating the false discovery from randomly recreating the original order of the time series and counting this as a “false discovery.”

D) Number of proteins identified as significantly decreasing from the distributions in S2B. The FDR is calculated by Protein ID, then combined by gene symbol afterward (i.e., we identify multiple “forms” of Cyclin B1, which may be splice isoforms of pseudoalleles, which are later averaged to one gene symbol). We identify 41 protein ID’s which significantly decrease (Table S1) which map to 29 unique gene symbols (Table S2).

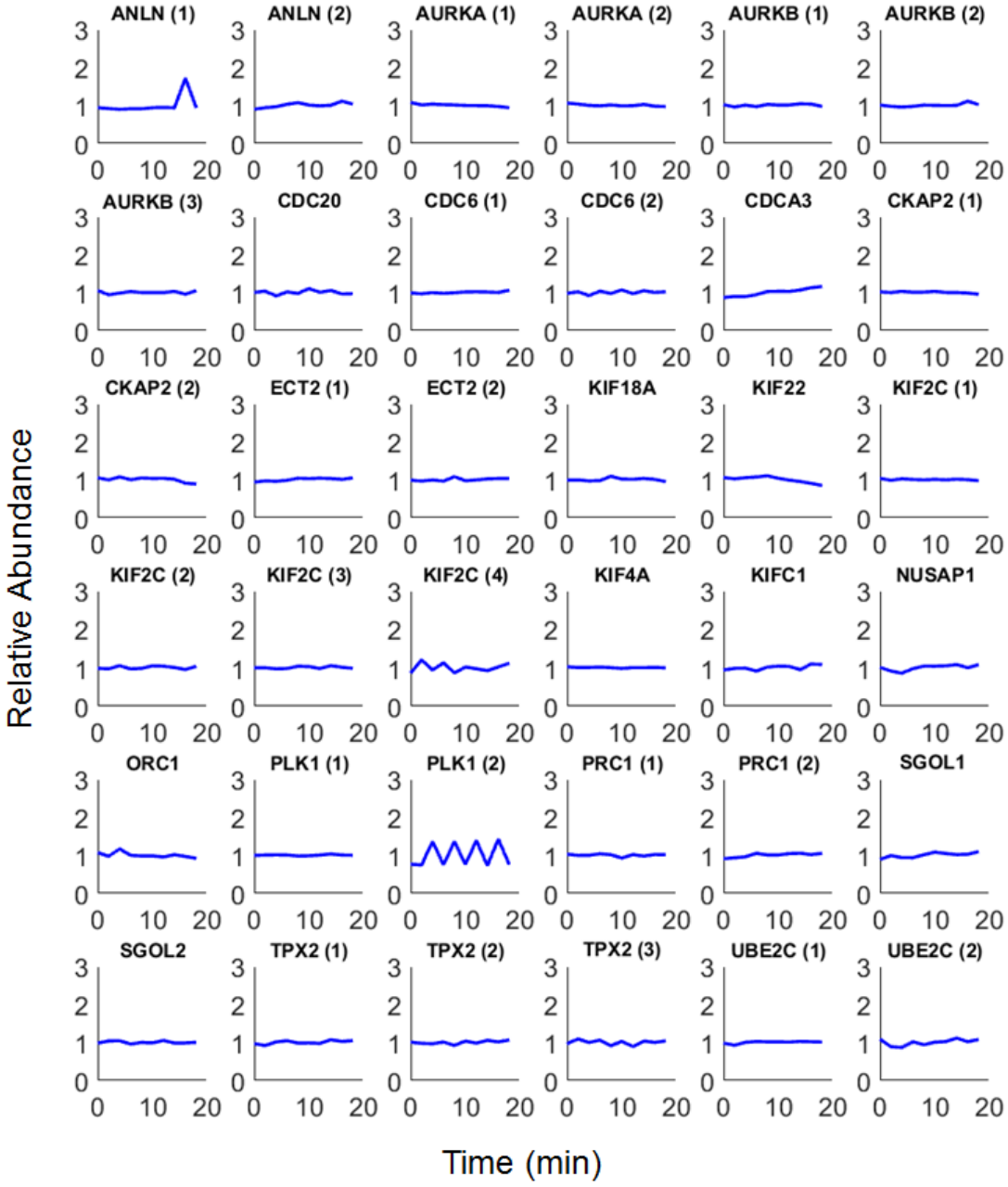

Figure S3: **Known APC/C^{Cdh1} substrates are stable post-fertilization**. Time series data of known targets of the APC/C^{Cdh1} following egg activation. The substrates are stable (i.e., do not pass the FDR threshold of 1% for decreasing, see Fig. S2). Repeated gene symbols indicate multiple protein IDs, which are likely alloalles or splice variants. The proteins are CDC20, Plk1, Aurora A/B (see citations in main text), ANLN (1), Cdc6 (2), , CDCA3 (TOME1) (3), Ckap2 (4), Ect2(5), KIF18A, KIF2C, KIF4A, KIFC1(6), NUSAP1 (7), ORC1 (8), Ase1/PRC1 (9), Shugoshin (10), TPX2 (11), UBE2C (12). KIF22 (13) is the only substrate that shows some marginal reduction in abundance, but it does not pass our FDR threshold and was previously shown in the literature to not interact with CDC20.

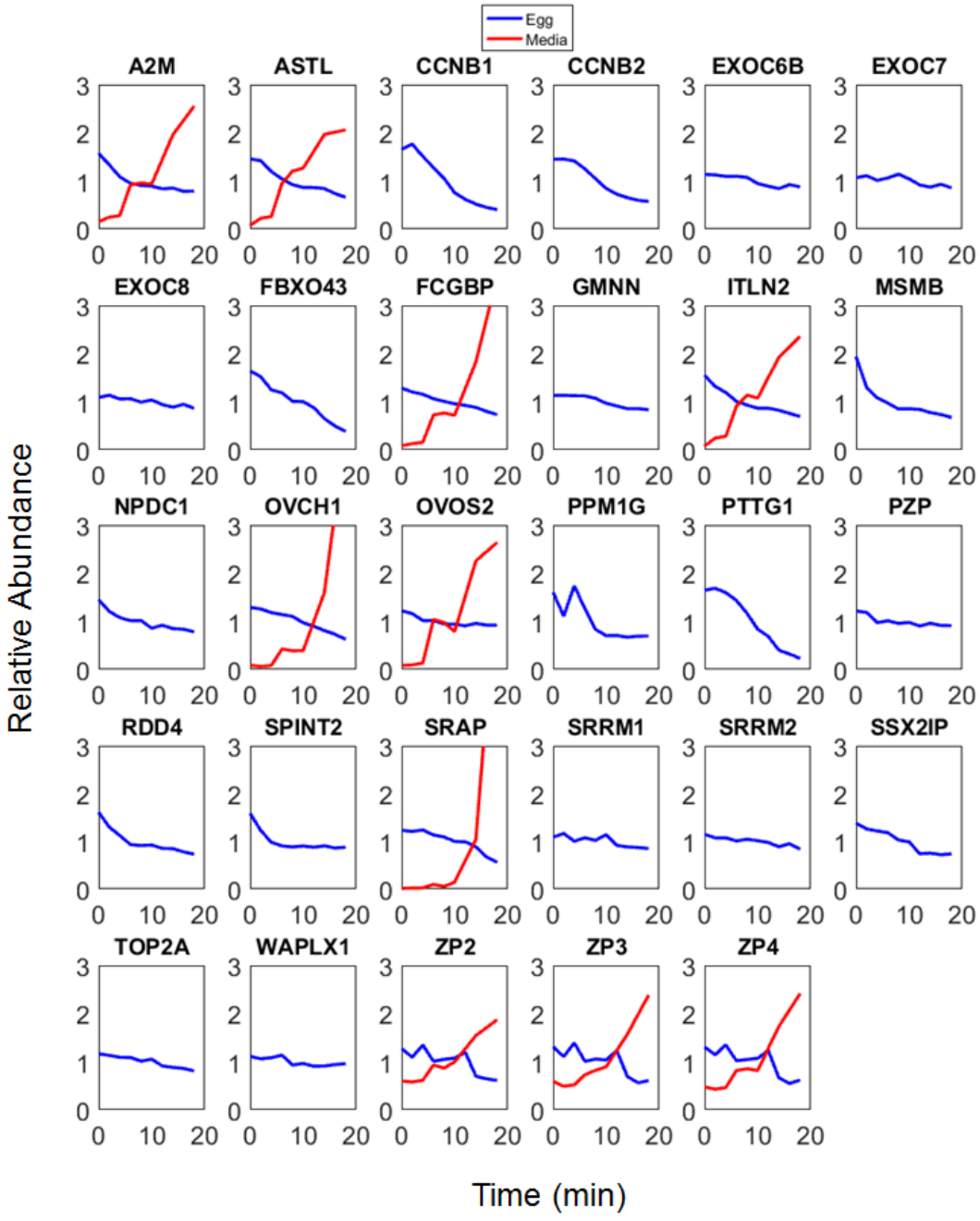

Figure S4: **Degradation and release of proteins at 1% FDR**. Time series data of the 29 genes which significantly decrease at 1% FDR following egg activation. CCNB1, CCNB2, GMNN, FBX043 (Emi2), and PPTG1 (Securin) are known APC/C substrates. Blue trends decrease abundance in the egg. Red trends increase in the supernatant; if no red trend is shown, it was not detected in the egg media. These are further classified in Table S3. For ZP2, 3, and 4, the jagged decrease from the egg data is not seen in its appearance into the supernatant. Therefore, the noise in the ZP trends is consistent with a technical rather than biological explanation. It is not clear why ZP proteins are more prone to this noise than other released proteins. We suspect this is because the ZP proteins are released by proteolysis rather than exocytosis. This causes greater inconsistencies in the removal of ZP fragments from the egg surface during harvesting.

