Supporting Information

Dephoure *et al.* 10.1073/pnas.0805139105

SI Materials and Methods

Cell Culture and Synchronization. Equal numbers of HeLa cells were cultured in 15-cm plates containing 25 ml of Dulbecco's modified Eagle's medium with either 80 mg/liter arginine-[$^{13}C_6$, $^{15}N_4$] and 50 mg/liter lysine-[$^{13}C_6$, $^{15}N_2$] (Cambridge Isotope Laboratories) or DMEM containing equal concentrations of conventional arginine and lysine. Both heavy and light DMEM were supplemented with 10% dialyzed FBS (HyClone). When the cells reached 70% confluence, heavy labeled cells were harvested by scraping. The cell pellet was washed with 5 ml of ice-cold PBS, snap-frozen in liquid nitrogen, and stored at −80°C. Thymidine was added to the unlabeled cells to a final concentration of 2.5 mM. After 19 h, the cells were washed three times with 25 ml of PBS and released into light media without added thymidine for 9 h. Thymidine was again added to 2.5 mM and the cells were allowed to grow for 16 h. Half of these cells were harvested as described. The remaining cells were washed three times with PBS and released into media containing 0.2 μg/ml nocodazole (Sigma–Aldrich) until 90% of cells were rounded up. The cells were collected by mitotic shake-off and processed as before. Approximately 5×10^7 thymidine or nocodazole-arrested cells were mixed with an equal number of asynchronous cells.

Protein Digestion. Proteins were digested with trypsin as described in ref. 1 with some modifications. Briefly, 10 ml of lysis buffer [8 M urea, 20 mM Hepes (pH 8.0), 1 mM glycerophosphate, 2.5 mM sodium pyrophosphate, 1 mM sodium vanadate] was added to each cell pellet and the samples were sonicated three times for 1 min. Lysates were reduced with 5 mM DTT at 60°C for 20 min and alkylated with 10 mM iodoacetamide in the dark at room temperature for 30 min. The samples were diluted to 2 M urea by the addition of 20 mM Hepes (pH 8.0). Sequencing grade modified trypsin (Promega) enzyme was added at a 1:100 ratio and digests were allowed to proceed overnight at room temperature with gentle agitation. Digestion was stopped by the addition of trifluoroacetic acid (TFA) to a final concentration of 1% and precipitates were removed by centrifugation at 4,000 rpm for 20 min. The supernatants were applied to preequilibrated Sep-Pak C18 columns (Waters), and the columns were washed with 0.1% TFA, 2.5% acetonitrile. Bound peptides were eluted by the stepwise addition of 10, 15, 20, 25, 30, 35, and 40% acetonitrile in 0.1% TFA. The eluates were combined and lyophilized.

SCX and Phosphopeptide Enrichment. Strong cation exchange (SCX) chromatography was performed as described in ref. 2. Briefly, 12 mg of a heavy/light (asynchronous/arrested peptide) mixture was resuspended in 800 μl of SCX buffer A (5 mM KH_2PO_4, pH 2.65, 30% acetonitrile) and separated on a 9.4 mm × 200 mm column packed with polysulfoethyl aspartamide (5 μm particles; 200 angstrom pores; PolyLC), using a 35 min gradient from 0% to 21% buffer B (5 mM KH_2PO_4, pH 2.65, 30% acetonitrile, 350 mM KCl) at a flow rate of 2 ml/min. Twelve fractions were collected, dried in a speed-vac evaporator (ThermoFisher), resuspended in 5% formic acid, and desalted using 100-mg tC18 SepPak cartridges (Waters). Eluted peptides from each fraction were split into two tubes each, dried, and stored at −20°C. One-half of each sample was resuspended in 200 μl of IMAC wash/binding buffer (250 mM acetic acid, 30% acetonitrile) and added to 50 μl of a 1:1 slurry of precharged IMAC resin (Phos-Select iron affinity gel, Sigma–Aldrich) washed 3 times in wash/binding buffer. Samples were agitated for 90 min at room temperature and washed three times with 200 μl of wash/binding buffer. Bound peptides were eluted with 2× 150 μl of 50 mM Tris (pH 10.0), 300 mM NH_4OH and added to 60 μl of 10% TFA. The remainder of each SCX fraction was dissolved in 200 μl of TiO_2 binding buffer (2 M dihydroxybenzoic acid, 50% acetonitrile, 0.1% TFA) and added to 50 μl of 1:1 slurry of Titansphere TiO_2 beads (GL Sciences) washed three times in the same buffer. Peptides were agitated at room temperature for 90 min, washed one time with 200 μl of binding buffer, two times with 200 μl of 50% acetonitrile, 0.1% TFA, and eluted and acidified as they were for IMAC. Eluted peptides from IMAC and TiO_2 samples were pooled with samples from the corresponding SCX fraction and dried down in a Speed-Vac. All samples were resuspended in 20 μl of 5% formic acid and desalted on C18 resin, using handmade STAGE tips (3). Peptides were eluted into glass inserts with 50% acetonitrile, 5% formic acid, dried once again, and stored at −20°C.

