- Supplementary Information –

A CDK-mediated phosphorylation switch of disordered protein condensation

Authors: Juan Manuel Valverde[1,2†], Geronimo Dubra[3,4†], Henk van den Toorn[1,2], Guido van Mierlo[5], Michiel Vermeulen[5], Albert J.R. Heck[1,2], Carlos Elena-Real[6], Aurélie Fournet[6], Emile Al Ghoul[7], Dhanvantri Chahar[3,4], Austin Haider[8], Matteo Paloni[6], Angelos Constantinou[7], Alessandro Barducci[6], Kingshuk Ghosh[8], Nathalie Sibille[6], Pau Bernado[6], Puck Knipscheer[9], Liliana Krasinska[3,4‡], Daniel Fisher[3,4‡*], Maarten Altelaar[1,2‡*]

Affiliations:

[1]Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Utrecht, 3584 CH Utrecht, Netherlands.

[2]Netherlands Proteomics Center, Padualaan 8, 3584 CH Utrecht, Netherlands.

[3]IGMM, University of Montpellier, CNRS, Inserm, Montpellier, France.

[4]Equipe Labellisée LIGUE 2018, Ligue Nationale Contre le Cancer, Paris, France.

[5]Department of Molecular Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, Oncode Institute, Radboud University Nijmegen, 6525 GA Nijmegen, the Netherlands.

[6]CBS, University of Montpellier, INSERM, CNRS, Montpellier, France.

[7]IGH, University of Montpellier, CNRS, Montpellier, France.

[8] Department of Physics and Astronomy, and Department of Molecular and Cellular Biophysics, University of Denver, Denver, Colorado 80208, USA.

[9]Oncode Institute, Hubrecht Institute–KNAW and University Medical Center, Utrecht, 3584 CT, Netherlands.

*Correspondence to: m.altelaar@uu.nl and daniel.fisher@igmm.cnrs.fr

†‡ Equal contributions

Supplementary Materials:

Extended data figures S1-S10.

Supplementary movie S1. Representative trajectory from PT-MD simulation of full chain Ki-67 at relative temperature 1.1.

Supplementary movie S2. Phase coexistence MD simulations of non-phosphorylated (A) and phosphorylated (B) monomers of the Ki-67 consensus repeat at a relative temperature of 0.98. This temperature is below the critical temperature (T_c) of the non-phosphorylated consensus repeat but above the T_c of the phosphorylated one. In fact, in the former case, two phases at equilibrium are observed (A), while in the latter, a single homogeneous phase is observed (B).