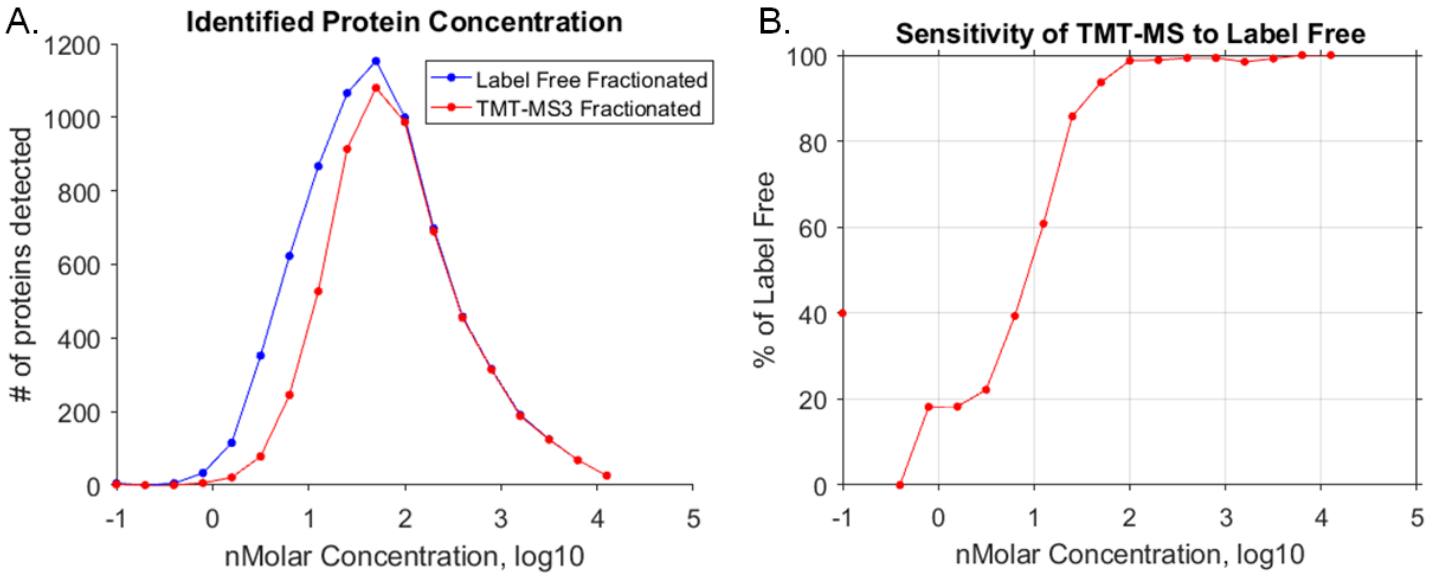

Figure S5: **Sensitivity of label-free and multiplex TMT-MS3 methods**.
A) Histograms comparing the distribution of abundances of proteins detected by label-free (blue) and fractionated TMT-MS3 (red) methods in the *Xenopus laevis* egg (all four replicates are pooled for the TMT distributions). The label-free method is the most sensitive technique available, but still does not detect the entire proteome. This is evident from by the fact that we can detect many more mRNA species via sequencing than protein species via MS (14). Additionally, there are proteins that are known to be expressed but are not detected in the label-free dataset (e.g., Axin). The fractionated TMT-MS3 method used in this study to quantify the proteins captures a subset of the label-free distribution, but with substantial overlap.

B) Plot comparing the percentage of proteins identified in the TMT-MS compared to the label-free as a function of concentration. The fractionated TMT data is essentially equivalent to 100nM, and at 50nM, ~95% of proteins identified with the label-free method are also seen with TMT-MS. All known APC substrates are at least 60nM. At 10nM, we detect 50% of the proteins with TMT-MS that we do with label-free. Though detection is less reliable at these lower concentrations, EMI2 (10nM), for example, is still detected in 2 of the 4 replicates. The 40% overlap seen at 0.1 nM (-1 in log10) is due to under-sampling, as the numbers are very small at those concentrations.

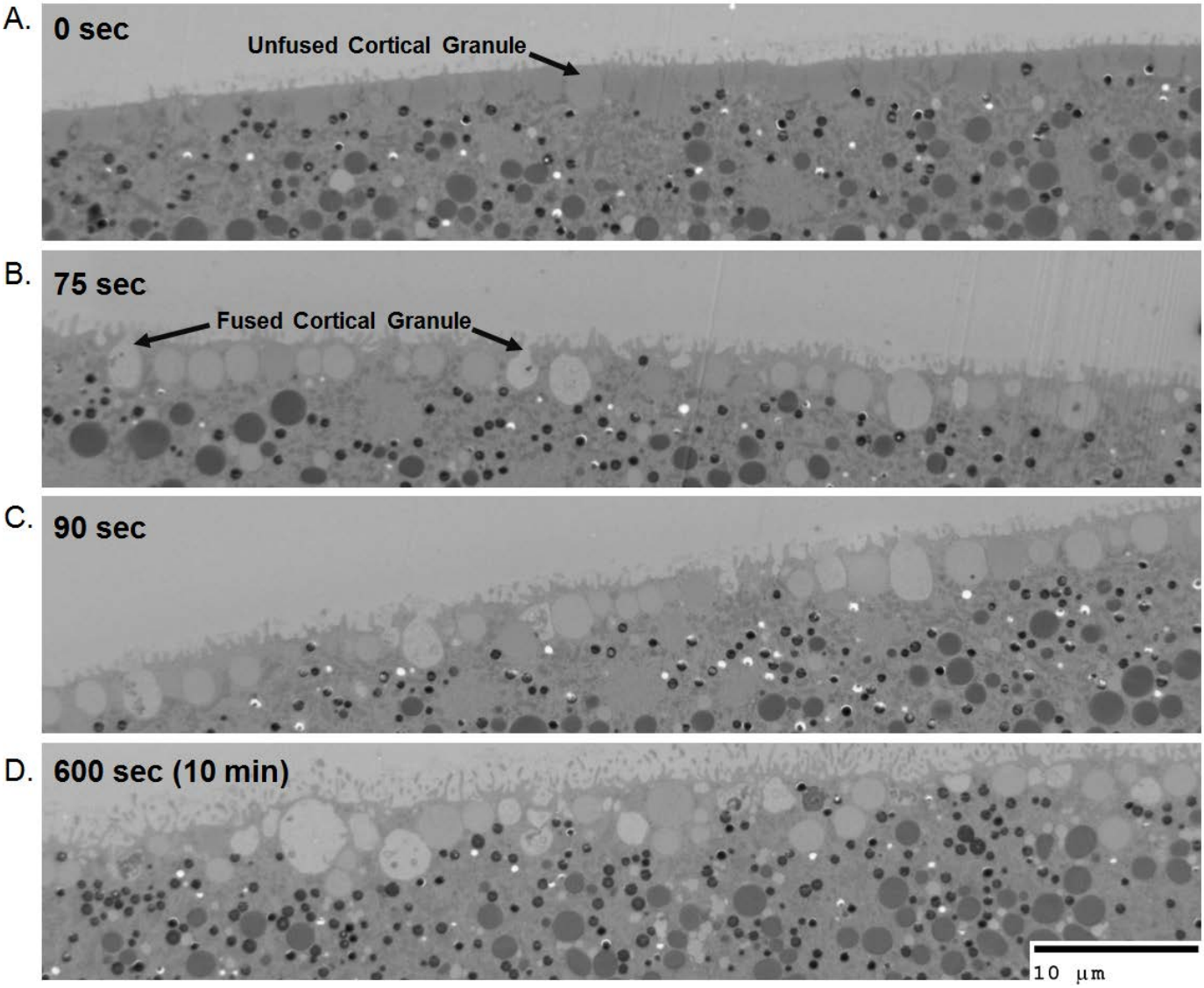

Figure S6: **Egg cortex time series imaging with electron microscopy**. Thin sections of the outer membrane of the animal pole of *Xenopus laevis* eggs at 690x magnification (scale bar indicated in bottom right of figure). Examples of unfused and fused cortical granules are indicated with arrows in A and B. The vitelline membrane is not present in these images (See Supplementary Methods for experimental details).

A) Outer membrane of the egg before activation occurs. No fused cortical granules are seen here.

B) 75 seconds post egg activation.

C) 90 seconds post egg activation.

D) 10 minutes post egg activation. ~50-60% of cortical granules depicted have released their contents through fusion with the outer membrane. At this time, fused cortical granules appear either as invaginations of the membrane or as low-density spheres (when the actual fusion event is not visible in the section).

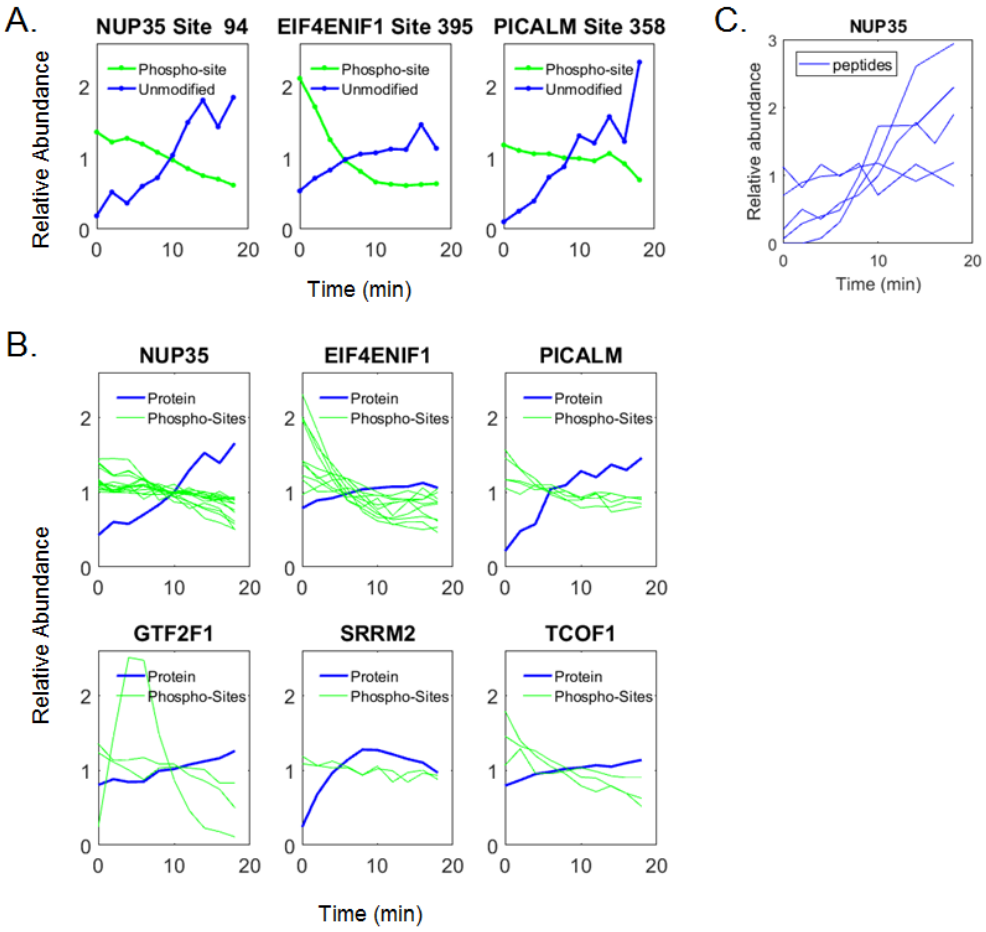

Figure S7: **Evidence that dephosphorylation causes specious increasing protein trends**. A small number of proteins significantly increase after fertilization (Table S4) at 1% FDR. Rather than synthesis, this can occur on stable proteins because of the loss of a modification on a particular peptide. The unmodified form rises reciprocally due to mass conservation (i.e., the unmodified form is created as the modified is destroyed). Since the protein trends are the sum of all unmodified peptides trends, a large change in some peptides can give an apparent increase despite the stability of the other peptides (e.g., S7C).

A) Times series of an unmodified peptide (blue) that increases in signal with matching a corresponding phospho-site residue (green) that decreases in signal for three of the proteins with apparently increasing proteins. This is direct evidence that the increasing signal is due to dephosphorylation rather than protein synthesis.