Database Searching and Filtering. MS/MS spectra were searched by using the SEQUEST (4) algorithm (v.27, rev.13) against a composite database containing the human IPI protein database (ftp.ebi.ac.uk/pub/databases/IPI/current, version 3.20) and its reversed complement. Search parameters allowed for three missed tryptic cleavages, a mass tolerance of ±100 ppm, a static modification of 57.02146 Da (carboxyamidomethylation) on cysteine, up to six total dynamic modifications, up to three of any particular dynamic modification, and dynamic modifications of 79.96633 Da (phosphorylation) on serine, threonine, and tyrosine, 15.99491 Da (oxidation) on methionine, 10.00827 Da on arginine, and 8.01420 Da on lysine. Additional searches were performed on the first six SCX fractions from each experimental condition with a mass tolerance of ± 15 ppm allowing up to five phosphorylations and up to eight total dynamic modifications. Search results were filtered to include <1% matches to reverse sequences by restricting the mass tolerance window, and setting thresholds for Xcorr and dCn′, defined as the dCn to the first nonidentical amino acid sequence for a match. Additional filtering was performed to remove sequences that contained both heavy and light isotopic variants of arginine and lysine. Combining all analyses from both experiments 200 reverse sequence hits remained in the final list, representing an estimated false-discovery rate of 0.3% after removal of the 200 reverse hits.

Phosphorylation Site Localization and Counting. To assign phosphorylation site localizations and measure the confidence of assignment, we applied a probabilistic algorithm (5) that considers all phosphoforms of a peptide and uses the presence or absence of experimental fragment ions unique to each to create an ambiguity score (Ascore). The Ascore indicates the likelihood that the best site match is correct when compared with the next best match. As in refs. 6 and 7, we considered sites with Ascore ≥13 ($P \leq 0.05$) to be confidently localized, and those with Ascore ≥19 ($P \leq 0.01$) to be localized with near certainty. In some cases, we detected the same peptide multiple times with different site localizations. Where we were unable to confidently assign both localizations, it was impossible to determine whether the two observations corresponded to one or two phosphorylation events. We used a conservative counting strategy, choosing the lower number of sites in such cases, to estimate the minimum number of phosphorylation sites in our dataset (2). The number of sites counted on a given peptide was only allowed to exceed

the maximum number of sites seen in a single observation of that peptide if the number of sites assigned with near certainty (Ascore $\geq$ 19, $P < 0.01$) allowed it.

Data Comparisons. To avoid errors due to variations in site position assignments, all comparisons to large-scale datasets were done by first generating 13-mers flanking the phosphorylated residues by six amino acids for all phosphorylation sites (shorter for some N- and C-terminal peptides). Mutual look-ups requiring site and protein matches were performed in Excel to determine the overlap. The set of known spindle proteins was as defined in (8). A comparison of all sites observed on heavy-labeled peptides from asynchronous cells when mixed with light G_1 samples or light M-phase samples yielded 9,768 sites in G_1 and 9,753 in M-phase. Of these 5,898 (60.5% of M-phase sites) sites were found in both.