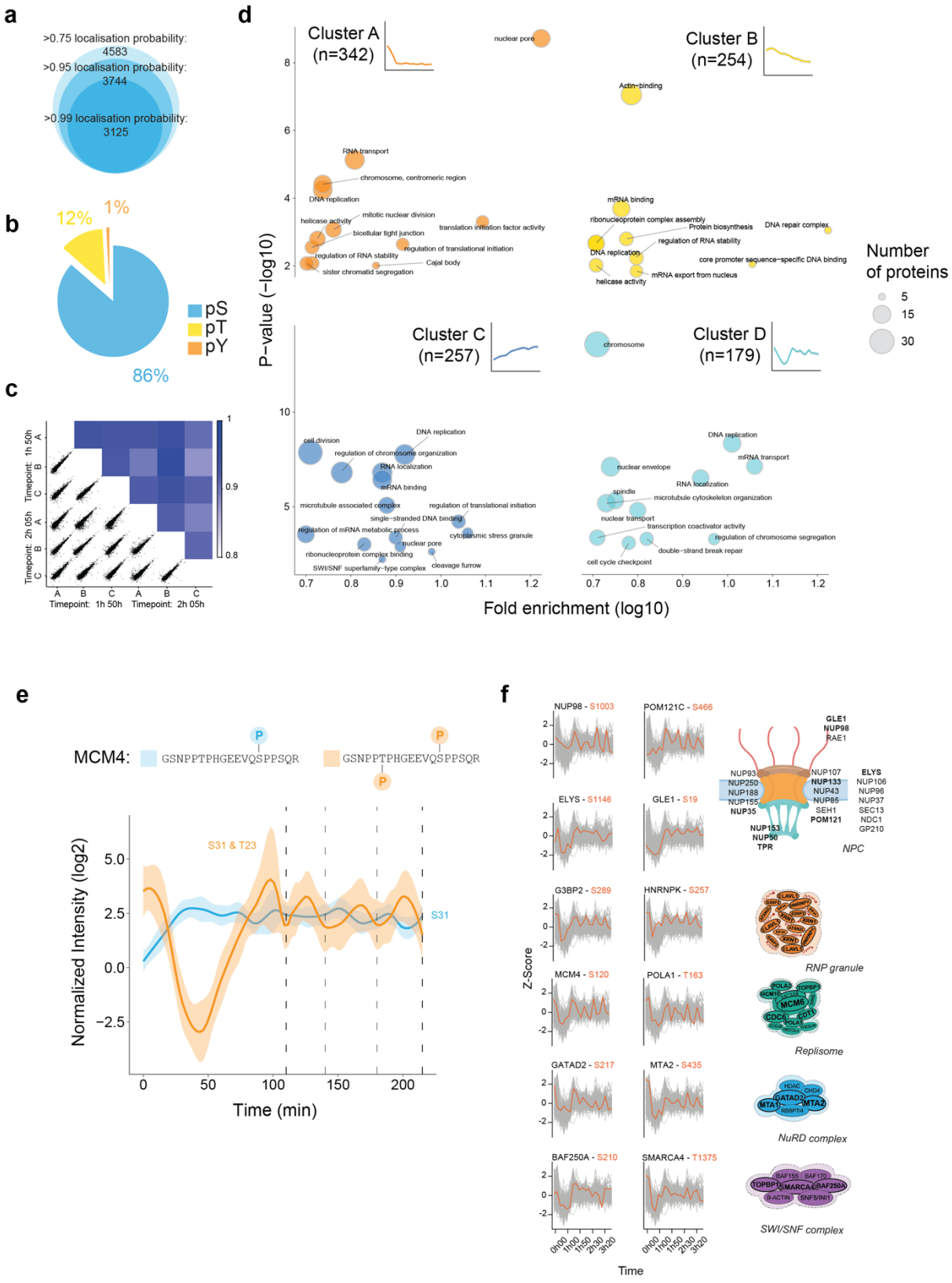
Data S1. Phosphoproteomics-data

Data S2. CDK-phosphorylation-data

Data S3. Protein-disorder-prediction

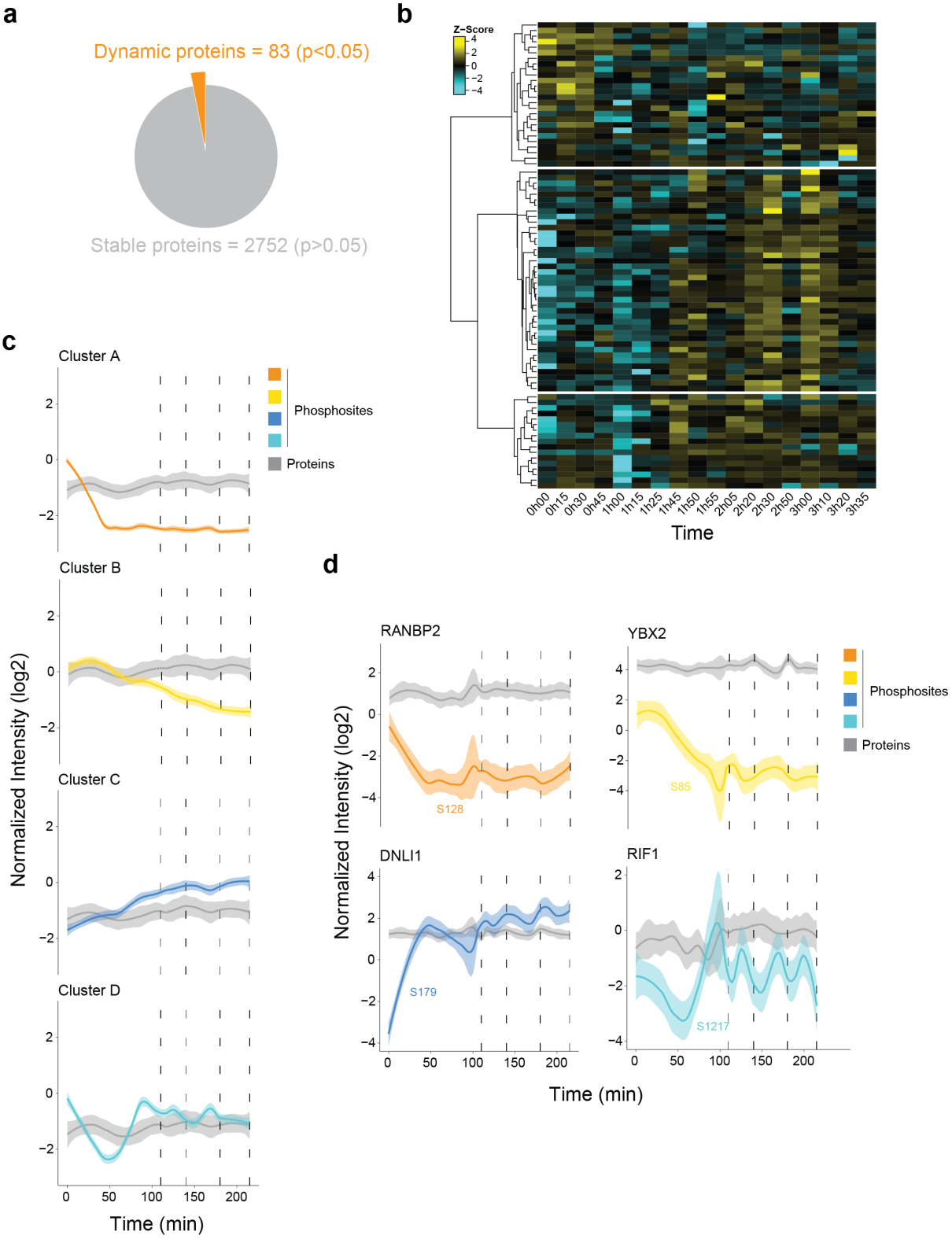
Data S4. Human-proteins-in-MLOs

Materials and Methods



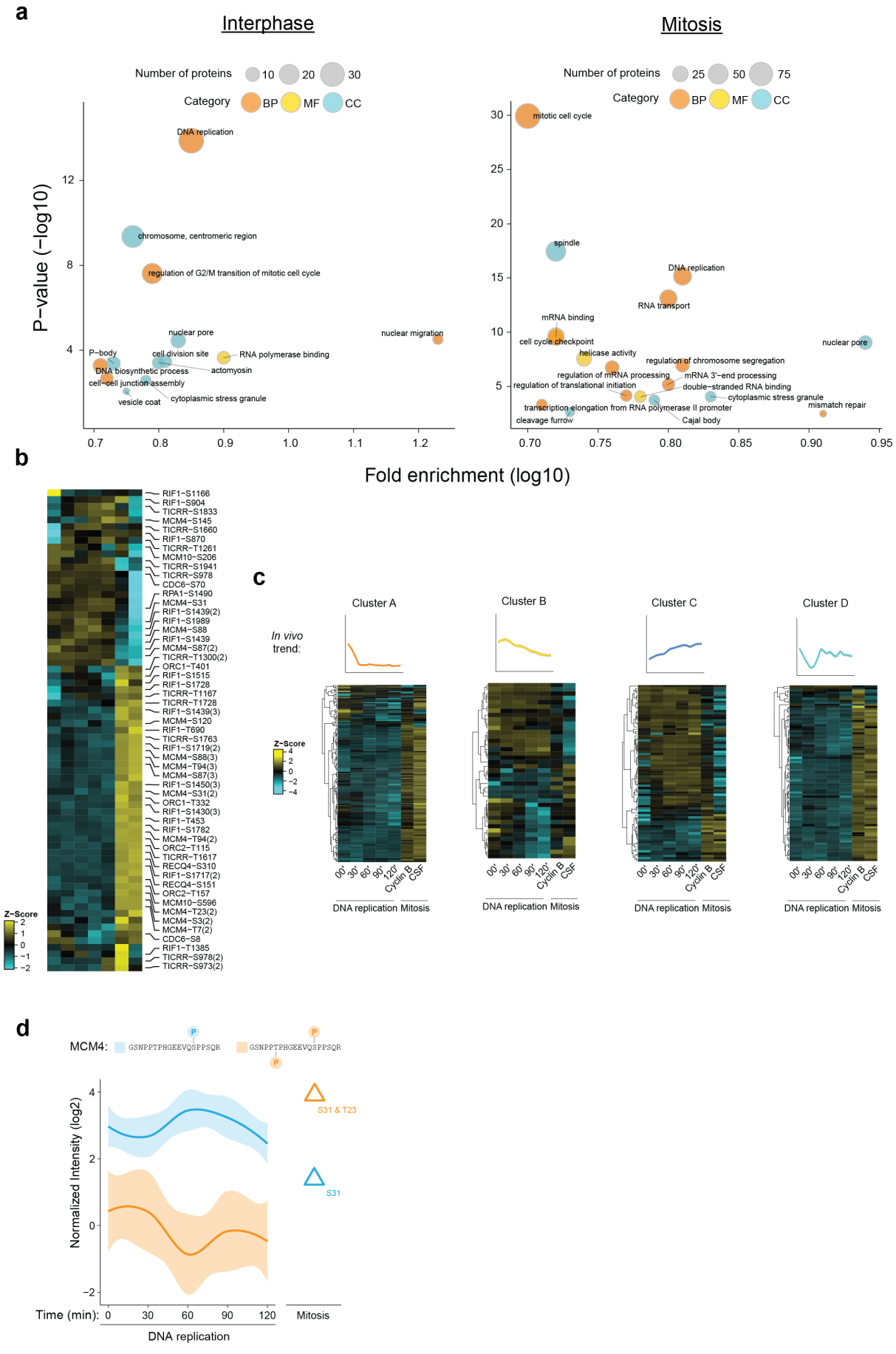
Extended data Fig. 1. Phosphosite dynamics correlates with cell cycle phases; related to Figure 1. (a) Total number of phosphosites detected and their distribution according to the site localisation probability score. (b) Distribution of phosphosites identified among serine, threonine and tyrosine residues. (c) Correlation coefficients for two randomly selected time

points. Scatter plots of significantly enriched (Fisher's exact test with Bonferroni correction, $p<0.05$) GO (BP, MF, CC, Uniprot keywords) terms for all dynamic phosphosites per cluster in the *in vivo* experiment, presenting the fold-enrichment of specific terms vs statistical significance. The size of the circles correlates with the number of proteins associated with the specific term. More details and the full list of enriched GO terms per cluster is found in Data S1. (d) *In vivo* reciprocal trends of singly- and multi-phosphorylated peptides carrying phosphorylated T23 and S31 of MCM4 (dashed lines depict the time points of cell division): orange curves, the trend of T23 and S31 in the multi-phosphorylated peptide; blue curve, the trend of S31 in the singly-phosphorylated peptide. (e) Examples of proteins with known association showing similar oscillating phosphorylation. Plots highlight the dynamic trend of the cluster (grey) and selected phosphosites (orange) over time. Right, illustrations of protein complexes formed by the proteins undergoing dynamic phosphorylation. Proteins highlighted in bold show at least one oscillating phosphosite in our dataset.

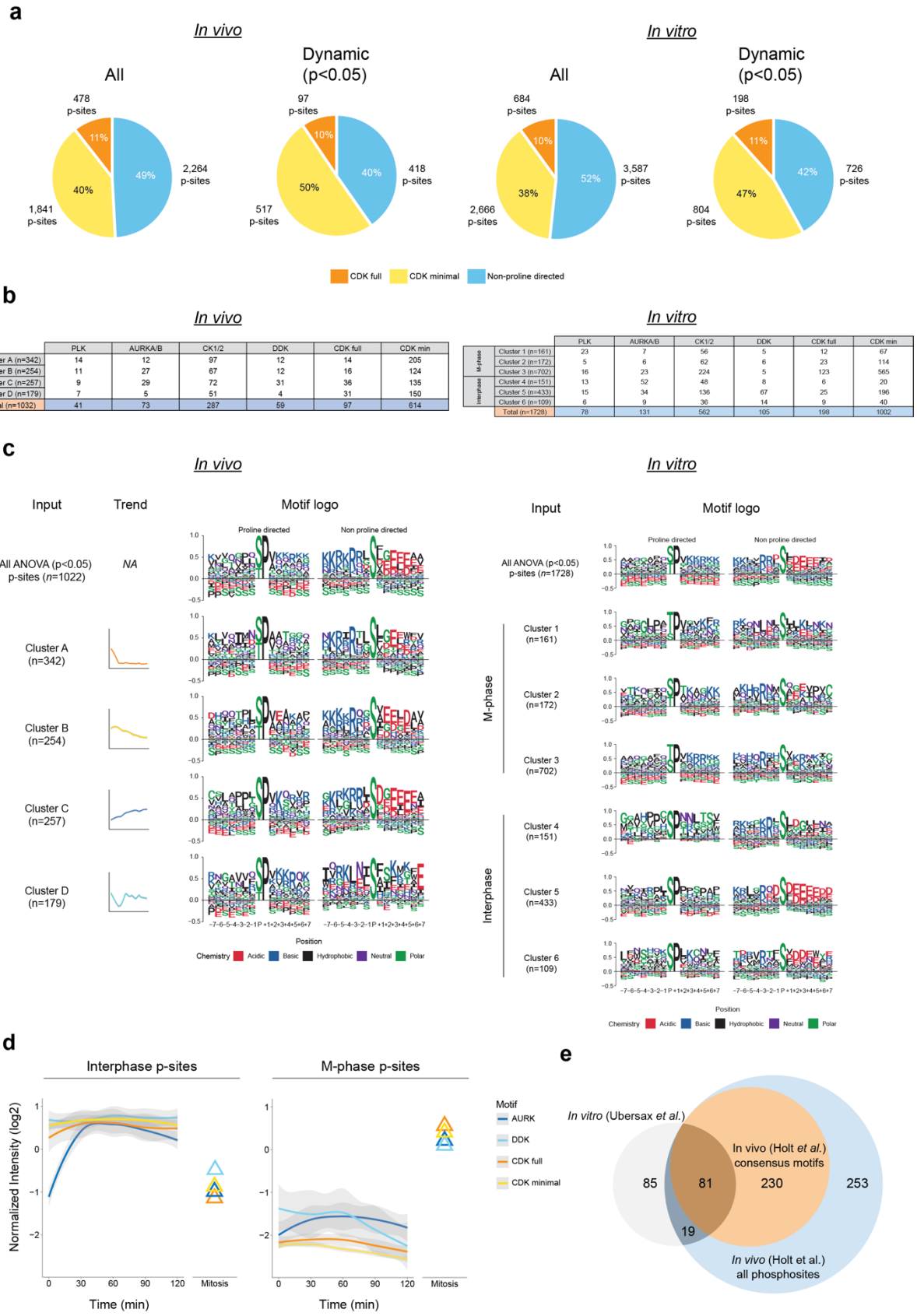


Extended data Fig. 2. Variation in the total proteome versus the phosphoproteome in early embryonic cell cycles; related to Figure 1. (a) Total proteome analysis reveals 83 proteins out of 2835 showing significant changes in abundance (ANOVA, Benjamini-Hochberg correction, FDR 0.05) over the time course. (b) Heat map showing abundance of the variable proteins over the time course. (c) Comparison of dynamic variations in total protein

compared to total phosphosites from the four clusters shown in Figure 1c (dashed lines depict the time points of cell division). (d) Examples of dynamics of individual phosphosites from the four clusters shown in Figure 1c and levels of the corresponding protein (dashed lines depict the time points of cell division).

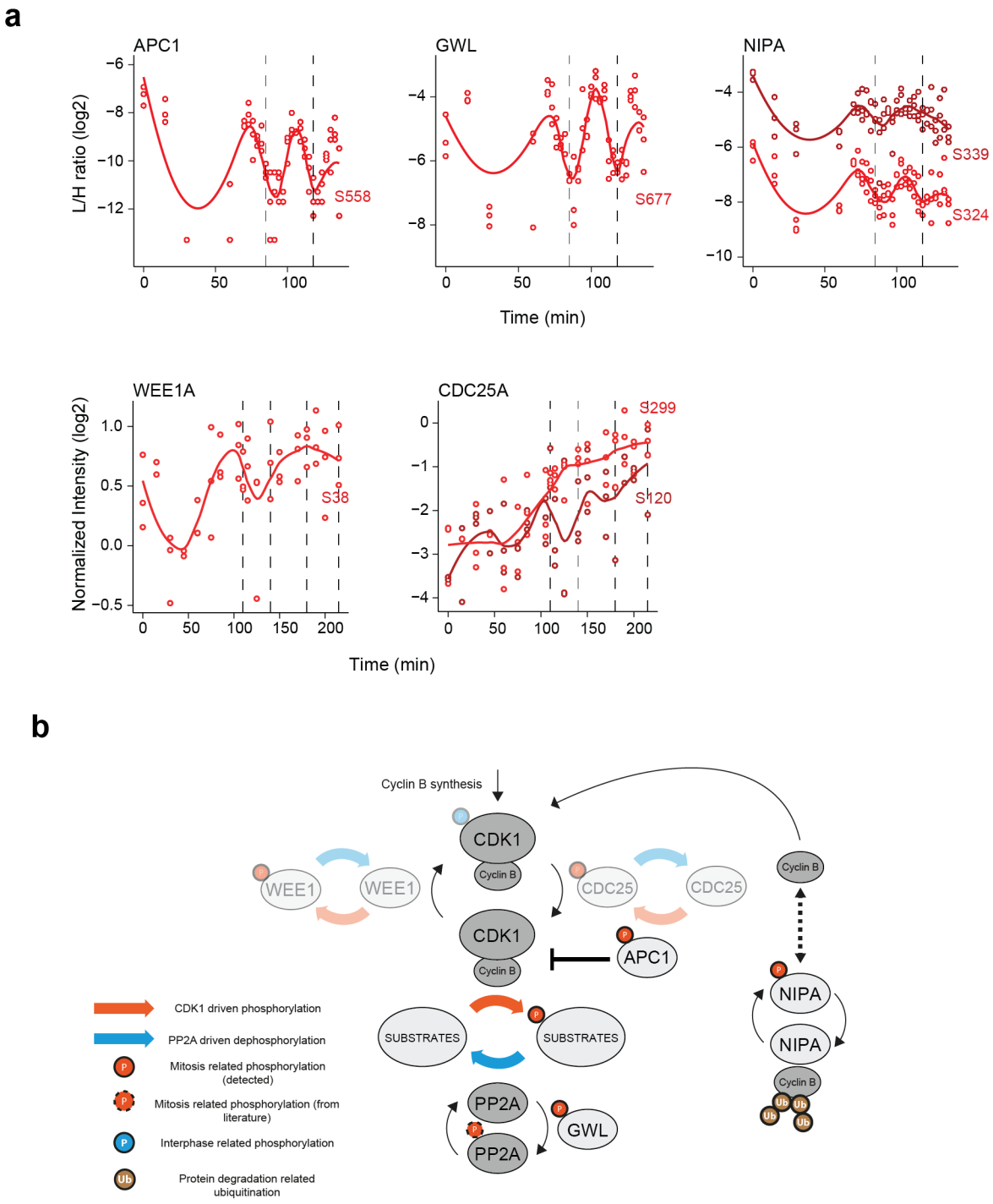

a
Interphase
Mitosis
Number of proteins
Category
BP
MF
CC
P−value (−log10)
DNA replication
chromosome, centromeric region
regulation of G2/M transition of mitotic cell cycle
nuclear pore
nuclear migration
P−body
cell division site
RNA polymerase binding
DNA biosynthetic process
actomyosin
cell−cell junction assembly
vesicle coat
cytoplasmic stress granule
mitotic cell cycle
spindle
DNA replication
RNA transport
mRNA binding
cell cycle checkpoint
helicase activity
nuclear pore
regulation of chromosome segregation
regulation of mRNA processing
mRNA 3'−end processing
regulation of translational initiation
double−stranded RNA binding
cytoplasmic stress granule
transcription elongation from RNA polymerase II promoter
mismatch repair
cleavage furrow
Cajal body
Fold enrichment (log10)
b
Z−Score
c
Cluster A
Cluster B
Cluster C
Cluster D
In vivo trend:
Cyclin B
CSF
DNA replication
Mitosis
d
MCM4:
GSNPPTPHGEEVQSPPSQR
S31 & T23
S31
Normalized Intensity (log2)
Time (min):
DNA replication
Mitosis