B) Time series of 6 of the 7 proteins (blue line) which apparently increase that have multiple dephosphorylated sites (green) detected on the proteins. The top 3 are the same proteins as S7A, but with additional sites shown. For the bottom three we do not detect exact matches of reciprocal modified and unmodified peptides. Nevertheless, the multiple phospho-sites that are dephosphorylated for each protein is consistent with the phenomenon discussed in S7A. For the 7th protein that increases (POLR2A, a RNA polymerase II subunit, see Table S4), we detect only one phospho-site, which is stable. However, as there are many other polymerases dephosphorylated in the dataset, it is likely POLR2A is dephosphorylated as well.

C) All the peptides for NUP35 identified in the study. This is an example of how the unmodified peptides (one of them shown in S7A) are dynamic from the loss of a modification, while the other two reflect the true flat trend of the overall protein. In this case and several others, using the median value at each time point, or even the median fold change, would still show a specious increase, as multiple peptides are affected by the same phenomenon from the loss of a modification. Importantly, the same general principle that produces this artifact can occur with different forms multiply phosphorylation peptides (Fig. S14) and is also important to the phospho-stoichiometry calculation (Fig. S8-10, Supplementary Methods).

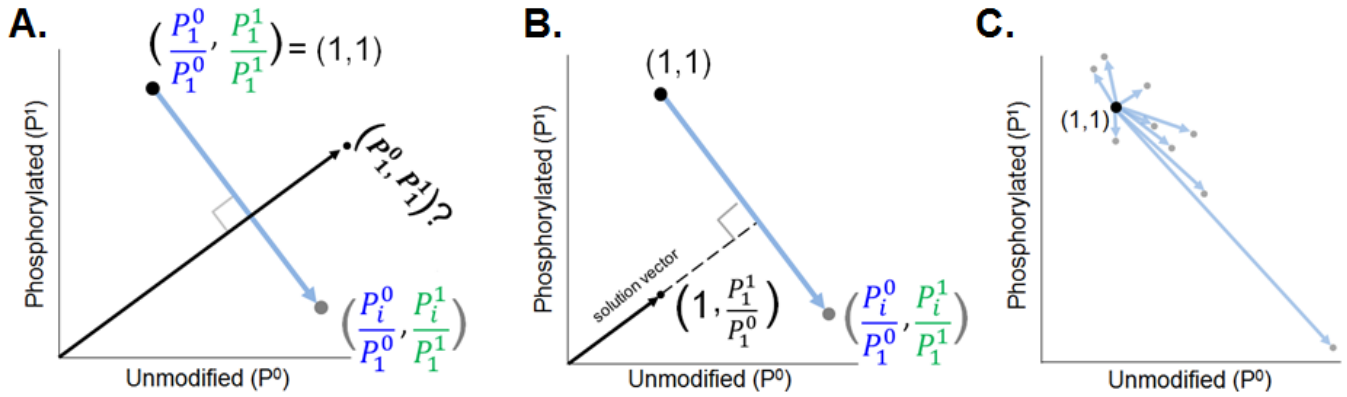

Figure S8: **Graphical representation of the relationship of relative ratios to absolute phospho-forms.** See Supplementary Methods for equations and term definitions.

A) In an example with 2 phospho-forms (unmodified and single phospho) and two conditions, the MS data for the unmodified (P^0) and phosphorylated (P^1) forms is normalized to the reference time point 1 (Eq. 1 from Supplement Methods) and plotted as coordinates on the P^0 vs P^1 plane. The subspace defined by the point (1, 1) and the point $\left(\frac{P_i^0}{P_1^0}, \frac{P_i^1}{P_1^1}\right)$ (shown as a light blue line) is orthogonal to the vector containing the absolute values of P^0 and P^1 forms at the reference time point (Eq. 4, 5; i.e., the dot product is zero). The values of P^0 and P^1 are not known.

B) The absolute values of the species $[P_1^0, P_1^1]$ lay on a line with the slope of $\frac{P_1^1}{P_1^0}$ passing through the origin. This can be seen graphically by normalizing the vector $[P_1^0, P_1^1]$ by it first element, to give the vector $\left[1, \frac{P_1^1}{P_1^0}\right]$ shown in black. The P^1/P^0 ratio is therefore the negative inverse of the slope of the line containing the points defined by the known ratios, and is sufficient to calculate stoichiometry (this step explained further Supplementary Methods).

C) Demonstrating the graphical representation of an over determined system when the number of measurements is greater than the number of phospho-forms. Each additional condition gives another vector (or linear equation). In this case, ten conditions give 9 vectors from the reference condition. The solution to this system can be estimated with regression (i.e., fitting a line). See S9 for demonstration.

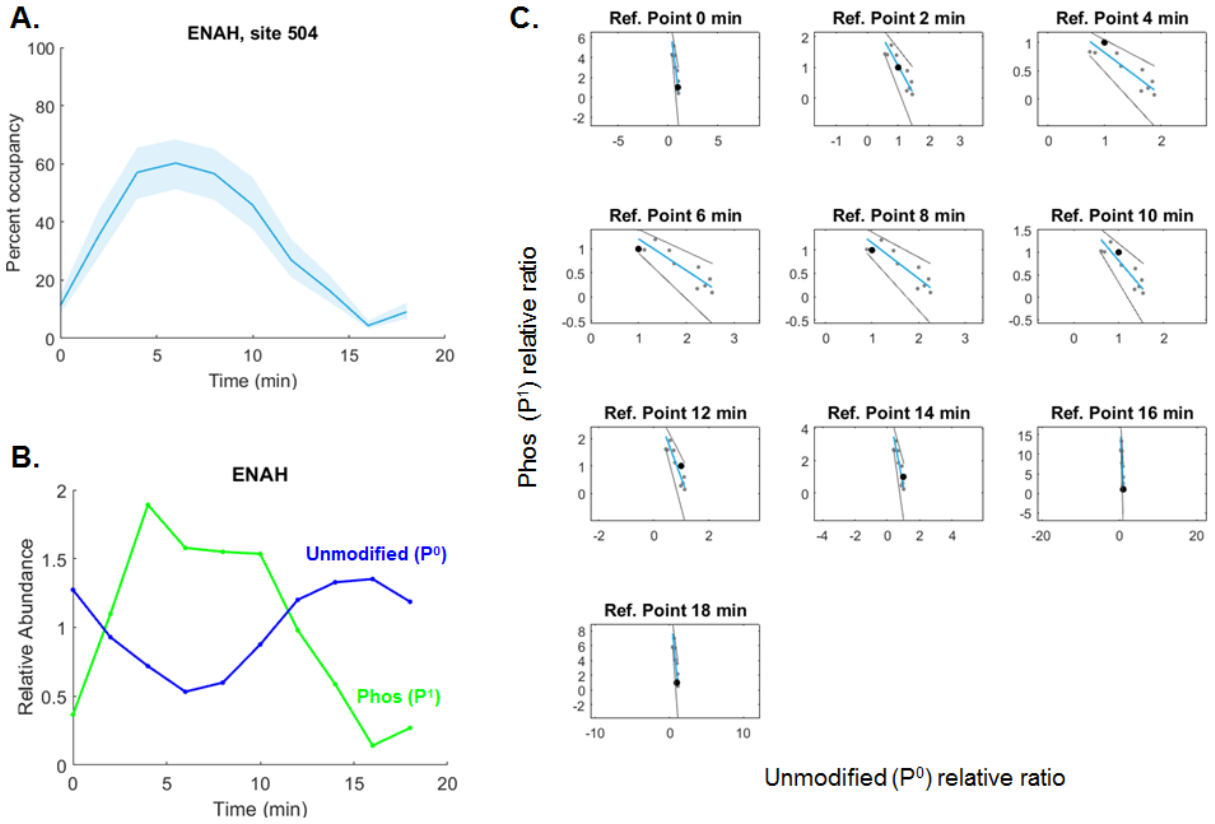

Figure S9: **Calculating phospho-stoichiometry for multiplexed measurements.** Visualization of how stoichiometry is solved for a single site with multiple time points.

A) Example time series data of phospho-site stoichiometry of the residue 504 of the protein ENAH. The occupancy with 95% confidence intervals (shaded) is shown in light blue.

B) The same data, but plotting the unmodified (P^0) and phosphorylated (P^1) trends (blue and green, respectively) that are used to calculate the stoichiometry.

C) Estimating stoichiometry for the whole time course. Each subplot is comparable how to the visualization in Fig. S8C.To obtain the P^1/P^0 ratio for the whole time course, the reference time point was switched and the analysis is repeated for each time point. The reference point (1,1) is plotted in black, while the rest of the data are plotted in grey. The optimal line, obtained with Total Least Squares regression, is shown in light blue. The 95% confidence intervals are obtained by bootstrapping at each time point and are plotted in gray. Note how the angle of the slope corresponds with the shifting stoichiometry over the time course. The negative inverse of these lines (i.e., orthogonal) gives the P^1/P^0 ratio and its confidence intervals (as in Fig. S8B).

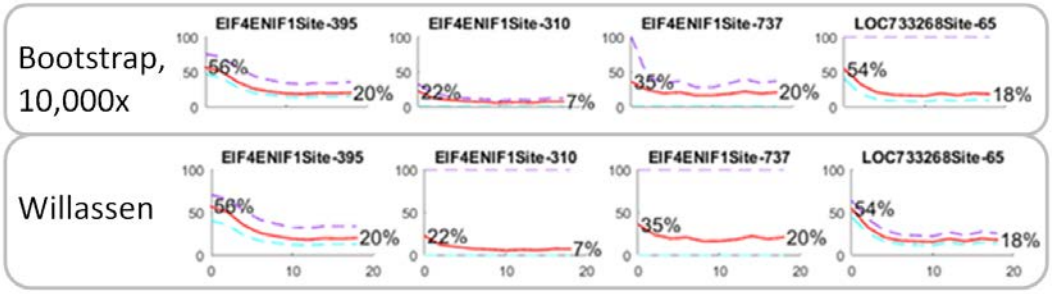

Figure S10. **Bootstrapped confidence intervals compared to Willassen confidence intervals.** Comparison of two different methods for calculating confidence intervals. Optimal estimate is shown in red, upper confidence intervals are shown in violet, and lower confidence intervals in cyan. For confidence interval validation, we minimized in the X and Y axis separately. Taking the wider confidence interval of the two gives an over-estimate of the 95% confidence interval of regressing in both axis (15, 16) (i.e., the Willassen solution). Bootstrapping confidence intervals are almost always narrower than the overestimate from the Willassen solution, as expected. The Willassen solution is only narrower when the fit of the line is determined by only one or two dynamic points. A very different slope results in the situation where these points are not sampled during bootstrapping, widening the resulting distribution. For the Willassen calculation, the confidence intervals are unaffected by one or two points defining the line as there is no subsampling.