Peptide Quantification. Automated peptide quantification was performed by using the Vista program (9). Briefly, the theoretical mass of both heavy and light variants of each peptide was calculated and used to identify ion peaks in the high mass accuracy precursor scans for each. The intensity of the peaks was used to construct ion chromatograms. Candidate peaks were required to fall within a tolerance window of $\pm$ 20 ppm from the calculated mass and were filtered to require the predicted isotopic distribution. For each isotopic variant, the background-subtracted area under the curve was determined as a function of elution time and used to calculate the light (G_1 or M-arrested) to heavy (asynchronous) abundance ratio. The S/N ratio of each heavy or light peptide species is defined as the ratio of the maximum peak intensity to the background noise baseline and where possible, is reported for every quantified peptide. We required S/N $\geq$ 3 for both heavy and light species for quantification. In some cases only the heavy or light version of a peptide was found and a S/N ratio only exists for one isotopic peptide species. For these peptides found exclusively in one sample, we have reported the peak S/N ratio or its inverse, as a proxy for relative abundance measurement. For such peptides, we required S/N $\geq$ 5 for the observed species. These peptides are noted in Table S3 Raw abundance ratios from each experiment were normalized based on the average abundance ratio of the unphosphorylated peptides. Phosphopeptides that exhibited a $\geq$2.5-fold increase in relative abundance were deemed up-regulated.

A nonredundant set of quantified peptides (Table S3) was produced by selecting the highest quality quantification for each. Peptides differing only in methionine oxidation state, or by the presence of amino or carboxyl-terminal missed cleavages, were considered redundant. Peptides were selected based on confidence score, minimum S/N, and total S/N. To generate motif specific $\log_2$ratio distribution plots we chose all quantified peptides harboring each of the three motif classes: acidophilic [pS/pT]-X-X-[D/E], basophilic [K/R]-X-X-[pS/pT], proline [pS/pT]-P, and those lacking one. We further required that each motif-specific phosphorylation site was localized with an Ascore $>$10 ($P < 0.1$) and that the peptide did not contain any potential motifs of another class.

Replicate Analysis. Changes between the two independent relative abundance measurements for each peptide were expressed as a ratio of ratios, by calculating the difference in the $\log_2$ abundance ratios. The dependence of these variations on S/N was examined by plotting $\Delta\log_2$ ratios as a function of the minimum S/N ratio of all four measured peptide species (heavy and light from each of two runs). Percentage deviation from the mean was calculated for each peptide pair as $|[\log_2\text{Ratio} - \text{Avg}(\log_2\text{Ratio})]|/\text{Avg}(\log_2\text{Ratio})$. Peptides with detectable levels of both "heavy" and "light" species in both analyses were grouped into subtractive bins based on minimum S/N thresholds, such that at S/N = 0 all peptides were selected and as S/N increased only those peptides meeting each threshold remained. The percentage of peptide pairs in each bin whose percentage deviation from the mean was $<$10%, 20%, and 30% was then calculated.

Regulated Phosphorylation Scores. RePh scores were calculated independently for G_1 and M-phase phosphopeptides. The score for a given site or combination of sites on a peptide is the median abundance ratio of all peptides harboring that unique combination of sites with scores capped at 100. Medians were calculated by using $\log_2$ transformed ratios and then converted back to values comparable to the original ratios as $2^{\text{median}(\log_2 \text{ratio})}$. Protein RePh scores (P-RePh) are the sum of all individual RePh scores for each protein. The fraction of P-RePh scores exceeding successively higher score thresholds was plotted for all G_1 and M-phase proteins and for two subsets of the M-phase data—proteins corresponding to genes annotated to the GO term "mitotic cell cycle" (GO:0000278), $n = 273$, and a set of proteins known to have mitotically regulated phosphorylation, $n = 153$. Only those proteins found in the mitotic dataset were included, 111 proteins for the GO associated collection and 113 known mitotic phosphoproteins. The set of known mitotic phosphoproteins was identified from searches of the literature. We chose to include only vertebrate proteins for which we could find direct evidence of increased phosphorylation in mitosis. To avoid bias, searches did not include gene names. The list is not comprehensive. Topo-plots of G_1 and M-phase P-RePh scores were generated in Matlab.