Extended data Fig. 3. *In vitro* phosphoproteomics discriminates interphase and mitotic phosphorylation; related to Figure 1. (a) Scatter plots of significantly enriched (Fisher's exact test with Bonferroni correction, $p<0.05$) GO terms for all dynamic phosphosites upregulated during interphase (left) and mitosis (right), presented as fold-enrichment of specific terms vs statistical significance. The size of the circles correlates with the number of proteins associated with the specific term, while the color corresponds to the GO term category. (b) Heatmap of dynamic phosphosites detected in DNA replication factors. (c) Behaviour of *in vivo* dynamic phosphosites (top) in *in vitro* experiments. (d) *In vitro* dynamics of T23 and S31 of MCM4. Orange curve shows upregulation of the multi-phosphorylated peptide (T23 and S31) during mitosis, while the blue curve shows the opposite trend for the singly-phosphorylated peptide (S31), as observed *in vivo*, in Extended data Fig. 1d.



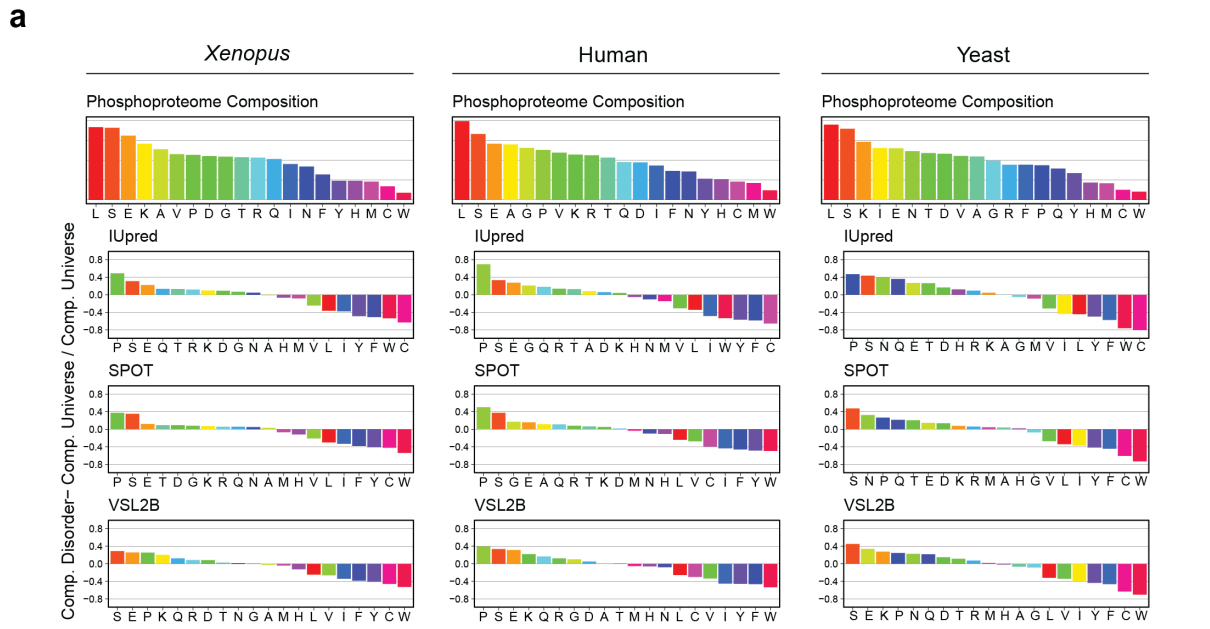
Extended data Fig. 4. CDK consensus phosphosites dominate the early embryo phosphoproteome; related to Figure 1. (a) Distribution of potential CDK targets among all detected phosphosites and dynamic phosphosites, *in vivo* (embryo, left) and *in vitro* (egg

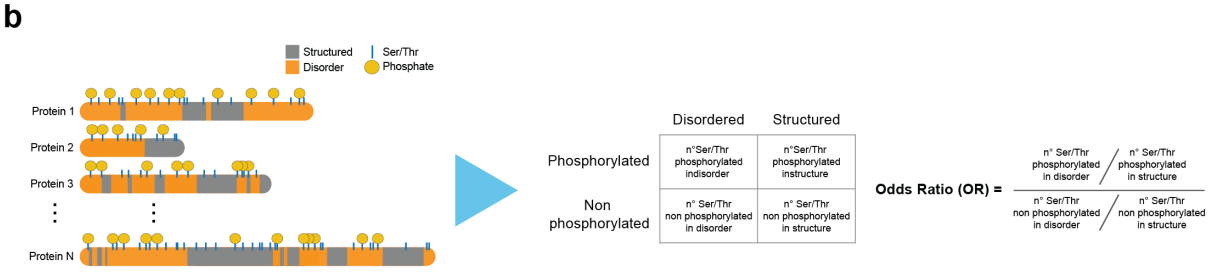
extract, right). (b) Observed phosphorylation motifs in the dynamic phosphoproteome *in vivo* (left) and *in vitro* (right). See methods for details. Note: in some cases, the sum of consensus sites exceeds the number of phosphosites due to redundancy between motif predictions. (c) Sequence motif logo for all dynamic phosphosites and for each of the clusters shown in Figure 1c, for the *in vivo* and *in vitro* experiments. Motifs are shown separately for proline-directed and non-proline directed phosphosites. (d) Dynamic trend of phosphorylations of potential kinase targets in egg extract S-phase clusters (4-6, left) and in mitotic clusters (1-3, right). (e) Venn diagram of observed *in vitro* (grey) and *in vivo* (blue) yeast CDK targets. *In vivo* targets showing CDK minimal consensus motif phosphorylations are highlighted in orange.



Extended data Fig. 5. Phosphorylation dynamics of the CDK1-oscillator network; related to Figure 2. (a) Single phosphosite plots of CDK1 regulators measured by targeted (top) or shotgun (below) phosphoproteomics. (b) CDK1-oscillator network: our data suggests that control of cyclin levels via positive (*e.g.* NIPA ubiquitin ligase) and negative (*e.g.* APC) feedback loops, accompanied by PP2A inactivation via GWL, can generate oscillation of CDK1 activity during early cell divisions. CDK1-Y15 regulation via feedback loops consisting

of CDC25 and WEE1A (greyed out) seems to be less important for switch-like mitotic phosphorylation after the first cell division.

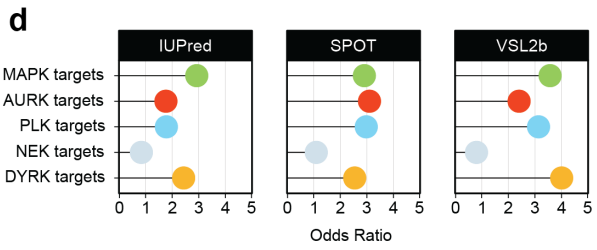


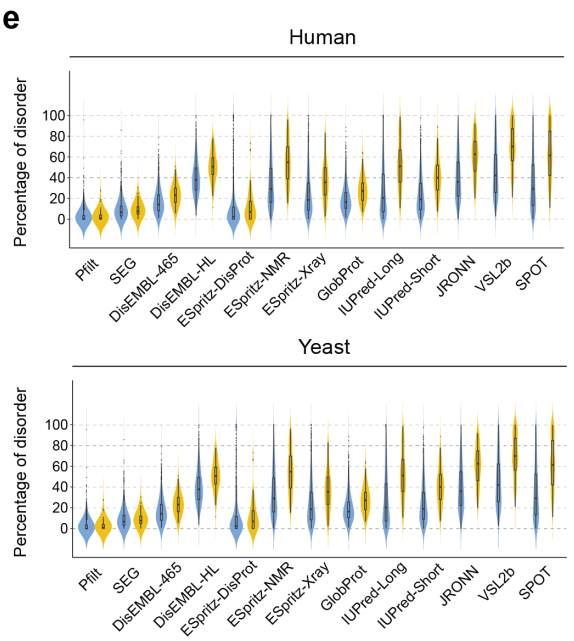


c

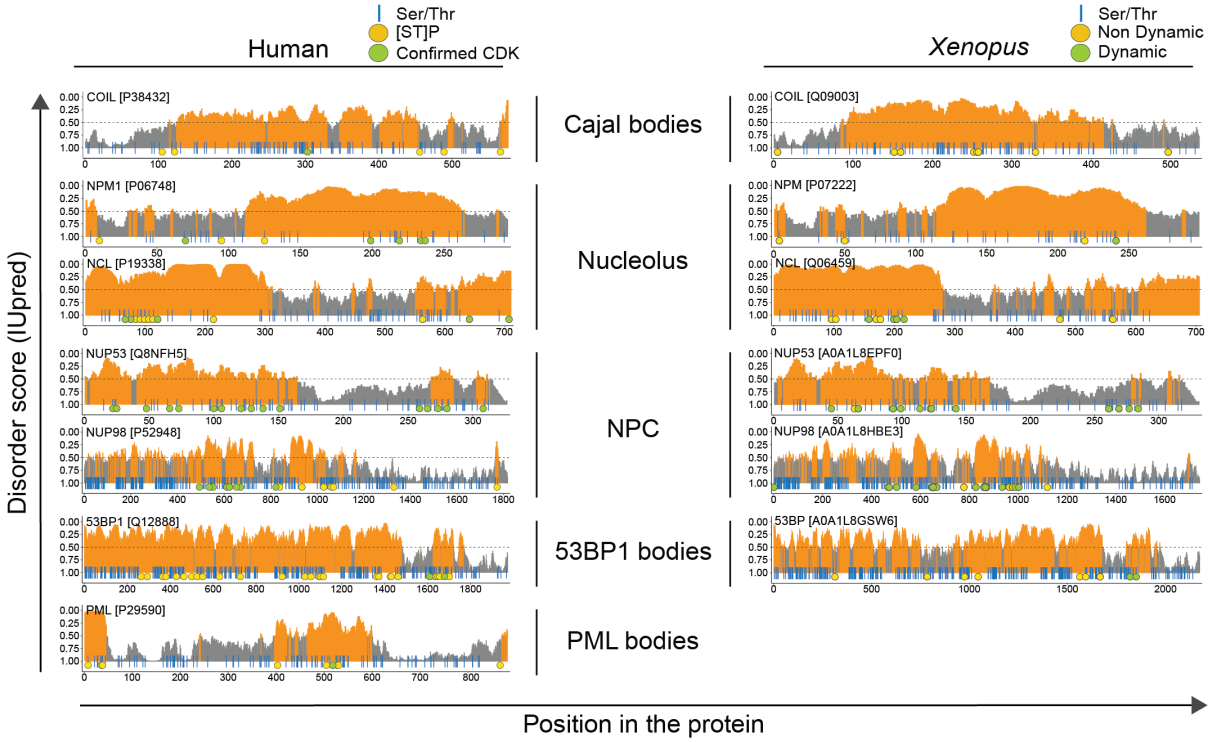
	IDR predictor	Disorder fraction	P-value	Odds ratio
Xenopus dynamic	IUPred	0.234	6.6E-51	2.70
	SPOT	0.277	1.3E-83	3.86
	VSL2b	0.381	2.2E-78	4.38
Human CDK targets	IUPred	0.259	4.9E-36	2.18
	SPOT	0.326	3.2E-56	3.05
	VSL2b	0.454	2.6E-54	3.84
Yeast CDK targets	IUPred	0.200	2.6E-31	4.89
	SPOT	0.277	7.8E-45	25.14
	VSL2b	0.380	8.6E-36	51.95