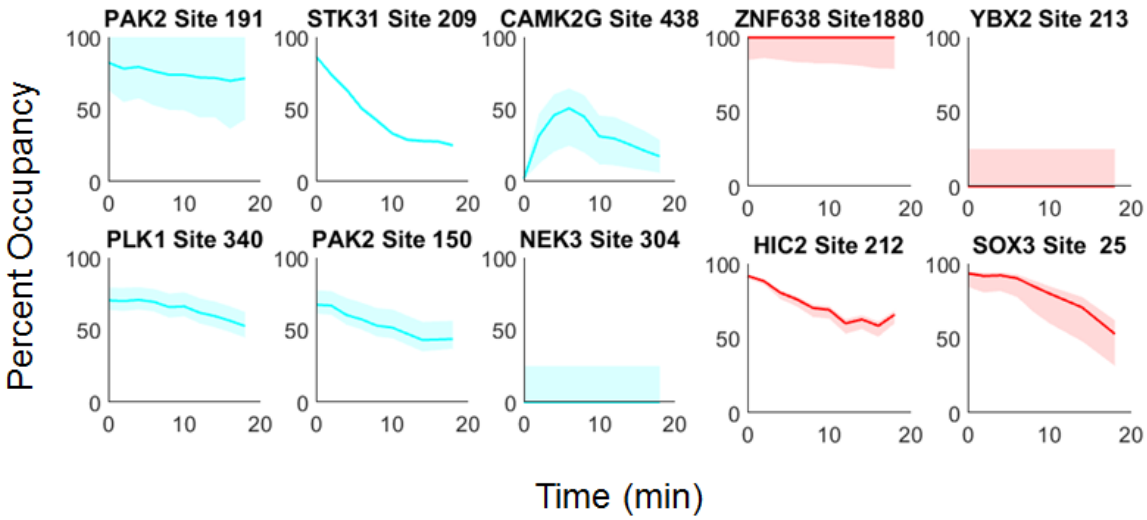

Figure S11: **Panel of kinase and transcription factor phospho-occupancy dynamics**. Time series of phospho-site occupancy dynamics on kinases (cyan) and transcription factors (red) following egg activation. Shaded areas are 95% confidence intervals.

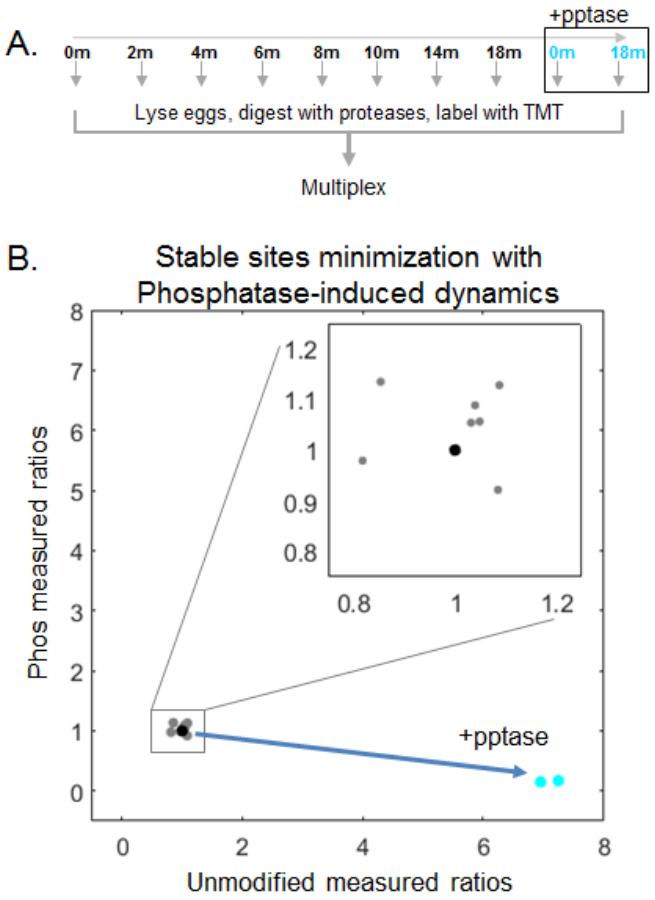

Figure S12: **Phosphatase treatment enables stoichiometry calculation for stable sites.**
A) Experimental schematic for the phosphatase treatment experiments. The time course was repeated as in Fig. 1B. The 12 and 16 minute time pointes were skipped so the experiment could be performed with one TMT 10plex. Replicates of the 0 and 18 min time points were collected and phosphatase treated in parallel with the protease digestions, multiplexed with the samples, and then analyzed with MS (see Methods). Given the multiplexing step, it was important to use a temperature-sensitive phosphatase that could be inactivated with heat treatment so untreated samples would remain unaffected after mixing without adding a purification step.

B) Visualization of data on the unmodified vs phosphoform plane for a stable site as in Fig. 4B. The bold point is the normalized reference time point (see Fig. S8) and the cyan points indicate phosphatase-treated conditions. With low dynamics, perfect data would collapse onto a single point, rather than forming a line. Noise will cause scatter (see inset), but no clear line (as in Fig. S8/9) will be formed. The phosphatase treated conditions allow for reliable minimization, though the fit is defined nearly exclusively by these two points.
For the confidence of the untreated conditions, the noise dominates when the phosphatase-treated conditions are not sampled during the bootstrapping. This can lead to two qualitatively similar stable sites giving very different confidence interval widths in the untreated conditions. For example, in perfect data, even very small fluctuations between the P^0 and P^1 forms are expected to be reciprocal to satisfy the conservation constraint. A “noisy” stable trend would have small fluctuations in P^0 and P^1 that change in the same direction. When the phosphatase conditions are not sampled during bootstrapping, the fit is extremely sensitive to this type of noise.

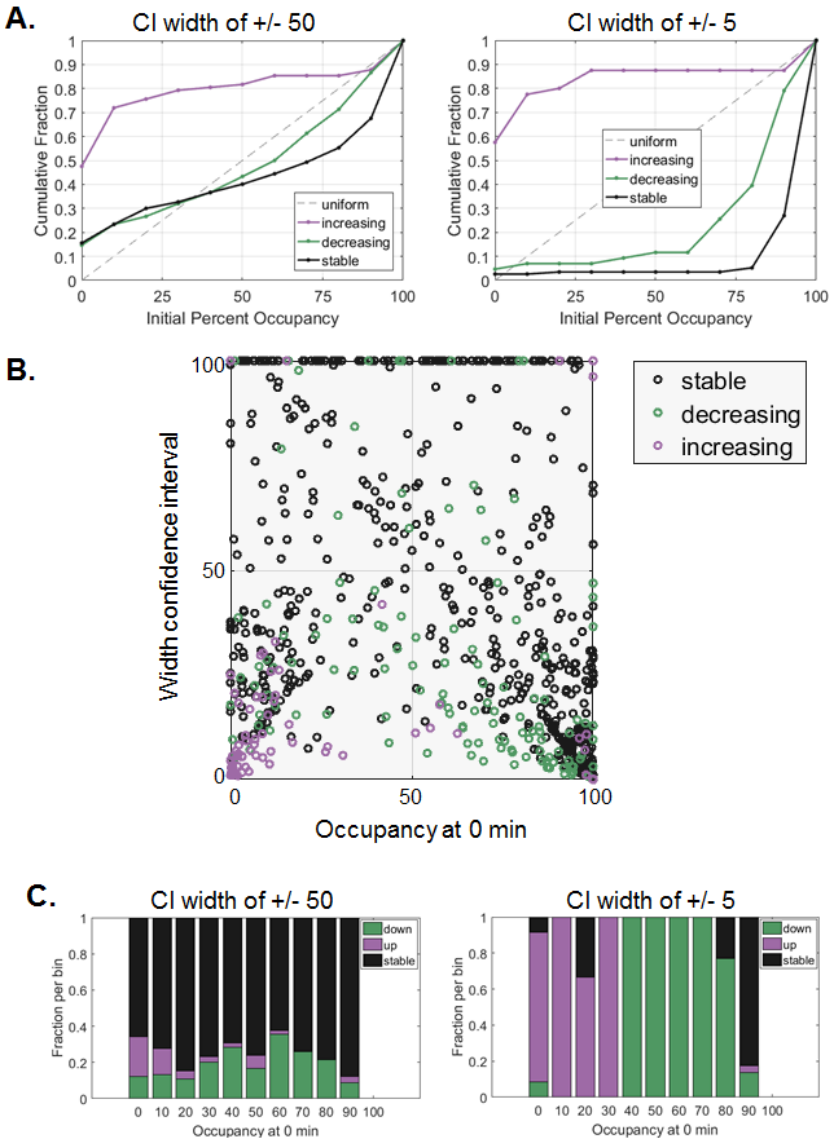

Figure S13: **Intermediate occupancies are preferentially depleted with narrow confidence interval cutoffs.**
A) Cumulative distributions of the occupancies at 0 minutes for phospho-site trends which increase (purple), decrease (green), or are stable (black) (Fig. 5D) with no cutoff confidence interval width (left) and a very stringent cutoff of +/- 5% (right). This demonstrates that narrow confidence interval cutoffs bias the distributions against intermediate occupancies. This is because the confidence intervals are limited to 0 and 100 (values outside this range are meaningless). Sites with intermediate occupancies have more potential for wide confidence intervals as they are less restricted by this limit. This is a problem for both dynamic and stable sites. However, because the estimate for stable sites is dominated by noise of the untreated conditions (see Fig. S12), the stable sites are particularly susceptible to this systematic bias against intermediate occupancies when applying cutoffs.

B) To demonstrate this bias, a scatterplot of the initial occupancy of a phospho-site at zero minutes is plotted against the width of the confidence interval for all sites where stoichiometry was calculated. The stable sites indeed show many sites with intermediate occupancies, but these sites often have more error (i.e., lie in the top half of the plot).

C) The same data as B is plotted as a bar graph showing the fraction of each trend per unit occupancy. Comparing the left and right panels, the intermediate occupancies for stable sites (black) are depleted at a higher fraction relative to the other trends (purple and green) if the confidence interval thresholds are set too aggressively. We therefore set the threshold for this study at +/-25%.