Gene Ontology and Protein Interaction Network Analysis. The top 1,000 scoring mitotic genes were submitted to the DAVID Bioinformatics Resources, 2008 (10). Separate searches were performed for biological process, cellular component, and molecular function. A minimum of five hits was required. Similar terms with highly redundant gene lists were removed. The most highly enriched terms are reported in Table S8. The 342 gene products with a G_1 P-RePh $>$0 were analyzed as above for GO biological process term enrichment. Protein–protein interaction network diagrams were generated with data from the complete human protein reference database interaction dataset (11) (version 7), using Cytoscape (12).

1. Matsuoka S, *et al.* (2007) ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. *Science* 316:1160–1166.
2. Villen J, Beausoleil SA, Gerber SA, Gygi SP (2007) Large-scale phosphorylation analysis of mouse liver. *Proc Natl Acad Sci USA* 104:1488–1493.
3. Rappsilber J, Ishihama Y, Mann M (2003) Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal Chem* 75:663–670.
4. Eng JK, McCormack AL, Yates JR, III (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* 5:976–989.
5. Beausoleil SA, Villen J, Gerber SA, Rush J, Gygi SP (2006) A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat Biotechnol* 24:1285–1292.
6. Wilson-Grady JT, Villen J, Gygi SP (in press) Phosphoproteome Analysis of Fission Yeast. *J Proteome Res.*
7. Li X, Gerber SA, Rudner AD, Beausoleil SA, Haas W, Villen J, Elias JE, Gygi SP (2007) Large-scale phosphorylation analysis of alpha-factor-arrested Saccharomyces cerevisiae. *J Proteome Res* 6:1190–1197.
8. Sauer G, Korner R, Hanisch A, Ries A, Nigg EA, Sillje HH (2005) Proteome analysis of the human mitotic spindle. *Mol Cell Proteomics* 4:35–43.
9. Bakalarski CE, *et al.* (2008) The impact of peptide abundance and dynamic range on proteome-scale quantitative analyses. *J Proteome Res*, in press.
10. Dennis G, Jr, *et al.* (2003) DAVID: Database for annotation, visualization, and integrated discovery. *Genome Biol* 4:P3.
11. Mishra GR, *et al.* (2006) Human protein reference database—2006 update. *Nucleic Acids Res* 34:D411–414.
12. Shannon P, *et al.* (2003) Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res* 13:2498–2504.