	IDR predictor	Disorder fraction	P-value	Odds ratio
Human MAPK targets	IUPred	0.259	6.2E-50	2.92
	SPOT	0.326	1.6E-45	2.91
	VSL2b	0.454	1.1E-45	3.57
Human AURK targets	IUPred	0.259	7.7E-15	1.77
	SPOT	0.326	6.0E-40	3.10
	VSL2b	0.454	1.6E-20	2.41
Human PLK targets	IUPred	0.259	1.0E-19	1.78
	SPOT	0.326	2.3E-51	2.98
	VSL2b	0.454	7.1E-42	3.14
Human NEK targets	IUPred	0.259	5.0E-01	0.84
	SPOT	0.326	7.4E-01	1.10
	VSL2b	0.454	3.5E-01	0.80
Human DYRK targets	IUPred	0.259	3.7E-04	2.43
	SPOT	0.326	3.2E-04	2.54
	VSL2b	0.454	6.2E-05	4.01

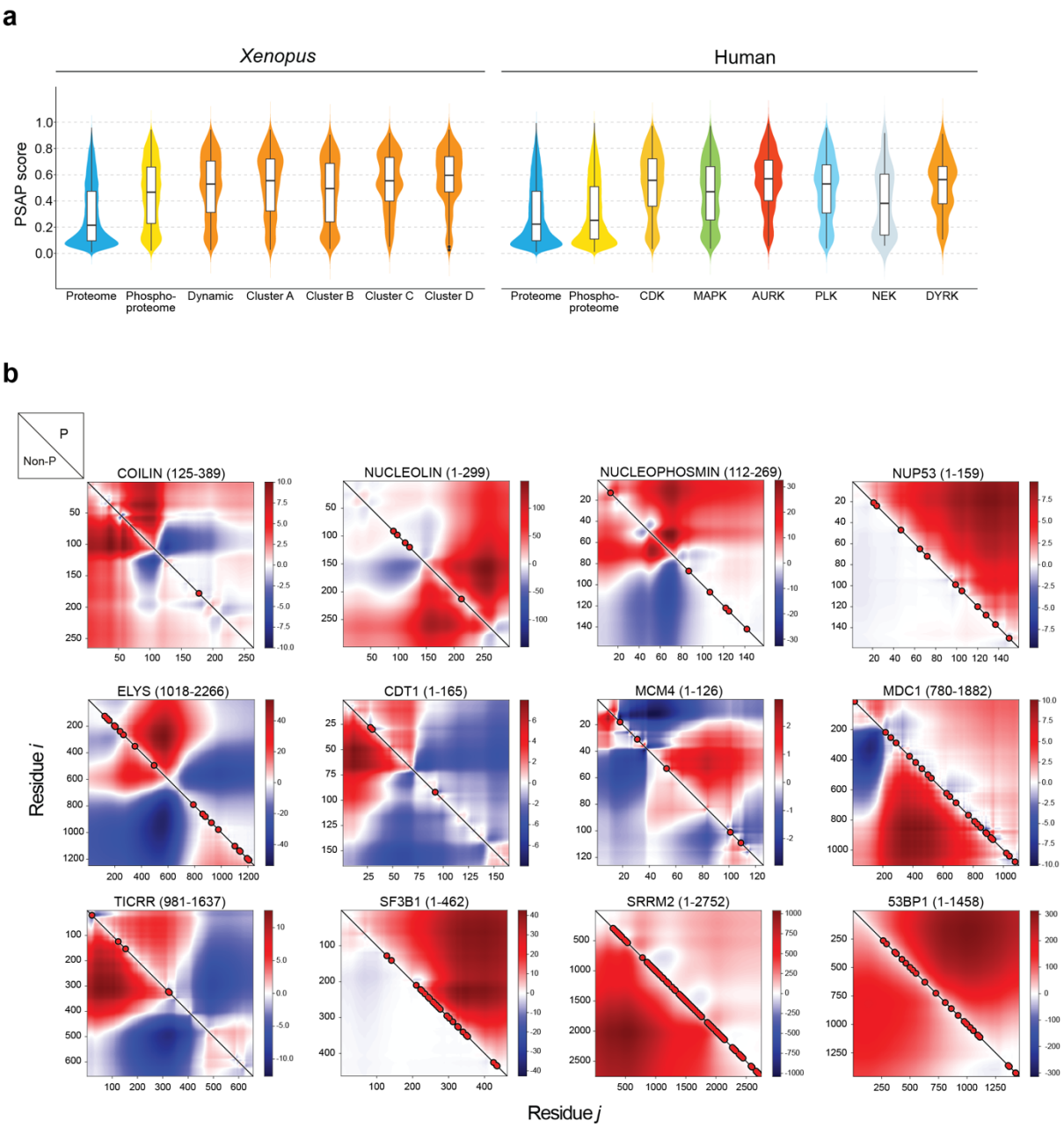




Extended data Fig. 6. Cell cycle phosphorylation occurs preferentially in IDRs; related to Figure 3. (a) Differential amino acid composition (see methods) in disordered regions for *Xenopus*, human and yeast determined with three IDR predictors. Amino acids are coloured in a rainbow pattern according to their relative abundance in each phosphoproteome. Disruptions of the rainbow pattern show specific compositional signatures for IDRs. (b) Scheme of the Odds Ratio analysis using contingency tables. The counts of phosphorylated Ser/Thr for all the proteins for each set (CDK-mediated in yeast/human, or dynamic in *Xenopus*, and other cell cycle-related kinases in human), in disordered and structured regions, are stored in a 2x2 table. (c) Tables showing results of statistical analysis of Odds Ratio with Fisher's test, using three disorder predictors. (d) Plots of the Odds Ratio for human cell cycle-related non-CDK kinases and MAPK. (e) Violin plots of the distribution of percentage of disordered residues per protein for CDK targets *vs* the rest of the phosphoproteome for human and yeast. Intrinsic disorder information of 13 different predictors was obtained from MobiDB, except for SPOT (calculated).

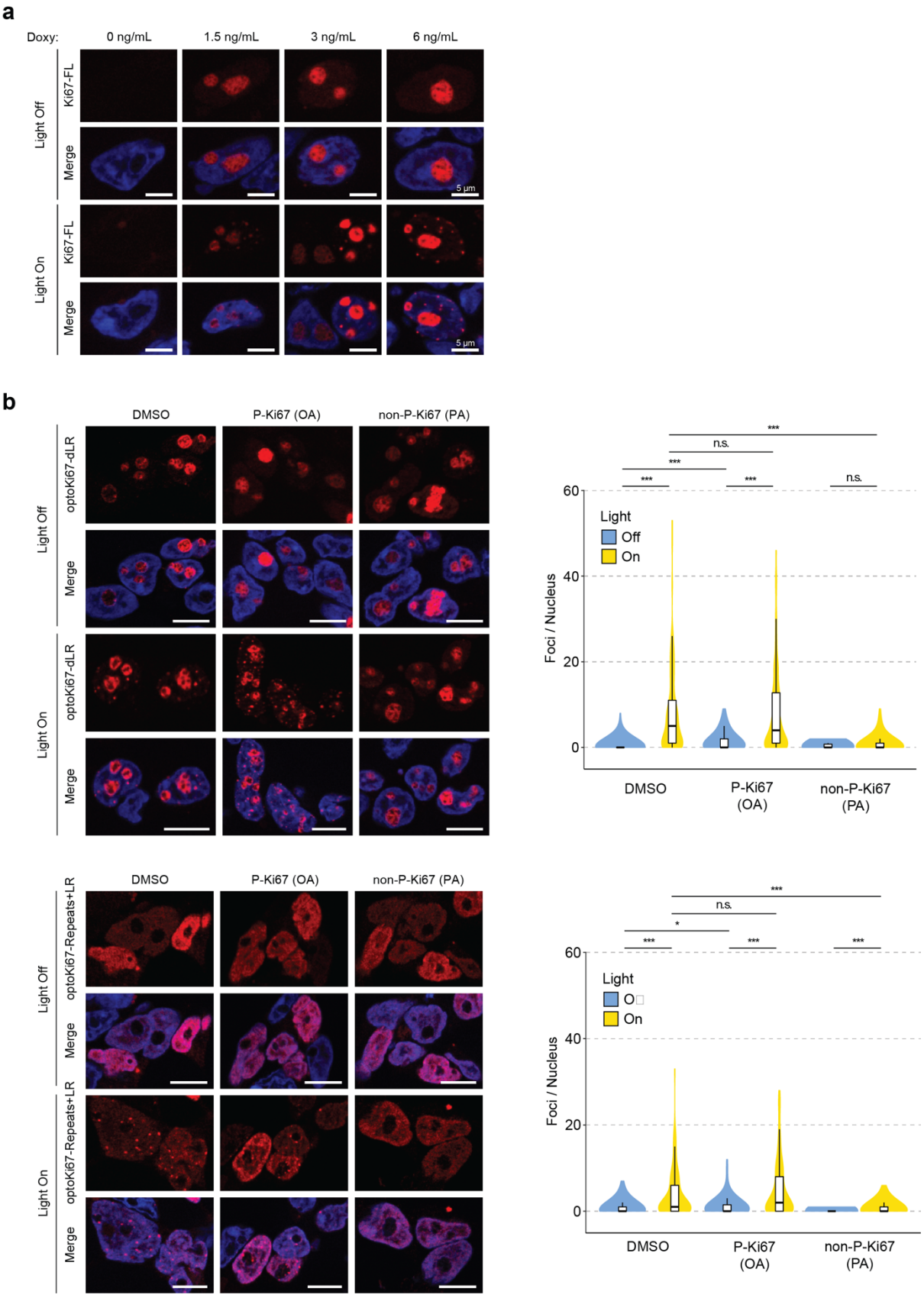


Extended data Fig. 7. Cell cycle phosphorylation of key MLO proteins; related to Figure 3. Diagrams of IUPred scores over the length of human CDK targets identified as primary components of MLOs in different studies, and their *Xenopus* homologues in this study. Regions with scores >0.5 (orange) are considered to be disordered, and <0.5 (grey) structured. Blue vertical lines indicate Ser and Thr residues; yellow circles, known Ser/Thr-Pro phosphosites (human) and non-dynamic phosphosites (*Xenopus*); green circles, confirmed CDK1 subfamily phosphorylations (human) and dynamic phosphorylations (*Xenopus*), from both embryos and egg extracts.