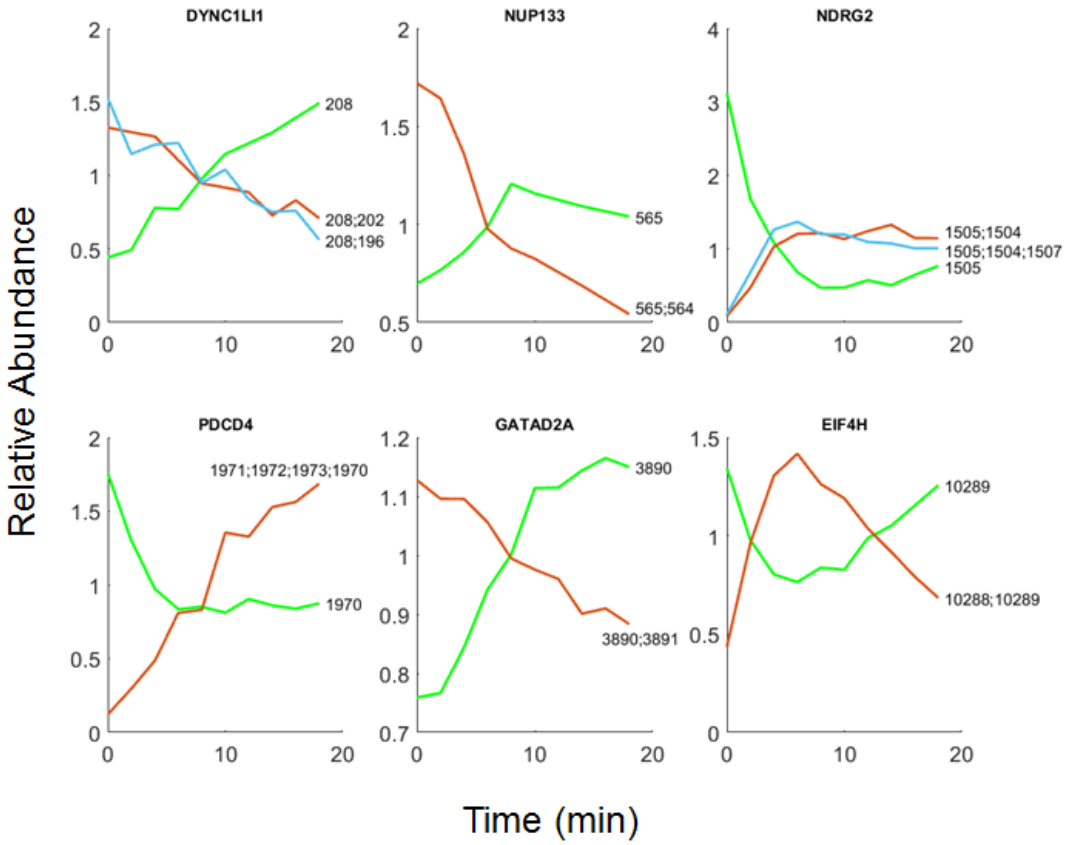

Figure S14: **Reciprocal trends occur between single and multi-phosphorylated forms of peptides.** Time series of proteins with multi-site phosphorylation that show reciprocal trends, labeled with the position of the modified residues. The single form is shown in green, while the first multi-site species (i.e., a double site) are shown in orange, and then the second in blue (e.g., triple sites) if present. These multi-sites (also called "composite sites" in the data) will always contain the single site shown. Just as with the reciprocal trends between the unmodified P^0 and P^1 forms, an otherwise non-dynamic P^1 form can appear to rise in abundance as the doubly phosphorylated P^2 form is dephosphorylated (or vice versa) due to conservation. If the P^0 form is detected, it is possible to calculate the site stoichiometry of each form (Fig. 4C, Fig. 5B), which is helpful for interpreting the trends. However, when interpreting relative trends, it is important to recognize that the true direction may be difficult to interpret for single sites that have multiple other sites on the same peptide. We detect this phenomenon for ~10% of the dynamic phospho-sites, which we correct for during clustering by discarding the single form and keeping the most-modified form (i.e., if the single form is phosphorylated the multi-site should also rise, whereas if the double site is dephosphorylated, the non-dynamic single site will increase due to conservation).

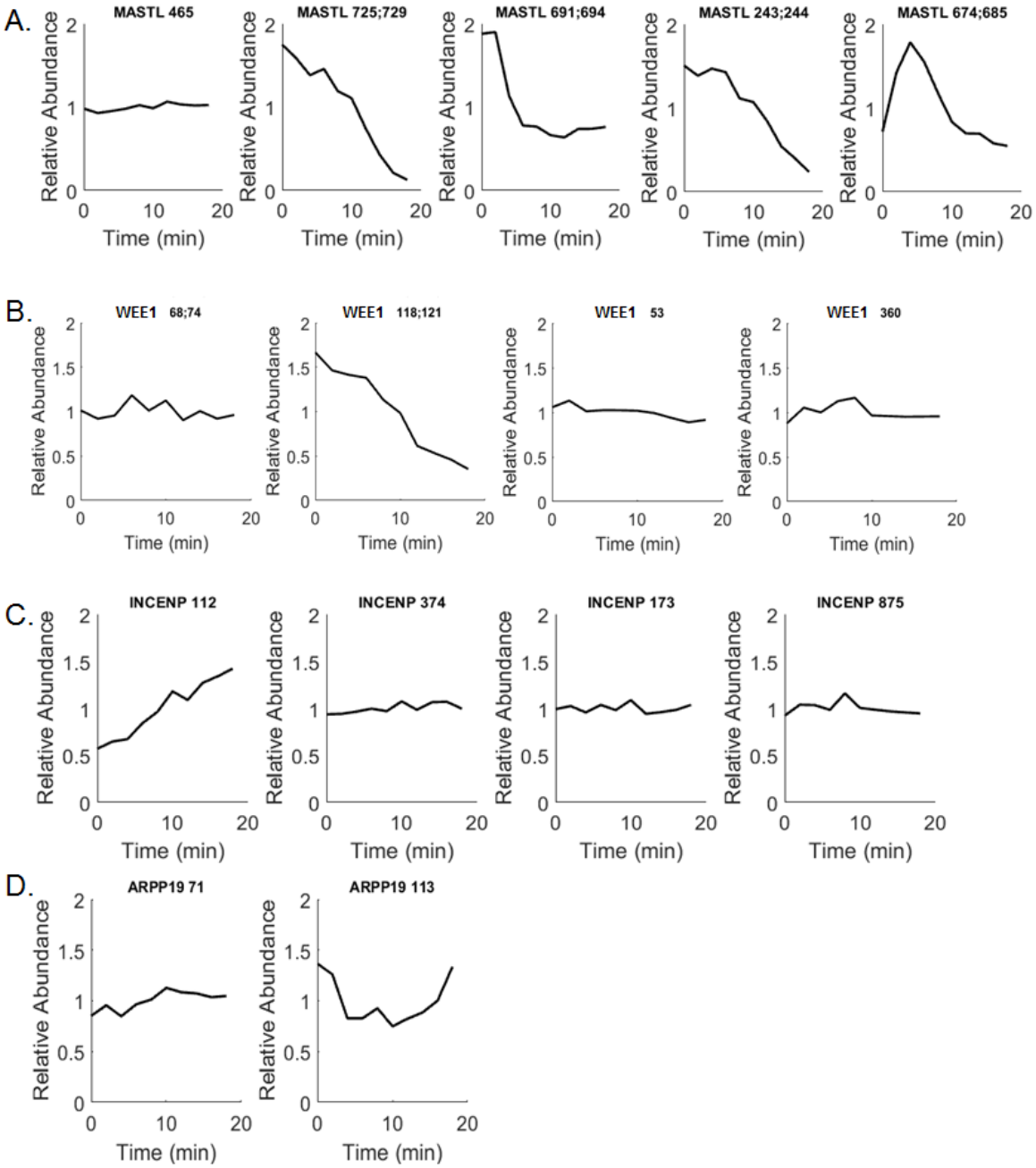

Figure S15: **Phospho-site dynamics on selected cell cycle proteins.** Time series of phospho-site relative abundance on known cell cycle regulators or targets. Site number indicated in title. A) Greatwall Kinase, B) WEE1 Kinase, C) INCENP, D) ENSA (PP2A inhibitor).

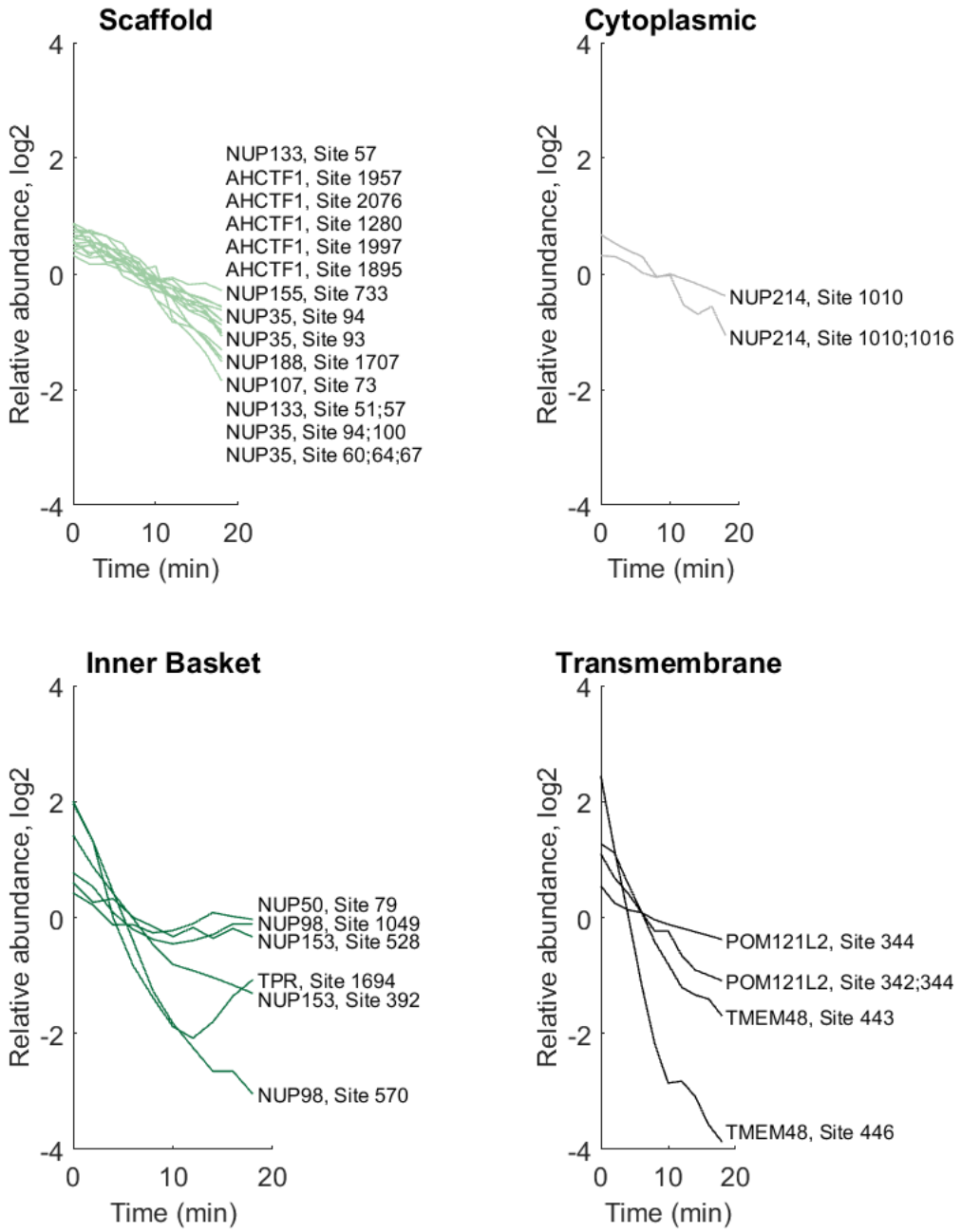

Figure S16: **Nucleoporins show differential dephosphorylation corresponding to nuclear pore sub-complexes.** Time series of nucleoporin phospho-site dynamics plotted by their respective sub-complexes. Many of the proteins have multiple phospho-sites, indicated by multiple entries in the legend. The phospho-occupancy plots for the subset of these sites that are available are shown in Fig. 6E. The slower dephosphorylation (~10 min half time) of the core scaffold (e.g., NUP188) and cytoplasmic components is consistent with their later recruitment. The fast dephosphorylation (~2-5 min half times) of inner basket/nuclear ring components (e.g., NUP153) is consistent with their early roles in NPC assembly of binding chromatin and nuclear envelope association (17). An exception is that trans-membrane nucleoporin recruitment is a secondary step, yet nearly all the sites that we detect in this region show fast dephosphorylation. Another exception is the scaffold component NUP107, which is recruited early, but shows slow dephosphorylation.