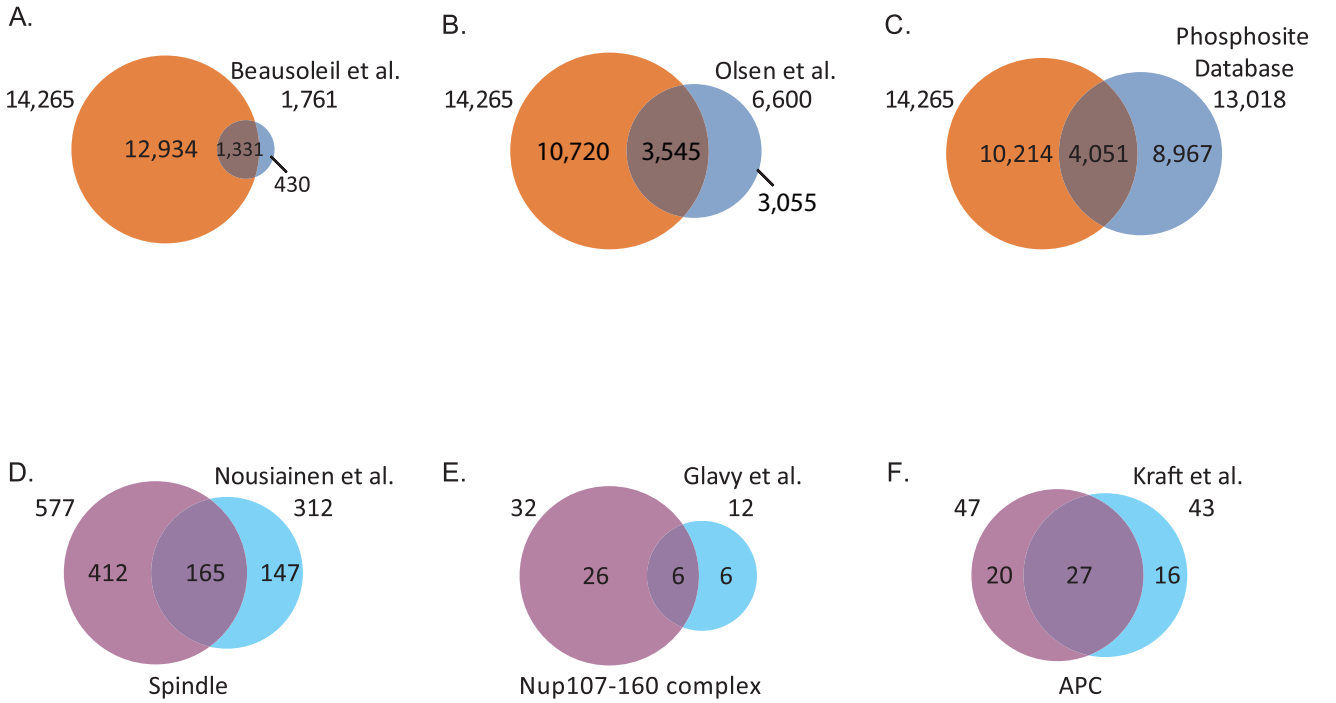

Fig. S1. Comparison to existing phosphoproteomic datasets. Shown are the total number of sites from and the overlap between our dataset and earlier analysis of nocodazole treated HeLa cells (9) (*A*), EGF stimulated HeLa cells (10) (*B*), a collection of sites curated from the literature (19) (*C*), purified spindle proteins from HeLa cells (20) (*D*), purified Nup107–160 complexes (21) (*E*), and purified anaphase promoting complexes (22) (*F*).