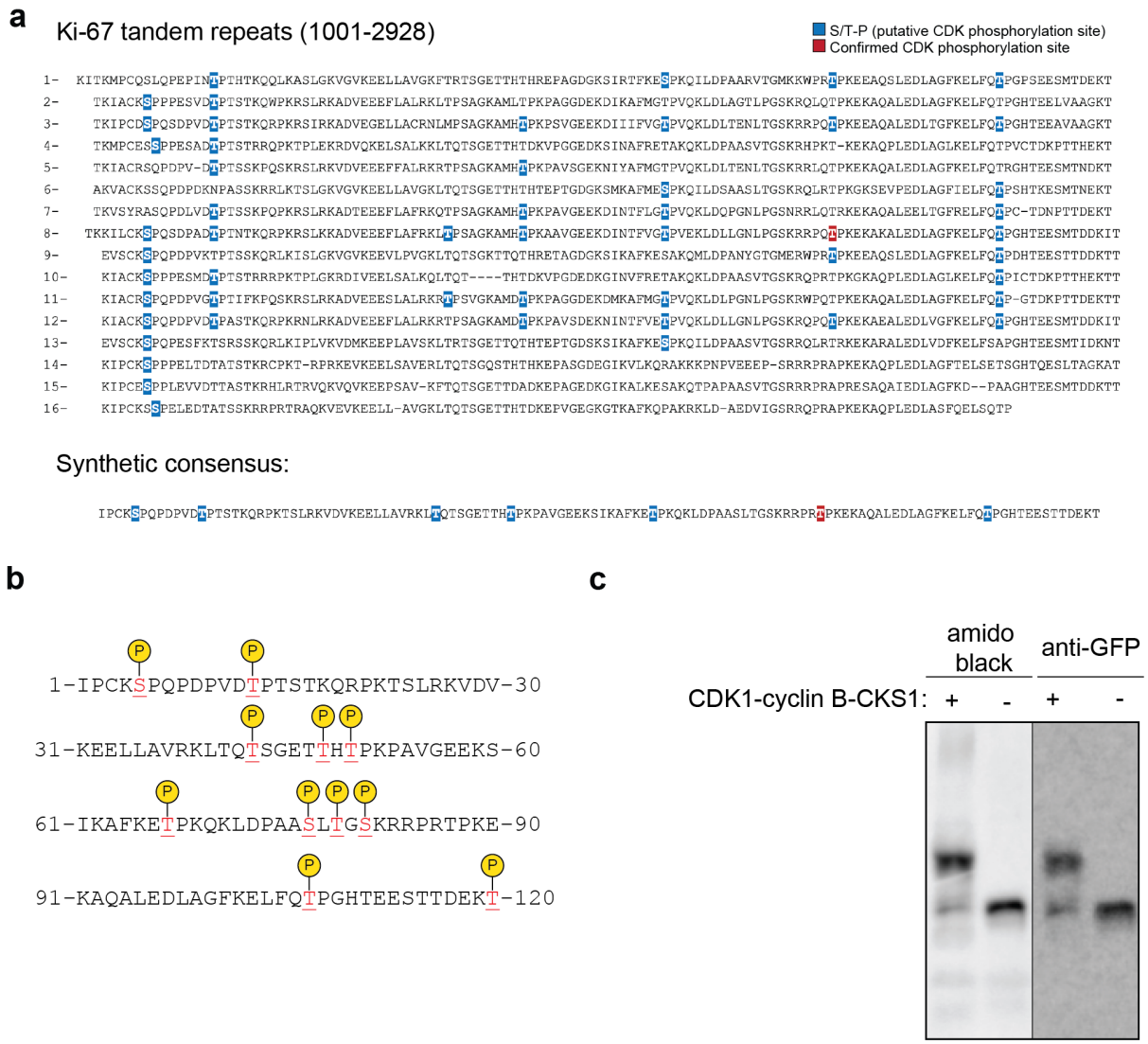
Extended data Fig. 8. CDK-mediated phosphorylation regulates IDR phase separation propensity; related to Figure 4. (a) Violin plots presenting PSAP score for *Xenopus* dynamic phosphoproteins (left), and human kinase targets, in comparison with total proteome and phosphoproteome. (b) Sequence Charge Decoration Matrix (SCDM) maps for a selection of human CDK targets and major MLO components (IDRs analysed are indicated), depicting the contribution of electrostatic interaction dictating the distance between two amino acid residues i and j (shown in x and y axes). The values of SCDM for different residue pairs (i,j) are shown using colour schemes with red and blue denoting positive (repulsive) and negative (attractive) values, respectively. The lower and upper triangles indicate SCDM map for the

unphosphorylated (non-P) and phosphorylated (P) sequences, respectively. Confirmed and putative (Ser/Thr-Pro) CDK phosphorylation sites are indicated with red circles.



Extended data Fig. 9. CDK-mediated phosphorylation regulates Ki-67 condensation; related to Figure 4. (a) Representative fluorescent images of HEK-293 cells expressing opto-

Ki-67 (FL) construct, induced by the indicated concentrations of doxycycline (Doxy), before (Light Off) and after (Light On) exposure to blue light. DNA was stained with Hoechst 33258. (b) Left, representative fluorescent images of cells expressing opto-Ki-67 constructs (with deleted LR domain, dLR, top; or with deleted N-terminus, Repeats+LR, bottom), before (Light Off) and after (Light On) exposure to blue light. Cells were pretreated for 1h with either vehicle (DMSO), 0.5 μM okadaic acid (OA), to inhibit protein phosphatase 2, or 5 μM purvalanol A (PA), to inhibit CDKs. DNA was stained with Hoechst 33258; scale bars, 10 μm. Right, quantification of results; the number of foci per nucleus was counted. Statistical significance was assessed by one-way ANOVA on ranks (Kruskal–Wallis test) and pairwise *post-hoc* comparisons using the Mann–Whitney test. P-values were adjusted by the Benjamini-Hochberg method.



Extended data Fig. 10. Ki-67 consensus repeat is stoichiometrically phosphorylated by CDK1 *in vitro*; related to Figure 4. (a) Alignment of human Ki-67 repeats, top, with the sequence of the consensus repeat depicted at the bottom. Confirmed and putative (Ser/Thr-Pro) CDK phosphorylation sites are highlighted in red and blue, respectively. (b,c) GFP-Ki-67 consensus repeat was phosphorylated *in vitro* using recombinant CDK1-cyclin B-CKS1 protein. Subsequently, the phosphosites were mapped by mass-spectrometry (b), and the stoichiometry of phosphorylation was analysed by Phos-Tag SDS-PAGE (c; amidoblack staining was used as loading control).

Materials and Methods

Egg collection and *in vitro* fertilisation.

Female *X. laevis* frogs were primed with 50 international units (IU) of human chorionic gonadotropin at least 2 days, and no more than 7-8 days before a secondary injection with 625 IU to induce ovulation. Roughly 16 hours after the second injection, fresh eggs were collected by pelvic massage and kept in 1x Marc's Modified Ringer's (MMR). Next, eggs were placed in a petri dish and checked under the microscope to keep only those that exhibited the healthy pigment pattern (dark animal pole and white vegetal pole).

To perform the *in vitro* fertilisation, around 1/3 of a full testis was cut into fine pieces and mixed with in 500 µl of 1x MMR. The suspension was pipetted up and down until big clumps were dissolved. Next, buffer was removed from the petri dish and the eggs were collected. Once eggs were well dispersed across the dish, the sperm suspension was added. The dish was then flooded with 0.1x MMR to induce fertilisation.

Sample collection.

The first time point was collected immediately before adding the sperm suspension (time 0' corresponds to the unfertilised egg). Eggs were kept at room temperature (18-20°) and under a dissecting microscope after fertilisation. At approximately 15 minutes, fertilised eggs underwent shrinkage of the animal hemisphere and rotation within the vitelline membrane, so that the animal hemisphere faced upwards. These changes are known indicators of successful fertilisation, so only the eggs that underwent these changes were used for the experiment.

Samples were collected approximately every 15 minutes. Eggs were rapidly placed in individual tubes and snap froze in liquid nitrogen, trying to preserve the phosphorylation events occurring at that specific time. Since the eggs were monitored under the microscope, we were able to determine if samples were collected before or after a cell division had occurred.

Xenopus egg extracts.

Interphase *Xenopus* egg extracts were prepared, and DNA replication time courses performed, as described previously[52]. Mitosis was induced in extracts by adding recombinant GST-Cyclin BΔ90 (40 ng/ml).

Cell lysis.

For the cell lysis, we used a similar approach to that described by Lindeboom *et al*[53]. Briefly, each sample was thawed and homogenised with 15 µL of ice-cold cell lysis buffer (20mM Tris-

HCl pH 8.0, 70mM KCl, 1mM EDTA, 10% glycerol, 5mM DTT, 0.125% Nonidet P-40, 1mM PMSF, 1x complete EDTA-free protease inhibitor, 1xPhoStop). Samples were subsequently centrifuged at max speed on a benchtop Eppendorf centrifuge. 10 µL of soluble material was recovered and snap-frozen in liquid nitrogen. Samples were stored at -80°C until further processing.

Protein digestion and phosphopeptide enrichment.

Cell lysates were digested using the FASP method[54]. Briefly: proteins were thawed and immediately reduced and alkylated with 10mM DTT and 0.05M iodoacetamide. Next, proteins were digested with Lys-C (overnight) at 37°C in a wet chamber, followed by addition of trypsin and further incubation under the same conditions for 4 hours. Both enzymes were used at 1:50 enzyme to protein ratio (protein quantification by Bradford assay showed that each individual egg provides ~20 µg of yolk free protein). For the egg extract experiment, each FASP filter was loaded with 200µg of protein. Peptides were cleaned using the Oasis HLB 96 well plates (Waters Corporation) and consequently subjected to phosphopeptide enrichment using Fe(III)-NTA 5µL cartridges in the automated AssayMAP Bravo Platform (Agilent Technologies), as described by Post *et al*[22]. Both flow through (peptides) and eluates (phosphopeptides) were dried down and stored at -80°C until further use.

LC-MS/MS analysis.

All samples for label-free shotgun proteomics were analysed using a UHPLC 1290 system (Agilent Technologies) coupled to an Orbitrap Q Exactive HF mass spectrometer (Thermo Fisher Scientific). Nano flow rate was achieved using a split flow setup aided by an external valve as described by Meiring *et al*[55]. Peptides were first trapped onto a pre-column (inner diameter [ID] of 100 µm and 2 cm length; packed in-house with 3µm C18 ReproSil particles [Dr. Maisch GmbH]) and eluted for separation into an analytical column (ID of 75 µm and 50cm length; packed in-house with 2.7 µm Poroshell EC-C18 particles (Agilent Technologies). The latter was done using a two-buffer system, consisting of buffer A (0.1% formic acid [FA] in water) and buffer B (0.1% FA in 80% ACN). Peptides were trapped during 5 minutes at 5 µL/min flow-rate with solvent A. For the measurement of the full proteome, we used a 155 min gradient from 10 to 36% of solvent B. For the phosphoproteome, we used a 95 min gradient from 8 to 32% of solvent B. Both methods included a wash with 100% solvent B for 5 minutes followed by a column equilibration with 100% solvent A for the last 10 minutes.

The mass spectrometer was operated in data dependent acquisition (DDA) mode. Full scan MS was acquired from 375-1600 m/z with a 60,000 resolution at 200 m/z. Accumulation target value was set to 3e6 ions with a maximum injection time of 20 ms. Up to 15 (12 for the phosphoproteome) of the most intense precursor ions were isolated (1.4m/z window) for fragmentation using high energy collision induced dissociation (HCD) with a normalised collision energy of 27. For MS2 scans an accumulation target value of 1e5 ions and a maximum injection time of 50 ms were selected. Scans were acquired from 200-2000m/z with a 30,000 resolution at 200m/z. Dynamic exclusion was set at 24s for the proteome and 12 s for the phosphoproteome.