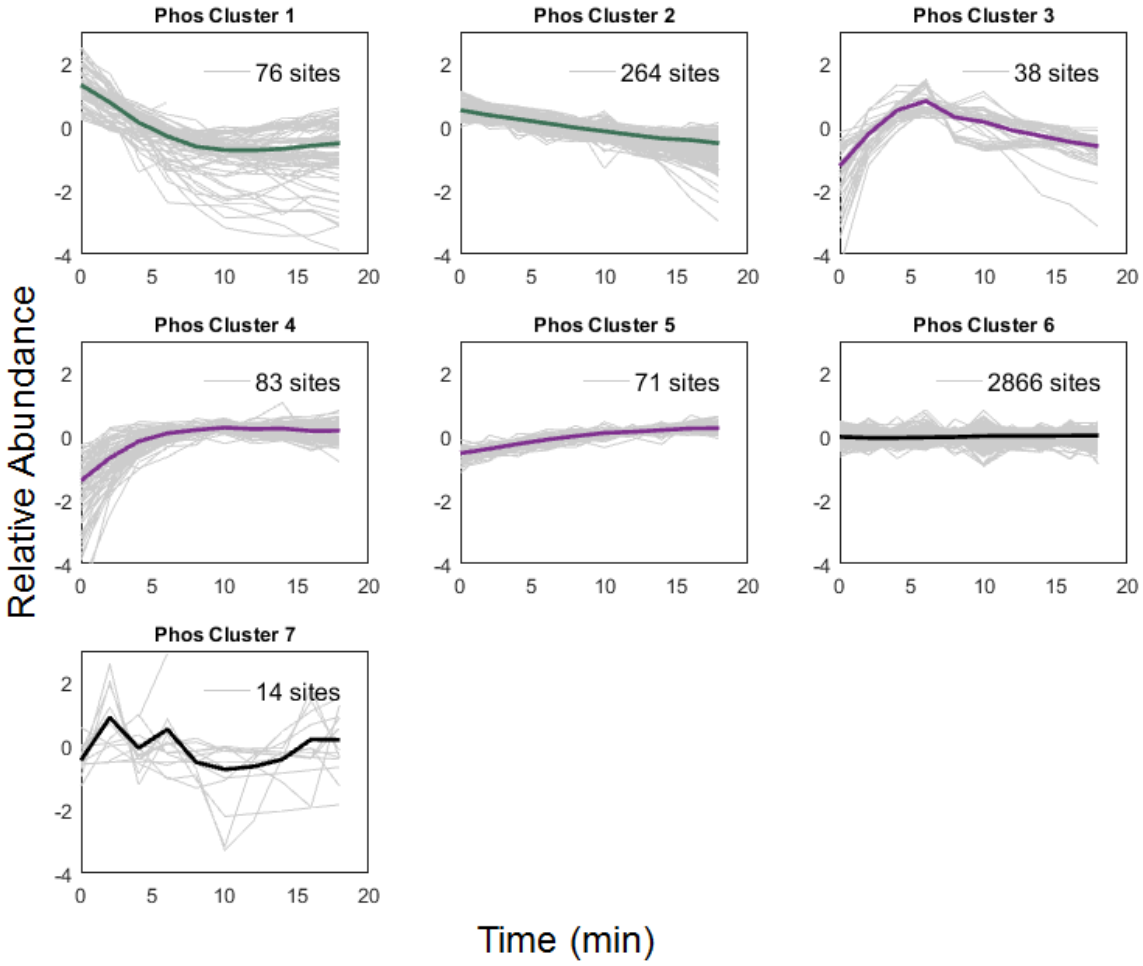

Figure S17: **K-means clustering of phosphorylation dynamics.** Time series of phospho-site cluster dynamics, including a 7th cluster of "noisy" trends excluded for display purposes only. These are included in all statistical analyses. Artifacts from single site dynamics are removed from clusters (as shown in Fig. S14). See Methods for further explanation of collapsing the clustering. Data provided in Table S8.

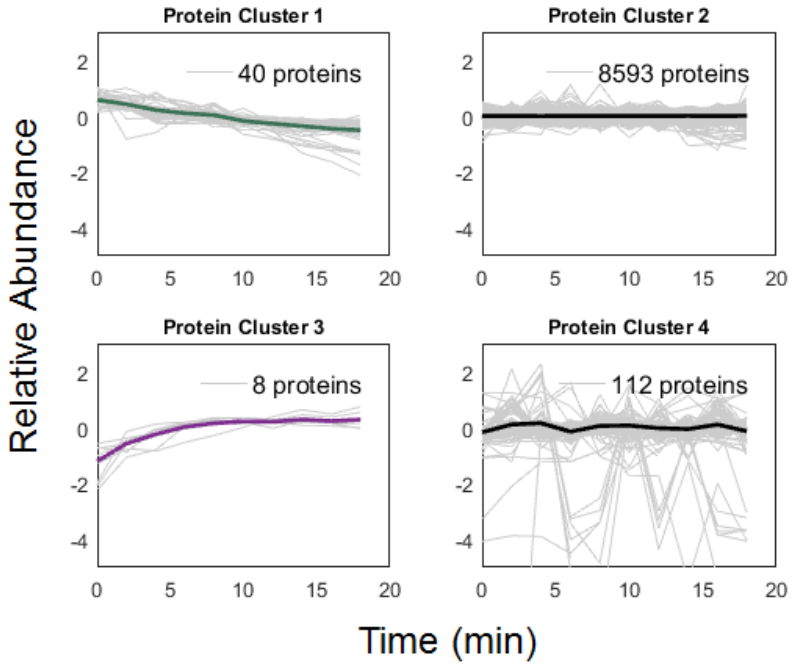

Figure S18: **K-means clustering of protein dynamics.** Time series of protein cluster dynamics, including a 4th cluster of "noisy" trends excluded for display purposes only. These are included to all statistical analyses (e.g., the FDR analysis). See Methods for further explanation of clustering. Data provided in Table S7.

Supplementary Methods

Principle of stoichiometry calculation

Here we demonstrate the calculation of phospho-site stoichiometry from relative trends for an arbitrary number of experimental conditions and phospho-forms. The following set of equations show the generalized geometric relationship between the measured relative ratios of phospho-form change and their unknown absolute values. Let the total number of species of a peptide be ***M.*** The species are the forms **P**, consisting of the phosphorylated residues ***j* = 0,1, 2, ... *M-1***, where *j* = 0 is the unmodified form. The ratios of change are the TMT-MS data from any experimental condition ***i*** normalized by the data at the reference condition, set here as ***i* = 1** (Eq. 1). Let the ratio of change of the total protein **T** for a given peptide be $\boldsymbol{c_{i,1}}$ (Eq. 2). This is either measured directly, or must be assumed as 1.

$$\text{Measured:} \qquad \frac{P_i^0}{P_1^0}, \frac{P_i^1}{P_1^1}, \ldots, \frac{P_i^j}{P_1^j} \qquad (1)$$

$$\text{Measured:} \qquad c_{i,1} = \frac{T_i}{T_1} \qquad (2)$$

Between any condition and the reference condition, the sum of all phosphorylated forms is either equal or related by the change in protein level $\boldsymbol{c_{i,1}}$ (Eq. 3).

$$\text{Conservation constraint:} \qquad \left(\sum_{j=0}^{M-1} P_1^j\right) c_{i,1} = \sum_{j=0}^{M-1} P_i^j \qquad (3)$$

Substituting Eq. (1) and Eq. (2) into Eq. (3) yields Eq. 4.

$$\text{Eq. (1), (2) in Eq. (3):} \qquad 0 = \sum_{j=0}^{M-1} P_1^j \left(\frac{P_i^j}{P_1^j} - c_{i,1}\right) \qquad (4)$$

Eq. (4) can be rearranged and written in vector form as Eq. (5).

$$\text{Eq. (4) in vector form:} \qquad 0 = [P_1^0, P_1^1, \ldots P_1^{M-1}] \cdot \left(\left[\frac{P_i^0}{P_1^0}, \frac{P_i^1}{P_1^1}, \ldots \frac{P_i^{M-1}}{P_1^{M-1}}\right] - \vec{1}c_{i,1}\right) \qquad (5)$$

Eq. (5) shows that the vector of absolute values $[P_1^0, P_1^1, \ldots P_1^{M-1}]$ is orthogonal to the *M*-1 dimensional subspace containing the measured relative values $\left(\left[\frac{P_i^0}{P_1^0}, \frac{P_i^1}{P_1^1}, \ldots \frac{P_i^{M-1}}{P_1^{M-1}}\right] - \vec{1}c_{i,1}\right)$, as the dot product is zero. When *M* = 2, this subspace is a 1-dimensional line containing the point (1, 1) scaled by $c_{i,1}$ (Fig. S8A). Since our aim is to solve for stoichiometry, it is sufficient that all possible absolute values lie on a line passing through the origin with a slope of $\frac{P_1^1}{P_1^0}$ (for intuition, see Fig. S8B). Stoichiometry is calculated by rewriting $\frac{P_1^1}{P_1^0+P_1^1}$ in terms of $\frac{P_1^1}{P_1^0}$, which is $(\frac{P_1^1}{P_1^0} / (1 + \frac{P_1^1}{P_1^0}))$. For *m* > 2, this subspace is an *M-1* dimensional plane, the orthogonal vector to this plane that is used to calculate the stoichiometry. When the number of conditions is greater number of conditions is greater than the number of species (*N* > *M*), the system is over-determined (Fig. S8C). In this case, the unique *M-1* dimensional subspace can be estimated by minimization (details discussed below).