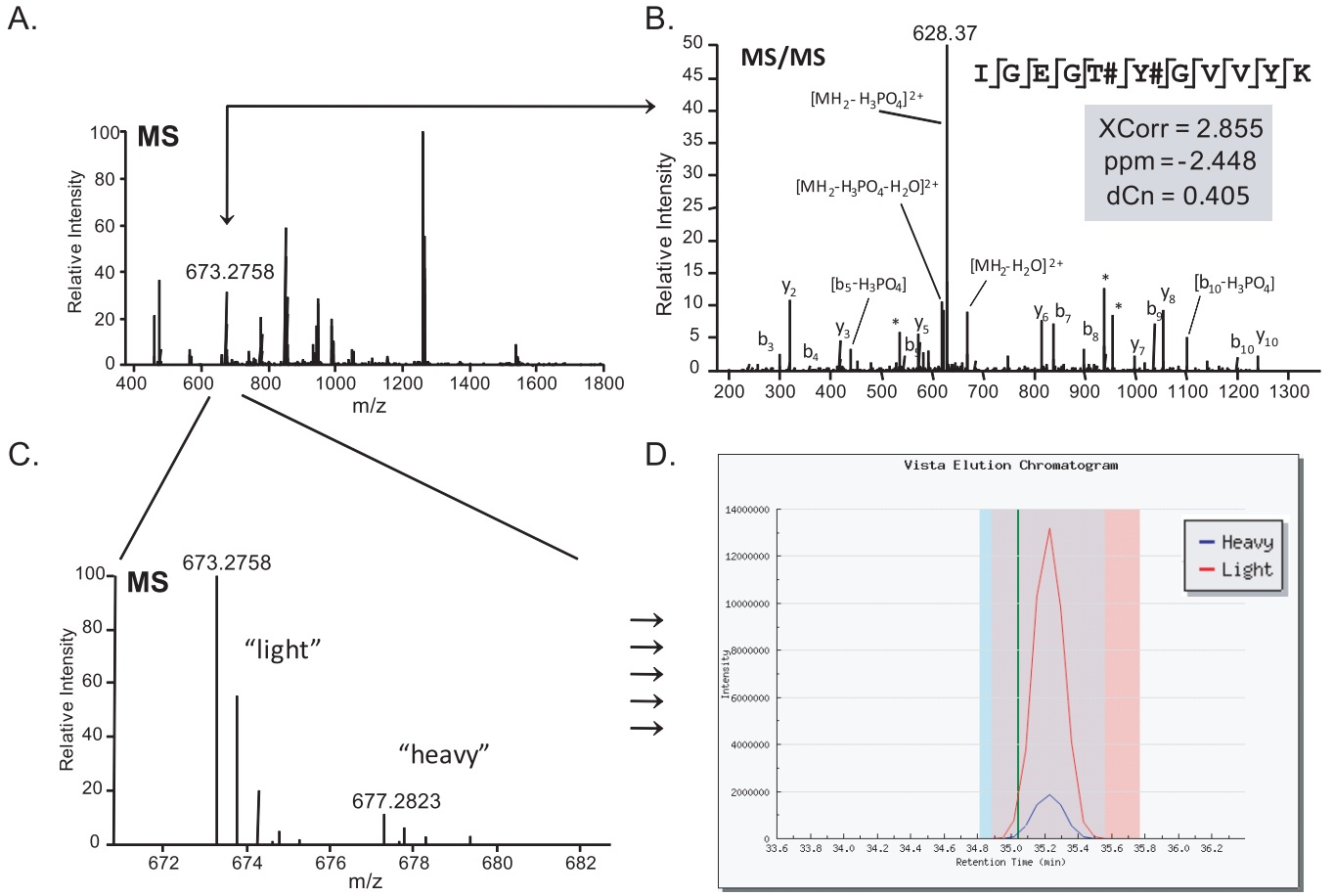

Fig. S2. Overview of automated quantification. (*A*) Shown is a full precursor MS scan performed in the Orbitrap from the G_1 sample. (*B*) The precursor ion with *m/z* = 673.2758 was chosen for fragmentation and MS2 analysis. An annotated spectrum is shown along with the matched peptides sequence from Cdc2, IGEGT#Y#GVVYK, where # denotes a phosphorylated residue. ⋆ indicates an unexplained peak in the spectrum. (*C*) An exploded section of the full MS scan shown in *A* reveals two isotopic envelopes corresponding to ''light'' and ''heavy'' forms of the unfragmented peptide. (*D*) For each peptide identified by searching MS/MS spectra, the software searches for the isotope envelopes corresponding to the precursor mass of both ''light'' and ''heavy'' peptide forms in full MS scans surrounding the point where the original MS/MS was collected. These intensities are used to construct extracted ion chromatograms for both species as shown in (*D*). Shown is an example of the quantification for a phosphopeptide containing inhibitory Cdc2 phosphorylations.