For targeted proteomics, an EASY-nLC 1200 System (Thermo Fisher Scientific) coupled to an Orbitrap Q Exactive HF was used. Peptides were separated using an EASY-Spray analytical column (ID of 75µm and 25cm length; packed with 2µm C18 particles with a 100 Å pore size) (Thermo Fisher Scientific). Gradient lengths were shortened to 60 minutes. Phosphopeptides of interest from the previous experiment were selected and heavy-labeled versions were synthesised (JPT Peptide Technologies). These synthetic standards were used during method development for retention time scheduling and instrument ion fill-time optimisation. Additionally, synthetic heavy peptides were pooled and combined with synthetic retention time peptide standards (iRT, Biognosys) to generate a spectral library, measured in DDA mode using the same LC-MS setup. This spectral library provided fragment intensity and retention time information for quality control assessment of targeted measurements. Samples were reconstituted in 2% FA containing ~200fmol of each synthetic standard. The mass spectrometer was operated in data independent acquisition mode with an inclusion list of targets for parallel reaction monitoring (PRM). The list included the m/z values for the heavy and light versions of the phosphopeptides. Optimal measurement parameters were determined using test samples spiked with the heavy-labeled standards in order to guarantee optimal sensitivity for detection of endogenous phosphopeptides. We measured the targets of interest in a scheduled fashion, during a four-minute window with a 120,000 resolution, maximum injection time of 246ms and an accumulation target value of 2e5 ions, to ensure maximum specificity and sensitivity.

Data processing.

DDA raw files were processed with MaxQuant[56] (v1.6.0.1) using a false discovery rate (FDR) <0.01. The default settings were used, with the following exceptions: variable modifications, specifically methionine oxidation, protein N-term acetylation and serine, threonine and

tyrosine phosphorylation were selected. Cysteine carbamidomethylation was selected as a fixed modification. We also enabled the 'match between runs' option with the default values. Fractions were set so that matching was done only among biological replicates and samples of consecutive time points. The database search was conducted against a database generated by Temu *et al* [57]. This was particularly helpful since other publicly available databases contained several incomplete and/or poorly annotated sequences, which proved to be impractical for further data analysis.

The data was uploaded to the Perseus platform[58] for further analysis. Briefly: decoy sequences and potential contaminants were filtered out. Only high confidence localisation (>0.75 localisation probability) phosphosites were conserved for further analysis. Intensities were $\log_2$ transformed and then normalised by subtracting the median intensity of each sample. Biological replicates were grouped accordingly by time point; this grouping allowed us to filter the data and keep only those phosphosites that could be detected in at least two out of three biological replicates in any of the time points. Missing values were imputed from a random normal distribution applying a downshift of 1.8 times the standard deviation of the dataset, and a width of 0.3 times the standard deviation. This effectively replaced missing values at the lower end of the intensity distribution. We then performed an ANOVA (Benjamini-Hochberg FDR <0.05) to determine which phosphosites displayed statistically significant changes through the time course. Average phosphosite intensities were grouped using a combination of k-means and hierarchical clustering using the ComplexHeatmap package[59] in R. Protein intensities were processed in a similar fashion, removing proteins that were only identified by peptides that carry one or more modified amino acids.

Next, the full list of proteins with significantly changing phosphosites were matched against the human Uniprot database using the Basic Local Alignment Search Tool (BLAST), to render Uniprot identifiers that were compatible with different Gene Ontology (GO) analysis tools. We used the STRING web tool[60] to gain insight into the relation amongst dynamically phosphorylated proteins. The full list of phosphoproteins was uploaded and analysed using default settings. Next, the protein network was loaded into Cytoscape for clustering and visualisation. Proteins were clustered using GLay community clustering[61] and enrichment of GO terms per cluster was obtained using BiNGO[62] (shown in Fig. 1b).

GO term enrichment was also acquired individually for each group (A-D) obtained after hierarchical clustering of dynamic phosphosites. For this we used STRING and filtered the enriched GO terms to keep only those with $p < 0.01$, fold enrichment >5 and a minimum of 5 proteins per term. The list of terms was further condensed by removal of redundant terms using

the Revigo web tool[63]. Remaining GO terms (including BP, MF and CC) were manually curated to further avoid redundancy. The final set of GO terms per cluster are shown in Extended Data Fig. 1. Following the same strategy, we analysed GO term enrichment for the interphase and mitotic clusters from the *in vitro* dataset separately (shown in Extended data Fig. 3).

PRM raw files were analysed with Skyline software[64]. Signal quality for each target of interest was assessed visually for all samples. Quality control of endogenous signals was done by confirming the perfect co-elution of both peptide forms (heavy and light), assessing their retention time and peak shape. We also used the similarity of the relative intensity of fragment ions (rdotp > 0.9) between light and heavy to exclude signals that showed poor correlation. Quantifications were done with a minimum of three fragments per phosphopeptide. The data was loaded into R for data cleanup and visualisation, using the Complex-Heatmap and ggplot2[65] packages.

Motif analysis.

Obtained phosphopeptides were aligned by centering them around the phosphosite detected and the conserved motifs for the different kinases were determined using regular expressions by applying the following rules:

PLK: [D/N/E/Y]-X-[S/T]-[Hydrophobic / ^P]

AURA/AURB: [K/R]-X-[S/T]*[^P]

NEK: [L|M|F|W]-X-[S|T]*-[A|V|I||L|F|W|Y|M]-[K|R]

Casein kinase 1: [D/E]-[D/E]-[D/E]-X-X-[S/T]* or [S/T]-X-X-[S/T]*

Casein kinase 2:[S/T]-[S/T]*-X-[E/D/S]

DDK: [S/T]*-[E/D]-X-[E/D] or [S/T]*-[S/T]-P

PKA: R-[R/K]-X-[S/T]*-[Hydrophobic]

Cdk full consensus motif: [S/T]*-P-X-[K/R]

Cdk minimal consensus motif: [S/T]*-P

Where [] groups multiple amino acids for one position, ^ at the left of a certain amino acid informs that it is forbidden for that position, and X represents any amino acid.

Data collection for human and yeast CDK1 targets.

Data of CDK1 substrates for *S. cerevisiae* were downloaded from online supplementary information of papers describing two different studies using *in vitro* (4) and *in vivo* (26) approaches, respectively. We defined high confidence yeast CDK1 targets as the intersection

of both datasets. Other phosphorylations detected in both studies for which there was no evidence for CDK1 involvement were considered as the non-CDK1-mediated phosphoproteome (universe). For human CDK1 subfamily targets, we extracted information available in the PhosphoSitePlus database[66]. An additional step of manual curation from the following studies[6,7,48,67–82] was performed to obtain a high confidence human CDK1 subfamily targets dataset. The phosphoproteome universe was constructed with all the phosphorylated proteins deposited in the PhosphoSitePlus database after subtraction of the CDK1 subfamily targets. Data is available in Data S2.

Data collection for MLO proteomes.

Data from proteomics studies of the composition of MLOs characterised by liquid-liquid phase separation was obtained from the following sources: stress granules[83,84], nuclear speckles[84,85], PML nuclear bodies[84,86], P-bodies[87], nucleoli[88,89], nuclear pore complexes[90], Cajal bodies[84,91], Super-enhancer-Mediator condensates[92]. Data is available in Data S4.

Prediction of intrinsically disordered regions.

For the UniProt proteomes of human, yeast and *Xenopus laevis*, disorder information was fetched from MobiDB[93] with the exception of SPOT disorder predictor, which was calculated for all the proteins of each dataset. For Xenopus proteomics studies, we used the available standalone software of IUPred, VSL2B, and SPOT to predict IDRs in all the proteins of the database. Data is available in Data S3.

Differential disorder composition.

For the three organisms analysed (*Xenopus,* human and yeast), the amino acid composition for the entire phosphoproteome and for the disordered regions of the phosphoproteome was calculated. For each amino acid, we estimated the differential disorder composition with the equation: (Comp. Disorder – Comp. Phosphoproteome) / Comp. Phosphoproteome

Positives values show aminoacids enriched in disordered regions while negative values represent aminoacids depleted in disordered regions.

Bioinformatic and statistical analysis of disorder and phosphorylation.

All the statistical analysis was performed with the R programming language (https://www.r-project.org/) using R studio as an integrated development environment (available at https://rstudio.com/). The packages Tidyverse[65] and Bioconductor[94] were used for cleaning,

manipulation, and graphical representation of the data. Sequence logos were generated using the information content as described[95]. IUPred scores were plotted with an *ad hoc* designed script, available upon request.

For the contingency table analysis, the disordered regions of CDK targets and dynamic phosphoproteins were calculated with three predictors (IUPred, VSL2B, and SPOT). For each combination of disorder predictor and phosphorylation dataset, a two-by-two table with the counts of phosphorylatable residues (Ser/Thr) phosphorylated or not, either located in IDRs or in structured regions (Extended data Fig. 6b), was generated. Each table was then analysed with the Fisher test for obtaining the odds ratio and the associated P-value.

The source code for all the analysis conducted in this publication is available upon request.

Calculation of SCDMs.

Elements (i,j) of SCDM were calculated using equation 1 of Huihui and Ghosh[44]. In this coarse grain model, each amino acid is considered a point with a charge q = -1 for Aspartic and Glutamic acid, charge q = 1 for Arginine and Lysine, charge q = 0 for all other amino acids. Phosphorylation is modeled by replacing neutral charge of Serine, Threonine to q = -2 to mimic the effect of double negative charge of phosphate groups. SCDM maps were made visually continuous between neighboring (i,j) pairs using spline-16 interpolation.

All atom simulation.

We performed all-atom Monte Carlo simulation of MCM4, TICRR, CDT1, Coilin (unphosphorylated and phosphorylated forms) to compute radius of gyration by using CAMPARI (version 2) based on the ABSINTH implicit solvation paradigm[96,97]. Simulations were carried out at 298 K. For CDT1, 16 trajectories were generated with each trajectory having 10 million steps. For MCM4, 27 trajectories were generated with each trajectory having 6.5 million steps. For Coilin, 54 trajectories were generated with each trajectory running for 4 million steps. For TICRR, 90 trajectories were generated with each trajectory having 3 million steps. For each trajectory, irrespective of the sequence, first 1.5 million steps were discarded due to equilibration yielding a cumulative (over all trajectories) of at least 135 million steps for each sequence. Each of the simulations were run using zero salt with only neutralizing Na^+ and/or Cl^- ions added in a droplet of 400 Angstrom. PDB files were generated every 5000 Monte Carlo steps. Phosphorylated sequences were modeled by replacing Serine or Threonine by two Aspartic acids. ACE cap and NME tail were added to each protein sequence.

Coarse-grained force field.