Implementation of the regression-based stoichiometry calculation

We provide an algorithm in MATLAB to calculate the site stoichiometry of a single phospho-site when multiple conditions create an over-determined system (https://github.com/marcpresler/OccupancyMS). For the purposes of this study, we assumed that proteins were stable, as the data (Fig. 1C,D) show that protein dynamics are overwhelmingly flat during egg activation (i.e., $c_{i,1}$ = 1). The main functions are to 1) estimate the *M-1* dimensional subspace (Eq. 5) through minimization and 2) establish confidence intervals for that estimate. For simplicity, we initially discuss the case where *M = 2*. Input to the algorithm is quantitative mass spectrometry data for a P^0 and matching P^1 peptide for all measured conditions. The data for each species is normalized to the initial reference condition (see Eq. 1). The solution is the stoichiometry for that reference condition. The data can be visualized as a P^0 vs P^1 plot that has one point per condition (Fig. 4B). Because of the normalization, the reference condition will lie at the point (1, 1) (see Eq. 5, Fig. S8), scaled by $c_{i,1}$. When the number of conditions *N = M*, minimization is not necessary as the subspace is uniquely defined (in this case, a line). However, with more conditions, there is no longer a single solution (i.e., the system is over-determined). We can estimate the most likely subspace fitting a linear model to the data by minimizing the total least-squares residuals (Fig. 4B). We chose total least-squares (TLS) regression because we expect measurement errors in both the P^0 and P^1 ratios. To satisfy the orthogonal relationship defined in Eq. 5, the $\frac{P_1^1}{P_1^0}$ ratio is the negative inverse of the minimized slope of the subspace. Once obtained, we calculate stoichiometry (see end of previous section) and express in terms of percent occupancy of the modified residue (or the stoichiometry x 100). A steep fit slope will give a low occupancy because the orthogonal P^1/ P^0 ratio is therefore small, while a shallow fit slope gives a high occupancy, as the orthogonal P^1/P^0 ratio is large. To obtain the stoichiometry for the rest of the conditions, the reference condition is switched and the analysis is repeated until the all conditions have been solved (see Fig. S9C).

For *M* > 2, the *M-1* dimensional subspaces are estimated using Principle Component Analysis (PCA), which provides a method to minimize the total least squares residuals in an arbitrary number of dimensions through Singular Value Decomposition. The last principle component is the orthogonal vector to the *M-1* plane. The phospho-form ratios are obtained by projections onto the P^0 dimension or by finding the vector contained by the span of the orthogonal vector whose elements sum to 1 (satisfying Eq. 3). The TLS and PCA package (18) is adapted from File ID: #31109 on MathWorks File Exchange. Note that for *M* > 2, this approach assumes that we can detect all phospho-forms significant to the calculation. For example, a doubly phosphorylated species (*M=4*) will theoretically have two single forms. We typically detect only one of these forms and must calculate in *M =3*. This assumes the other form does not exist, or that it has a very small occupancy.

Calculating confidence intervals for the stoichiometry estimates

To aid in interpreting the quality of the estimated stoichiometry calculation, we sought to calculate a confidence interval of the minimized slope coefficient. Standard techniques for calculating confidence intervals rely on normally distributed error (19), which does not hold for total least squares minimizations. We therefore used bootstrapping (20) to estimate the confidence interval of the minimization. Specifically, the confidence interval is calculated from the distribution of slopes that is produced by iteratively fitting data subsampled with replacement. We used the MATLAB function 'bootci' and the bias corrected and accelerated percentile method to calculate confidence intervals. We performed 10,000 iterations per condition. For validation, we calculated an upper-bound estimate of the confidence interval by regressing against error in the Y-axis and then the X-axis separately, and taking the wider confidence interval of the two for each axis (15, 16). As expected, the bootstrapped confidence intervals were nearly always narrower than the upper-bound estimate (Fig. S10). Rarely, the subsampling with replacement will result in only one condition selected. In this case, fitting is not possible, so a random value between 0 and 100 is assigned. If a trend has repeated values in either the P^0 or P^1 forms (e.g., multiple zeros), a line with the slope of 0 or undefined can occur if only these points are subsampled during the bootstrapping, and fitting will fail. In these cases, the code returns either 0 or 100% occupancy automatically, depending whether the slope is undefined or zero, respectively. For the data where the initial best-fit slope results in an occupancy estimate that falls outside the 0 to 100% bounds, the confidence intervals are automatically set to 0 and 100% to reflect the high amount of error. This bootstrapping approach scales to *n*-dimensional space for multi-phosphorylated species.

Additional considerations

Corresponding P^0 and all P^j forms must have a matching Protein ID (i.e., same gene symbol and isoform, if relevant) and an identical sequence to the peptide used to quantify the phospho-site occupancy. A given phospho-site may contain several peptides (e.g., missed cleaved species). Any of these peptides are used as a match for the P^0 species and are summed. The user must make sure that the each condition is appropriately normalized before import. Because of reference condition normalizations (Eq. 1), any value of 0 in the data is replaced with 1E-9 as to not result in undefined ratios. All replicates are averaged into one trend.

Pseudo-code for computing occupancy with confidence intervals from multiplexed-MS data

Load data

Load the .xlxs or other file containing TMT signal for phospho sites and matched unmodified peptides

For each site in the data set:

Replace any zeros with 1e-9

For each condition:

%% Calculate occupancy value %%
Set the current condition as the reference
Normalize phospho site and unmodified data to the reference
Pass ratios to total least squares regression function and fit slope
Calculate occupancy from negative inverse of slope

End

%% Calculate confidence interval for estimated occupancy value %%
If, for all conditions, the best-fit occupancy is between 0 and 100
For each condition
Check for repeated values within phospho site and unmodified peptide data and flag
For each bootstrap iteration
Randomly choose conditions with replacement for fitting
Check that the same condition has not been chosen every time. If it has, choose random integer between 0 and 100 and move to next iteration
Create new input values using the conditions chosen above
(If flag is present, check to make sure the new input does not define a line with infinite or zero slope and if so set occupancy to 0 or 100, respectively and move to next iteration)
Pass new input ratios to total least squares regression function and fit slope
Calculate occupancy from negative inverse of slope
End
Use the 'bias corrected and accelerated percentile method' determine the upper and lower 95% confidence intervals
End
Else
Report upper and lower confidence interval bounds as 0 and 100, respectively, for all conditions

End

End

False discovery rate (FDR) calculation

The approach to calculate the FDR is visualized in Fig. S2. In brief, the experimental distances are measured to a 10-dimensional, idealized degradation vector and then randomized iteratively. In practice, the data were measured in up to four replicates with ten conditions each. To account for this, all protein trends were represented as 40-dimensional vectors. The idealized vector was repeated four times to match this. Proteins were ranked according to their distance from the repeated idealized vector by Cosine distance. While non-linear degradation trends exist in the data, they are much closer to the idealized vector than the flat trends. Thus, they are still ranked highly, with a small underestimation in distance. If a protein was missing from a given replicate, the mean trend of that replicate (which is essentially flat) was substituted for the missing data in the 40-dimensional experimental vector. The data were then randomized 10,000 times with the cosine distance measured each iteration. The cumulative distributions of distances for all proteins (experimental) and pooled randomized trends were used to calculate the FDR at a given cosine distance as the idealized vector, or (Randomized/(Experimental + Randomized)) *100. Since the randomized values are pooled we apply one cutoff to the entire dataset, so multiple hypothesis correction is not required. The FDR is calculated for each Protein ID (Table S1), and then combined by gene symbol afterward (Table S2).

Phospho-peptide enrichment

We chose to multiplex peptides and enrich one on column to improve the data quality by eliminating variability that comes from multiple parallel enrichment columns. There is a tradeoff of decreased yield and therefore depth, as it is not economical to label more than a few milligrams of material. We used 2.5-4 mgs of TMT-labeled peptides per replicate. These were desalted and dried down into a pellet, which was resuspended in Binding Buffer (85% 11.1M Lactic Acid, 50% acetonitrile, pH ~1.5-2.0) at 10 mg/mL and spun for 30 minutes at 21K RCF at RT. 5 µm Titanium Dioxide microspheres (GL Sciences 5020-75000) were added at 8:1 peptide to bead ratio (by weight) and washed 3X with 1 mL of Binding Buffer. Peptides were added to dried Titanium Dioxide beads and incubated for 1 hr at RT with constant agitation. PTFE membrane filters (0.2 µm low-binding hydrophilic, Millipore UFC30LG25) were washed three times with Binding Buffer, and then used to filter the TiO_2 microspheres, spinning at no more than 10K rcf until dryness. The first flow-through was used for a second serial enrichment as described below (A and B enrichments). The final flow-through was saved and used as a protein replicate. The A and B enrichments were each washed twice with two column-volumes of 50% acetonitrile, 0.1% Trifluoroacetic Acid and then twice with 25% acetonitrile, 0.1% Trifluoroacetic Acid. Phosphopeptides were eluted from the beads 3 times with 1 column volume of Elution Buffer (200mM of KH_2PO_4, pH 10) each. Beads were incubated in Elution Buffer for 5 minutes with occasional agitation before spinning the beads to dryness; the 3 eluates were combined. The A and B enrichments remained separated. All samples were acidified with 10% formic acid and 0.5% TFA and each desalted on C18 SepPak. We found if more than ~15 mg of beads were used in one filter, the specificity of the enrichment suffered as the column clogs and causes the washes and elution to be less efficient. Therefore, for the second two experiments, we split the sample onto 10 columns processed in parallel for the A enrichment. Since the peptides are pre-labeled and already multiplexed, the measured peptide ratios are unaffected by splitting the samples. For the B enrichment, the washing steps were performed in a single Eppendorf tube and the beads were washed by pelleting (a practical step to reduce the use of filter tubes). The A and B enrichments and the flow-through were desalted again. A typical yield of 50-80 µg of peptides eluted from the column for the A and B enrichments combined, with a median specificity of ~80% between biological replicates. The flow-through sample was fractionated with HPLC as in the main text. For the enriched samples, peptides were subjected to an offline Strong Cation Exchange fractionation adapted from Dephoure and Gygi (21). Peptides were re-supsended in Buffer A (7mM KH2PO4, pH 2.65 and 30% ACN) and loaded onto 20um, 30 angstrom PolySulfoethyl A beads (Nest Group, BMSE2003). Salt cuts were made by mixing appropriate ratios of Buffer A with Buffer B (7mM KH2PO4, pH 2.65, 30% ACN, and 350 mM KCl) for Enrichment A: 0, 5, 10, 20, 40, 60, and 350mM, for Enrichment B: 0, 5, 10, 40, 350. Fractions were desalted on STAGE tips and analyzed by mass spectrometry.

EM imaging time series

Eggs were prepared and electro-activated as described in the main text. Time points collected were 0 sec, 8 sec, 15 sec, then at 15 second intervals until 105 seconds, then 2 min, 4 min, and 10 min. 8 sec time

point was the earliest condition we were able to collect. 3-5 eggs were collected in 200 µL of volume with a cut pipette tip. The eggs and buffer were added to 2X fixative solution (see below) and quickly mixed to final 1X concentration. We are confident the fixation occurred quickly at the cortex, where we focused our imaging. Eggs were fixed with 3% gluteraldehyde in 0.1X MMR for 30 minutes. The fixed eggs were cut and the animal and vegetal poles separated and placed into fix in separate tubes to maintain orientation in later steps. In most cases, halving the eggs led to the loss of the vitelline membrane, and the extracellular morphology is therefore not present in the images. Eggs were rinsed 2 times in 0.1X MMR and 3 times in 0.05M cacoydylate buffer pH 7.0. They were then post-fixed with 1% osmium in 0.8% $K_3Fe(CN)_6$ in cacodylate buffer on ice in a fume hood for 1 hour (22). They were then washed 3 times with cacodylate buffer and 2 times in distilled water prior to staining overnight in 1% aqueous uranyl acetate in the dark. Samples were washed in distilled water prior to dehydration in an ethanol series using the progressive lowering of temperature method followed by exchanges of 100% ethanol and propylene oxide at room temperature (23). Infiltration was done in 2:1, 1:2 mixtures of propylene oxide epon araldite for 1 hour each step followed by 100% epon araldite for 2 hours. Samples were embedded and polymerized at 65°C for 48 hours. The animal poles of the eggs were thin-sectioned on a Reichert Ultracut S microtome and the sections viewed on a Technai G2 Spirit Bio TWIN Transmission Electron Microscope and imaged with an AMT 2k CCD camera.