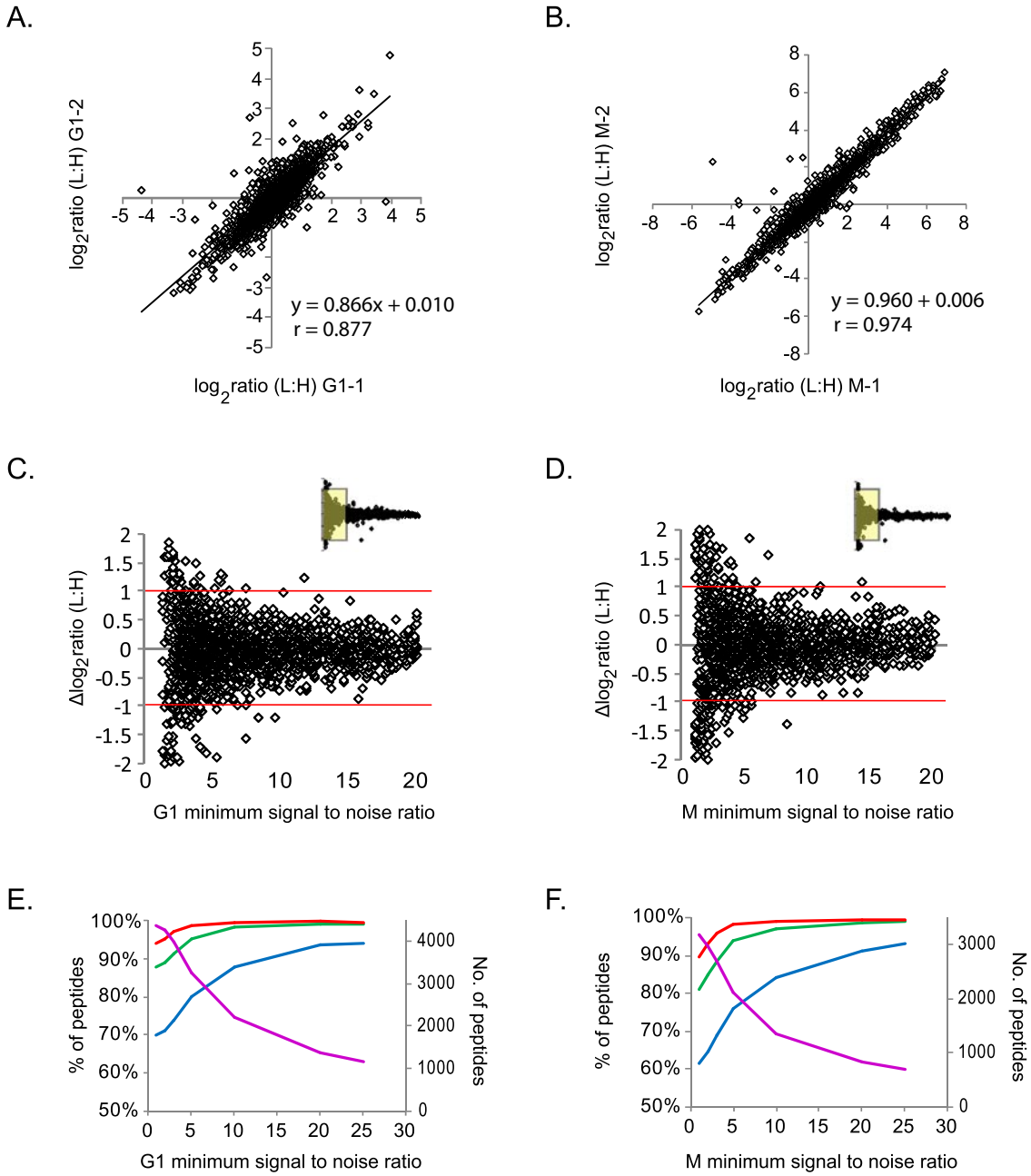

Fig. S3. Reproducibility of abundance ratio measurements. All samples were run and analyzed twice. (*A* and *B*) For peptides identified in both analyses with minimum signal to noise ratio ≥3, the $\log_2$ transformed abundance ratio (L:H) from run 1 was plotted against the that from run 2 for G_1 (*A*) and M phase samples (*B*). Linear regression analysis was used to fit the data to a line. The resulting equations and *r* values are shown. (*C* and *D*) The $\Delta\log_2$ transformed ratio $|[\log_2(\text{Ratio}_1) - \log_2(\text{Ratio}_2)]|$ for all peptides quantified in both analyses is plotted as a function of the minimum signal to noise ratio observed for either isotopic variant in both analyses. (*Inset*) Area of detail highlighted in the context of the larger dataset. $\Delta\log_2$ falls sharply as the minimum signal to noise ratio increases. Red lines indicate the position of a 2-fold change. Vista ratios are highly reproducible when the peak signal to noise ratio is ≥3 for both species. (*E* and *F*) The fraction of quantified peptide pairs with a given minimum peak signal to noise ratio that deviate from the mean no more than 10% (blue trace), 20% (green trace), or 30% (red trace). The number of qualifying peptide pairs at each signal to noise cutoff is shown in purple and graphed on the secondary axis.