We adopted a one-bead-per-residue coarse-grained (CG) model that has been shown to capture the structural and phase separation properties of flexible proteins as a function of their sequence[98] and phosphorylation pattern[99]. In this framework, bonded interactions between neighboring residues were modeled using a harmonic potential with a constant of 1000 kJ/mol/nm2 and a bond length of 0.38 nm. Electrostatic interactions between charged residues, i.e. Asp, Glu (-1e); Lys, Arg (+1e), His (+0.5e); ph-Ser, ph-Thr (-2e), were computed with a screened Coulomb potential, using a Debye-Huckel length of 1 nm. Short-range interactions were modeled by Lennard-Jones (LJ) potentials defined by

$$V_{ij}(r) = 4\lambda_{ij}\epsilon\left((\sigma_{ij}/r)^{12} - (\sigma_{ij}/r)^{6}\right)$$

where values of λ_{ij}, ϵ, and σ_{ij} used reported values[98,99] with the exception of λ_{ArgArg} that was set equal to 0.01, instead of 0.00, in order not to neglect excluded volume effects while using LJ functional form. In the simulation of full-length Ki-67, the conformation of the N-terminal folded domain (1-128) was restrained by an elastic network based on experimental structure (PDB: 1R21) with a distance cutoff of 2 nm and an elastic constant of 5000 kJ/mol/nm2. All MD simulations were performed with GROMACS 2018.3 (ref [100]). Simulations were run in the NVT ensemble, controlling the temperature by means of a Langevin thermostat with a friction constant of 25 ps^{-1} and a time step of 10 fs.

Single chain MD simulations of full-length Ki-67.

The structure with PDB id 1R21 was used to model the conformation of the first 128 residues of the full-length Ki-67, while TraDES[101,102] was used to generate the initial conformation of the remaining disordered part of the non-phosphorylated molecule. The models were then joined using UCSF Chimera 1.14 (ref [103]). The initial configuration for the phosphorylated full-length Ki-67 was obtained by 'mutating' (i.e. using parameters that have been tuned specifically for ph-Ser/ph-Thr) the amino acids in the non-phosphorylated structure. Initial configurations were inserted in a cubic simulation box with a side of 100 nm with periodic boundary conditions. We employed a Parallel Tempering (PT-MD) scheme, with eleven replicas in the 300-400K range in order to enhance the conformational sampling and obtain a reliable estimate of the radius of gyration of the protein as a function of the temperature. PT-MD simulation ran for 1e8 steps, attempting Monte Carlo exchanges between neighboring replicas every 1e2 steps and saving snapshots every 1e4 steps for further analysis.

Single chain MD simulations of consensus repeats.

The initial extended configuration of the non-phosphorylated consensus repeat of Ki-67 was generated with the tLEAP tool available in AmberTools18 (ref [104]), while the phosphorylated structure was generated by 'mutating' (i.e. using parameters that have been tuned specifically for ph-Ser/ph-Thr) the residues in the non-phosphorylated configuration. PT-MD simulation followed the same protocol employed for the simulations of full-length Ki-67, with the only difference being the range of temperatures, between 200K and 500K, with intervals of 10K.

Phase Coexistence MD simulations.

Phase coexistence simulations of monomers and dimers of the non-phosphorylated consensus repeat and of monomers of the phosphorylated consensus repeat were carried out employing the slab method[98]. Initial configurations of the slab simulations for each system were generated by inserting 200 copies of the respective molecules in a 20x20x30nm box in random positions and orientations with the *gmx insert-molecules* tool available in GROMACS 2018.3 (ref [100]), and extending the simulation box in the z direction to 200 nm. Simulations were run for 1.5e8 steps, saving frames every 10000 steps for further analysis, discarding the first 0.5e8 steps of the simulations as equilibration.

Estimation of theta temperatures.

Coil-to-globule transition temperatures (T_θ) were estimated for non-phosphorylated full-length Ki-67 and the monomer of the non-phosphorylated consensus repeat from single chain PT-MD simulations, following a reported approach[105]. Intramolecular distances as a function of the chain separation |i-j| were computed for each temperature replica and fitted with the following expression:

$$R_{ij} = 0.55|i-j|^{\nu}$$

where the scaling exponent ν is the fitting parameter. We then determined T_θ as the temperature corresponding to $\nu = 0.5$

Estimation of binodal curves.

Density profiles from phase coexistence simulations were estimated by means of the gmx density tool available in GROMACS 2018.3, centering the dense phase in the middle of the z-axis. The concentration of the diluted phase was computed by averaging the density in the box at 0nm<z<60nm and 140nm<z<200nm, while the concentration of the dense phase was

obtained by averaging the density at 90nm<z<110nm. Following the approach of [98] the critical temperatures, T_c, for the three systems investigated with phase coexistence simulations, have been evaluated by fitting the following equation:

$$\rho_H - \rho_L = A(T_c - T)^{0.325}$$

Human opto-Ki-67 plasmid construction.

pCDNA5_FRT_Ki67-FL-mCherry-Cry2 was generated by PCR amplification of human Ki67. The fragment was cloned into the AflII/KpnI digested pCDNA5_FRT_TO_TurboID-mCherry-Cry2 (Addgene plasmid # 166504) using the In-Fusion HD Cloning Kit protocol.

Generation of Flp-In T-REx 293 opto-Ki67 cell lines.

Flp-In T-REx 293 (Termo Fisher Scientifc, Darmstadt, Germany) cell line was grown under standard conditions (37˚C, 5% CO2) in Dulbecco's modified Eagle's medium (Merck-Sigma-Aldrich, D5796). The medium was supplemented with 10% fetal bovine serum (FBS), 100 µg/ml Zeocin and 15 µg/ml Blasticidin. One million HEK-293 T-REx cells were plated in a 6-well plate 24 hours before transfection. The next day, 500 ng of each optoKi67 expression plasmid is combined with 3.5 µg pOG44 encoding the Flp recombinase. Transfection was made with 8 µl Lipofectamine 2000 according to the manufacturer's instructions. 48 hours post transfection, cells were transferred to a 100 mm petri dish. On the next day, selection was performed by adding hygromycin B at a final concentration of 50 µg/mL. Around 14 days after selection, clones were pooled and expanded. The cells were tested for the expression of the construct by immunofluorescence. Flp-InT-Rex 293 opto-Ki67 stable cell lines were maintained with 15 µg/mL Blasticidin and 15 µg/mL Hygromycin.

Opto-Ki-67 activation.

Cells were plated in DMEM on coverslips a day prior to activation. Expression of opto-Ki67 was induced with 2 µg/ml doxycycline for 16 hours. For light activation, plates were transferred into a custom-made illumination box containing an array of 24 LEDs (488nm) delivering 10 mW/cm2 (light intensity measured using a ThorLabs-PM16-121-power meter). Cry2 activation was induced using 4 min of blue light cycles: 4s On followed by 10s Off. Cells were fixed with 4% paraformaldehyde (PFA) for 15 min at RT, counterstained with Hoechst (Invitrogen, Cat H21491) in PBS-TritonX (0.2%), and mounted on glass slides using Prolong Gold antifade reagent (Invitrogen, Cat P36930). Images were captured using a 63x objective (NA 1.46 oil).

To check the effect of inhibitors on Ki-67 foci formation, the cells were incubated for 1h with 0.5 µM okadaic acid, 5 µM purvalanol A or vehicle (DMSO) for 1 hour prior to light activation and fixation. Foci number was analysed using FIJI Software and statistical significance was assessed by one-way ANOVA on ranks (Kruskal–Wallis test) and pairwise post-hoc comparisons using the Mann–Whitney test. P-values were adjusted by the Benjamini-Hochberg method. Plots were generated using the ggplot2 library in R.

Ki-67 consensus repeat DNA construct.

The cDNA sequence coding for Ki-67 consensus repeat (Ki67-CR) was ordered from IDT® gene synthesis. It was subsequently cloned into pDB-GFP plasmid between HindIII and XhoI sites to obtain the pDB-GFP-Ki-67-CR vector. In this construct, Ki-67-CR was fused with a (his)6-GFP N-terminal tag. GFP sequence is followed by the HRV 3C (3C) protease recognition site (Leu-Glu-Val-Leu-Phe-Gln/Gly-Pro). Specific cleavage can occur between Gln and Gly, with Gly-Pro remaining at the C terminus Ki-67-CR without any tag.

Ki-67 consensus repeat expression and purification.

The pDB-GFP-Ki-67-CR plasmid was transformed into *E. coli* BL21(DE3); transformed cells were grown overnight at 25°C in N-5052 auto-induced medium[106], supplemented with 50 µg/ml kanamycin and 15N NH4Cl. Cells were harvested by 20 min centrifugation at 6000g at 4°C. The pellet was resuspended in 20 mM Tris-HCl pH 7.5, 300 mM NaCl and 2mM DTT (buffer A) and stored at -80°C. Cells were supplemented with a Complete® EDTA free tablet (Roche), lysed by sonication, insoluble proteins and cell debris were sedimented by centrifugation at 40000g at 4°C for 30 min. Supernatant was supplemented with imidazole to 5 mM final and loaded onto 5ml gravity affinity columns (Ni Sepharose Excel 5ml, Cytiva), equilibrated with buffer A. Columns were washed with 50 ml of buffer A and proteins were eluted with a one-step gradient of buffer B (buffer A containing 500 mM imidazole). The peak fractions were analysed by SDS-PAGE. Fractions containing tagged Ki-67-CR were pooled and dialysed overnight at 4°C against buffer C (50 mM Bis-Tris pH 6.7, 50 mM NaCl, 2 mM DTT). The dialysed protein was then loaded on a Superdex S200 16/60 (HiLoad 16/600 Superdex 200pg, Cytiva) equilibrated with buffer C. Fractions containing the protein of interest were analysed by SDS-PAGE and pooled. The purified GFP-Ki-67-CR protein was concentrated to 5 mg/ml with Vivaspin centrifuge concentrator (Sartorius Stedim Biotech).

***In vitro* Ki-67 peptide phosphorylation assay.**

Protein was desalted by using PD10 Mini-Trap column in 50 mM Hepes 7.5, 50 mM NaCl, 2 mM DTT. Phosphorylation reaction was performed in a total volume of 220 µl, and contained 140 µl of GFP-Ki-67-CR (400 µg), 5mM $MgCl_2$, 500 µM ATP, 50mM β-glycerophosphate, recombinant CDK1-cyclin B/CKS1 (5 µg; gift from Jane Endicott and Tony Ly) in 50 mM Hepes pH 7.5, 50 mM NaCl, 2 mM DTT. Reaction was incubated at 30°C for 18 hours and stopped by adding 10 mM EDTA. It was subsequently desalted by using PD10 Mini-Trap column in 50 mM Bis-Tris pH 6.7, 50 mM NaCl, 2 mM DTT, and used to perform NMR experiments, Phos-tag SDS-PAGE and phosphoproteomics.

NMR experiments and data analysis.

All NMR samples contained final concentrations of 10% D_2O and 0.5 mM 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS). Experiments were performed at 293 K on a Bruker Avance III spectrometer equipped with a cryogenic triple resonance probe and Z gradient coil, operating at a 1H frequency of 700 and 800 MHz. 15N-HSQC was acquired for each sample in order to determine amide (1HN and 15N) chemical shifts of non-phosphorylated and phosphorylated GFP-tagged Ki-67-CR. 15N-HSQC spectra were acquired for respectively non-phosphorylated (and phosphorylated) proteins at 800 (700) MHz using 32 (128) scans, 128 (256) increments and a spectral width of 22.5 ppm in the indirect dimension. All spectra were processed with TopSpin v3.5 (Bruker Biospin) and analysed using CCPN-Analysis software[107]. Chemical shifts were referenced with respect to the H_2O signal relative to DSS using the 1H/X frequency ratio of the zero point according to Markley et al[108].

Total internal reflection fluorescence (TIRF) experiments.