Data normalization

Every protein or phospho-site was required to have a summed TMT Signal to Noise value across the ten TMT channels of greater than or equal to 386 on the Thermo Fusion or Lumos, or 189 for the Orbitrap Elite. Each condition was then normalized to the condition median to correct for pipetting errors. Each trend was then normalized to its mean. Replicates trends were averaged. For proteins and phospho-sites that appear in the phosphatase-treated experiment, the 12 min and 16 min time points are missing. In this case, these time points are replaced by averaging the two flanking conditions. For the supernatant experiments, median normalization per time point is not appropriate, as the time series are not expected to be flat. We therefore estimated an accumulation curve and normalized each trend to this curve. Similarly, median normalization is not appropriate for the phosphatase treated samples on the phospho-level as the signal is expected to be lower than the rest for these conditions. We normalize by the mean of the averages of all untreated conditions. This correction is unnecessary on the protein level for the phosphatase-treated conditions, as the majority of proteins are not affected by the treatment.

Establishment of absolute changes of protein and phosphorylation level

Proteins were matched by gene symbol to our previous dataset of protein abundance in the *Xenopus* egg. For the protein loss, the abundance of candidates that passed the 1% FDR threshold were multiplied by the fold change from the first condition, which is the unactivated egg that is equivalent to the sample used in the previous study. Securin was not seen in our previous absolute abundance data set. We estimated its abundance from its stoichiometric binding partner, Separase. The rates of degradation for the known cell cycle targets were calculated with a linear fit using Matlab “polyfit” function. Confidence intervals were calculated using File ID: 39126 on The MathWorks File Exchange. For obtaining absolute values of phospho site, the same procedure was followed for proteins for which we could estimate phospho-stoichiometry. In this case, the abundance was multiplied by the percent occupancy at each time point.

Estimating maximal degradation rates of the APC/Proteasome

The overall rate of ubiquitination-dependent degradation is dependent on the rates of ubiquitination, deubiquitination, and proteasome reaction. It is difficult to estimate the K_{cat} for the APC or deubiquitination activity. However, we can approximate the rate of the proteasome reaction. We estimate $1/ K_{degradation}$ as ~80 seconds, given that $1/ K_{degradation} = N \times (1/ ([Proteasome] \times K_{on}) + 1/K_{cat}$. N, number of times a substrate must bind before degradation occurs, is ~6; K_{on} was measured at ~10^5 M^{-1} s^{-1}; Kcat is ~20 seconds per substrate bound (24). The proteasome concentration is ~1µM in the frog egg (14). For $[Substrate_{Ubiq}]$, though we cannot infer from rates 1 and 2, we can estimate that 1% or at most 10% of the substrate concentration is modified at steady state (25). The total concentration of known APC substrates is ~400nM, therefore $[Substrate_{Ubiq}]$ = 4-40 nM. The rate of degradation is ~3-30nM /min (or $[Substrate_{Ubiq}] / K_{degradation}$). Our estimated rate of degradation

rate from the data is ~14 nM/min, or 150 pg/min (Fig. 2A), and the original concentration estimates contain ~2 fold error (14). Thus, with these parameters, an estimate of the upper bound of the degradation rate overlaps with the range measured from the data.

K-means clustering and multi-site artifact correction

K-means clustering was implemented in MATLAB using the 'kmeans' function with K = 56, which was chosen because we found it gave stable clusters. The analysis was performed on the mean normalized, replicate-averaged data for both protein and phosphorylation data. The 56 clusters were manually collapsed into 7 clusters for phosphorylation and 4 clusters for the protein that qualitatively represented the data. The last cluster was reserved for sites that were consistently clustered individually, unstably, or into clusters with less than 10 protein/sites. For the purposes of display, these are not included in Fig. 1 but are provided in Fig. S17 and Fig. S18 and were included in any statistical analyses. These cases constitute <1% of the data. The modified Waterfall plots in Fig. 1 are normalized to the first time point and ordered by their cluster and then then ascending within each cluster by the value of the last time point. For the phosphorylation trends, we corrected for possibly misleading trends of single sites with corresponding composite sites. A double site and corresponding single site can often show the same reciprocal behavior as shown between unmodified and singly modified species (see Fig. S14). Therefore, a trend that appears as phosphorylation is really dephosphorylation of the double form. To avoid being misled by this trend, we systematically identified single sites with reciprocal trends to their corresponding composite sites (as evaluated by a Spearman correlation of <-0.5). These sites were removed from the analysis, and the most-modified composite site was kept, as this is the form most likely to give the true direction of the sites. This correction totaled ~50 dynamic sites. This phenomenon has been previously discussed in the literature (26).

Motif enrichment analysis

Using localized sites with possible artifacts from reciprocal trends removed, we evaluated the enrichment of various motifs -S#/T#-P- or -RxxS- using regular expression searches in each cluster discussed above and in Fig. 1E. A p-value was calculated with the Fisher's exact test (27).

Gene set enrichment analysis

Gene Symbols classified by their artifact corrected K-means clusters were used for gene set enrichment analysis performed with WebGestalt (28). Enrichment was assessed with the hypergeometric test, and multiple-hypothesis testing correction was performed with the Benjamini-Hochberg procedure (29).

Supplementary Information Appendix References

1. Zhao WM & Fang G (2005) Anillin is a substrate of anaphase-promoting complex/cyclosome (APC/C) that controls spatial contractility of myosin during late cytokinesis. *J Biol Chem* 280(39):33516-33524.
2. Petersen BO*, et al.* (2000) Cell cycle- and cell growth-regulated proteolysis of mammalian CDC6 is dependent on APC-CDH1. *Genes & development* 14(18):2330-2343.
3. Ayad NG*, et al.* (2003) Tome-1, a trigger of mitotic entry, is degraded during G1 via the APC. *Cell* 113(1):101-113.
4. Seki A & Fang G (2007) CKAP2 is a spindle-associated protein degraded by APC/C-Cdh1 during mitotic exit. *J Biol Chem* 282(20):15103-15113.
5. Liot C*, et al.* (2011) APC(cdh1) mediates degradation of the oncogenic Rho-GEF Ect2 after mitosis. *PLoS One* 6(8):e23676.
6. Singh SA*, et al.* (2014) Co-regulation proteomics reveals substrates and mechanisms of APC/C-dependent degradation. *The EMBO journal* 33(4):385-399.
7. Li L*, et al.* (2007) NuSAP is degraded by APC/C-Cdh1 and its overexpression results in mitotic arrest dependent of its microtubules' affinity. *Cell Signal* 19(10):2046-2055.
8. Araki M, Wharton RP, Tang Z, Yu H, & Asano M (2003) Degradation of origin recognition complex large subunit by the anaphase-promoting complex in Drosophila. *The EMBO journal* 22(22):6115-6126.
9. Juang YL*, et al.* (1997) APC-mediated proteolysis of Ase1 and the morphogenesis of the mitotic spindle. *Science* 275(5304):1311-1314.

10. Eshleman HD & Morgan DO (2014) Sgo1 recruits PP2A to chromosomes to ensure sister chromatid bi-orientation during mitosis. *Journal of cell science* 127(Pt 22):4974-4983.
11. Stewart S & Fang G (2005) Anaphase-promoting complex/cyclosome controls the stability of TPX2 during mitotic exit. *Molecular and cellular biology* 25(23):10516-10527.
12. Rape M & Kirschner MW (2004) Autonomous regulation of the anaphase-promoting complex couples mitosis to S-phase entry. *Nature* 432(7017):588-595.
13. Feine O, Zur A, Mahbubani H, & Brandeis M (2007) Human Kid is degraded by the APC/C(Cdh1) but not by the APC/C(Cdc20). *Cell cycle* 6(20):2516-2523.
14. Wuhr M*, et al.* (2014) Deep Proteomics of the Xenopus laevis Egg using an mRNA-Derived Reference Database. *Current biology : CB* 24(13):1467-1475.
15. Willassen Y (1984) Testing hypotheses on the unindetifiable structural parameters in the classical 'errors-in-variables- model with application to Friedman's permanent income model *Economics Letters* 14:221-228.
16. Moran PAP (1971) Estimating structural and functional relationships. *J. Multivariate Anal.,* 1:232-255.
17. Meszaros N*, et al.* (2015) Nuclear pore basket proteins are tethered to the nuclear envelope and can regulate membrane curvature. *Developmental cell* 33(3):285-298.
18. Bednarova IPaD (2010) Total Least Squares Approach to Modeling: A Matlab Toolbox. *Acta Montanistica Slovaca* 15(2):pp. 158-170.
19. Ness C-LCaJV (1997) Structural and Functional Models Revisited *Recent Advances in Total Least Squares Techniques and Errors-In-Variables Modeling* ed Huffel SV (Society for Industrial and Applied Mathematics), p 37.
20. Shao J, Tu D (1995) *The Jackknife and Bootstrap* (Springer, New York) 1st Ed.
21. Dephoure N & Gygi SP (2011) A solid phase extraction-based platform for rapid phosphoproteomic analysis. *Methods* 54(4):379-386.
22. McDonald K (1984) Osmium ferricyanide fixation improves microfilament preservation and membrane visualization in a variety of animal cell types. *J Ultrastruct Res* 86(2):107-118.
23. Carlemalm E, Villiger W, Hobot JA, Acetarin JD, & Kellenberger E (1985) Low temperature embedding with Lowicryl resins: two new formulations and some applications. *J Microsc* 140(Pt 1):55-63.
24. Lu Y, Lee BH, King RW, Finley D, & Kirschner MW (2015) Substrate degradation by the proteasome: a single-molecule kinetic analysis. *Science* 348(6231):1250834.
25. Dimova NV*, et al.* (2012) APC/C-mediated multiple monoubiquitylation provides an alternative degradation signal for cyclin B1. *Nature cell biology* 14(2):168-176.
26. Solari FA, Dell'Aica M, Sickmann A, & Zahedi RP (2015) Why phosphoproteomics is still a challenge. *Mol Biosyst* 11(6):1487-1493.
27. Fisher R (1922) On the interpretation of X^2 from contingency tables, and the calculation of P. *Journal of the Royal Statistical Society* 85(1):87-94.
28. Wang J, Duncan D, Shi Z, & Zhang B (2013) WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. *Nucleic acids research* 41(Web Server issue):W77-83.
29. Benjamini YH, Yosef (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 57(1):289-300.