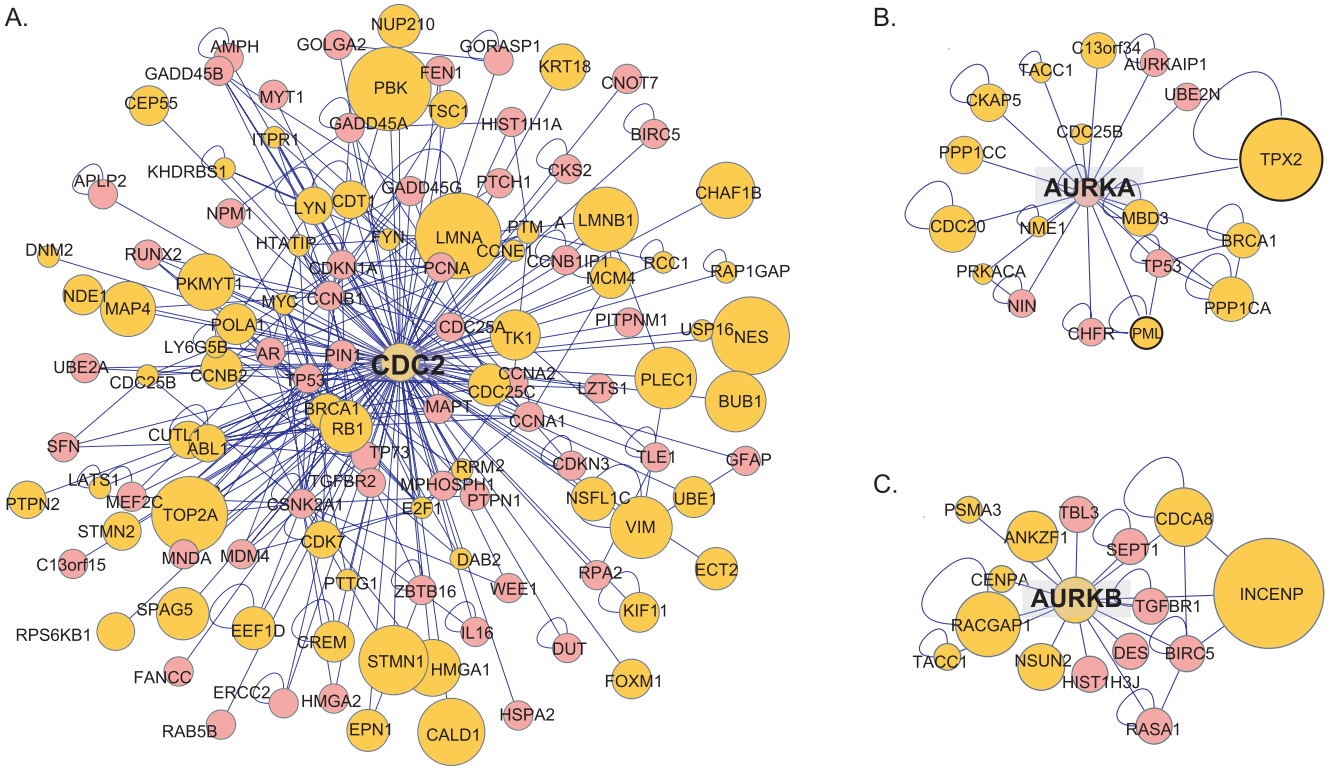

Fig. S4. Additional mitotic kinase interaction networks. Shown are the first-neighbor human protein interaction networks (41) for Cdc2 (*A*), Aurora kinase A (*B*), Aurora kinase B (*C*). Yellow nodes were found phosphorylated in our mitotic dataset. The size of each node corresponds to its mitotic P-RePh score, larger nodes are more heavily phosphorylated in mitosis.

Other Supporting Information Files

Table S1
Table S2
Table S3
Table S4
Table S5
Table S6
Table S7
Table S8