In order to measure the ability of the phosphorylated and non-phosphorylated Ki-67 repeat to liquid-liquid phase separate, 50 µM protein was prepared in 50 mM Bis-Tris pH 6.7, 50 mM NaCl, 2 mM DTT. Dextran was added just before preparing samples on circular glass coverslips (2.5 cm, 165 µm thick, Marienfeld). Coverslips were cleaned with a 15 min cycle of sonication with ultrasounds in 1M KOH, followed by a second cycle of sonication in deionized water. Samples were deposited into wells of Press-to-seal silicone isolater with adhesive (Invitrogen), and covered with a second coverslip to avoid evaporation. Images were acquired with a custom-made TIRF microscope using a LX 488-50 OBIS laser source (Coherent). Oil immersion objective with a 1.4 numerical aperture (Plan-Apochromat 100x, Zeiss) was used. Fluorescence was collected with an EmCCD iXon Ultra897 (Andor) camera. The setup includes a 1.5x telescope to obtain a final imaging magnification of 150-fold, corresponding to

a camera pixel size of 81.3 nm. Fluorescence images at different time points were obtained by averaging 150 individual images, each acquired over 50 ms exposure time.

Phos-tag SDS-PAGE.

For Phos-tag SDS-PAGE (12.5% Phos-tag™ SuperSep™ pre-cast gel 50 μmol/L; Fujifilm Wako Chemicals #193-16571), 500 ng of unphosphorylated and 500 ng of CDK1-cyclin B-CKS1 phosphorylated GFP-Ki-67-CR protein were loaded. SDS-PAGE was performed following standard protocol, with the exception of two additional washes 20 min each in transfer buffer containing 10 mM EDTA, followed by a wash in transfer buffer, preceding wet transfer. For Western blot, anti-GFP antibody (rabbit polyclonal Chromotek PABG1; 1:10 000) was used.

Mapping of Ki-67-CR phosphorylation sites by mass spectrometry

6.5 μg of GFP-Ki-67-CR protein phosphorylated by CDK1-cyclin B-CKS1 were digested in a FASP filter as described earlier for the other samples. Peptides were subsequently cleaned using C18 cartridges and phosphor-enriched using Fe(III)-NTA cartridges in the AssayMAP Bravo. Phosphopeptides were measured in a technical duplicate, acquiring in DDA mode with an Ultimate 3000 uHPLC system coupled to an Orbitrap Exploris 480 (Thermo Fisher Scientific) during a 60 minutes gradient. Raw files were searched against the Ki-67-CR sequence using MaxQuant with the same parameters as applied to the other phosphoproteomics experiments.

References

52. Parisis, N. *et al.* Initiation of DNA replication requires actin dynamics and formin activity. *EMBO J.* (2017) doi:10.15252/embj.201796585.
53. Lindeboom, R. G. H., Smits, A. H., Perino, M., Veenstra, G. J. C. & Vermeulen, M. Mass Spectrometry-Based Absolute Quantification of Single Xenopus Embryo Proteomes. *Cold Spring Harb Protoc* (2018) doi:10.1101/pdb.prot098376.
54. Wisniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat Methods* **6**, 359–62 (2009).
55. Meiring, H. D., van der Heeft, E., ten Hove, G. J. & de Jong, A. P. J. M. Nanoscale LC-MS(n): technical design and applications to peptide and protein analysis. *J Sep Sci* **25**, 557–568 (2002).

56. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
57. Temu, T., Mann, M., Raschle, M. & Cox, J. Homology-driven assembly of NOn-redundant protEin sequence sets (NOmESS) for mass spectrometry. *Bioinformatics* **32**, 1417–9 (2016).
58. Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods* **13**, 731–40 (2016).
59. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).
60. Szklarczyk, D. *et al.* STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* **47**, D607–D613 (2019).
61. Su, G., Kuchinsky, A., Morris, J. H., States, D. J. & Meng, F. GLay: community structure analysis of biological networks. *Bioinformatics* **26**, 3135–7 (2010).
62. Maere, S., Heymans, K. & Kuiper, M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* **21**, 3448–9 (2005).
63. Supek, F., Bosnjak, M., Skunca, N. & Smuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* **6**, e21800 (2011).
64. MacLean, B. *et al.* Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**, 966–8 (2010).
65. Wickham, H. *et al.* Welcome to the Tidyverse. *Journal of Open Source Software* **4**, 1686 (2019).
66. Hornbeck, P. V. *et al.* PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res* **43**, D512–D520 (2015).
67. Orthwein, A. *et al.* Mitosis Inhibits DNA Double-Strand Break Repair to Guard Against Telomere Fusions. *Science* **344**, 189–193 (2014).
68. Wyatt, H. D. M., Sarbajna, S., Matos, J. & West, S. C. Coordinated Actions of SLX1-SLX4 and MUS81-EME1 for Holliday Junction Resolution in Human Cells. *Molecular Cell* **52**, 234–247 (2013).
69. Linder, M. I. *et al.* Mitotic Disassembly of Nuclear Pore Complexes Involves CDK1- and PLK1-Mediated Phosphorylation of Key Interconnecting Nucleoporins. *Developmental Cell* **43**, 141-156.e7 (2017).
70. Liu, J. *et al.* Cell cycle-dependent localization of the CDK2-cyclin E complex in Cajal (coiled) bodies. *J Cell Sci* **113**, 1543–1552 (2000).

71. Chi, Y. *et al.* A novel landscape of nuclear human CDK2 substrates revealed by in situ phosphorylation. *Science Advances* **6**, eaaz9899 (2020).
72. Klein, U. R., Haindl, M., Nigg, E. A. & Muller, S. RanBP2 and SENP3 Function in a Mitotic SUMO2/3 Conjugation-Deconjugation Cycle on Borealin. *MBoC* **20**, 410–418 (2008).
73. Goto, H. *et al.* Complex formation of Plk1 and INCENP required for metaphase–anaphase transition. *Nature Cell Biology* **8**, 180–187 (2006).
74. Bartsch, O., Horstmann, S., Toprak, K., Klempnauer, K.-H. & Ferrari, S. Identification of cyclin A/Cdk2 phosphorylation sites in B-Myb. *European Journal of Biochemistry* **260**, 384–391 (1999).
75. Curtis, M., Nikolopoulos, S. N. & Turner, C. E. Actopaxin is phosphorylated during mitosis and is a substrate for cyclin B1/cdc2 kinase. *Biochem J* **363**, 233–242 (2002).
76. Fourest-Lieuvin, A. *et al.* Microtubule Regulation in Mitosis: Tubulin Phosphorylation by the Cyclin-dependent Kinase Cdk1. *Molecular Biology of the Cell* **17**, 1041 (2006).
77. Milner, R. E., Busaan, J. L., Holmes, C. F., Wang, J. H. & Michalak, M. Phosphorylation of dystrophin. The carboxyl-terminal region of dystrophin is a substrate for in vitro phosphorylation by p34cdc2 protein kinase. *J. Biol. Chem.* **268**, 21901–21905 (1993).
78. Lowe, M. *et al.* Cdc2 Kinase Directly Phosphorylates the cis-Golgi Matrix Protein GM130 and Is Required for Golgi Fragmentation in Mitosis. *Cell* **94**, 783–793 (1998).
79. Yun, J. *et al.* Cdk2-dependent Phosphorylation of the NF-Y Transcription Factor and Its Involvement in the p53-p21 Signaling Pathway. *J. Biol. Chem.* **278**, 36966–36972 (2003).
80. Kitzmann, M. *et al.* cdk1- and cdk2-Mediated Phosphorylation of MyoD Ser200 in Growing C2 Myoblasts: Role in Modulating MyoD Half-Life and Myogenic Activity. *Molecular and Cellular Biology* **19**, 3167–3176 (1999).
81. Thiel, D. A. *et al.* Cell Cycle-Regulated Phosphorylation of p21-Activated Kinase 1. *Current Biology* **12**, 1227–1232 (2002).
82. Li, M., Stefansson, B., Wang, W., Schaefer, E. M. & Brautigan, D. L. Phosphorylation of the Pro-X-Thr-Pro site in phosphatase inhibitor-2 by cyclin-dependent protein kinase during M-phase of the cell cycle. *Cellular Signalling* **18**, 1318–1326 (2006).
83. Jain, S. *et al.* ATPase-Modulated Stress Granules Contain a Diverse Proteome and Substructure. *Cell* **164**, 487–498 (2016).

84. Fong, K. *et al.* Whole-genome screening identifies proteins localized to distinct nuclear bodies. *J Cell Biol* **203**, 149–164 (2013).
85. Dopie, J., Sweredoski, M. J., Moradian, A. & Belmont, A. S. Tyramide signal amplification mass spectrometry (TSA-MS) ratio identifies nuclear speckle proteins. *J Cell Biol* **219**, (2020).
86. Liu, J. *et al.* Functional proteomic analysis of promyelocytic leukaemia nuclear bodies in irradiation-induced MCF-7 cells. *J Biochem* **148**, 659–667 (2010).
87. Hubstenberger, A. *et al.* P-Body Purification Reveals the Condensation of Repressed mRNA Regulons. *Molecular Cell* **68**, 144-157.e5 (2017).
88. Stenström, L. *et al.* Mapping the nucleolar proteome reveals a spatiotemporal organization related to intrinsic protein disorder. *Molecular Systems Biology* **16**, e9469 (2020).
89. Tafforeau, L. *et al.* The complexity of human ribosome biogenesis revealed by systematic nucleolar screening of Pre-rRNA processing factors. *Mol Cell* **51**, 539–51 (2013).
90. Lin, D. H. & Hoelz, A. The Structure of the Nuclear Pore Complex (An Update). *Annu. Rev. Biochem.* **88**, 725–783 (2019).
91. Machyna, M., Heyn, P. & Neugebauer, K. M. Cajal bodies: where form meets function. *WIREs RNA* **4**, 17–34 (2013).
92. Quevedo, M. *et al.* Mediator complex interaction partners organize the transcriptional network that defines neural stem cells. *Nature Communications* **10**, 1–15 (2019).
93. Piovesan, D. *et al.* MobiDB: intrinsically disordered proteins in 2021. *Nucleic Acids Res* doi:10.1093/nar/gkaa1058.
94. Gentleman, R. C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* **5**, 1–16 (2004).
95. Douglass, J. *et al.* Identifying protein kinase target preferences using mass spectrometry. *Am J Physiol Cell Physiol* **303**, C715–C727 (2012).
96. Vitalis, A. & Pappu, R. V. Methods for Monte Carlo simulations of biomacromolecules. *Annu Rep Comput Chem* **5**, 49–76 (2009).
97. Vitalis, A. & Pappu, R. V. ABSINTH: a new continuum solvation model for simulations of polypeptides in aqueous solutions. *J Comput Chem* **30**, 673–699 (2009).
98. Dignon, G. L., Zheng, W., Kim, Y. C., Best, R. B. & Mittal, J. Sequence determinants of protein phase behavior from a coarse-grained model. *PLOS Computational Biology* **14**, e1005941 (2018).

99. Regy, R. M., Thompson, J., Kim, Y. C. & Mittal, J. Improved coarse-grained model for studying sequence dependent phase separation of disordered proteins. *Protein Sci* **30**, 1371–1379 (2021).
100. Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).
101. Feldman, H. J. & Hogue, C. W. A fast method to sample real protein conformational space. *Proteins* **39**, 112–131 (2000).
102. Feldman, H. J. & Hogue, C. W. V. Probabilistic sampling of protein conformations: new hope for brute force? *Proteins* **46**, 8–23 (2002).
103. Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605–1612 (2004).
104. Case, D. A. *et al.* The Amber biomolecular simulation programs. *J Comput Chem* **26**, 1668–1688 (2005).
105. Dignon, G. L., Zheng, W., Best, R. B., Kim, Y. C. & Mittal, J. Relation between single-molecule properties and phase behavior of intrinsically disordered proteins. *Proc Natl Acad Sci U S A* **115**, 9929–9934 (2018).
106. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* **41**, 207–234 (2005).
107. Vranken, W. F. *et al.* The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59**, 687–696 (2005).
108. Markley, J. L. *et al.* Recommendations for the presentation of NMR structures of proteins and nucleic acids. *J Mol Biol* **280**, 933–952 (1998).