Cell, Volume *129*

Supplemental Data

Systematic Discovery of In Vivo Phosphorylation Networks

Rune Linding[1,2*†], Lars Juhl Jensen[3*], Gerard J. Ostheimer[2,4*], Marcel A.T.M. van Vugt[2,5], Claus Jørgensen[1], Ioana M. Miron[1], Francesca Diella[3], Karen Colwill[1], Lorne Taylor[1], Kelly Elder[1], Pavel Metalnikov[1], Vivian Nguyen[1], Adrian Pasculescu[1], Jing Jin[1], Jin Gyoon Park[1], Leona D. Samson[4], James R. Woodgett[1], Robert B. Russell[3], Peer Bork[3,6†], Michael B. Yaffe[2†] and Tony Pawson[1†]

[1] Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Canada.

[2] Center for Cancer Research, Massachusetts Institute of Technology, Cambridge, USA

[3] European Molecular Biology Laboratory, Heidelberg, Germany

[4] Center for Environmental Health Sciences, Massachusetts Institute of Technology, Cambridge, USA

[5] Department of Cell Biology and Genetics, Erasmus University, Rotterdam, The Netherlands

[6] Max-Delbrück-Centre for Molecular Medicine, Berlin, Germany

[*] These authors contributed equally to this work

[†] To whom correspondence should be addressed; E-mail: linding@mshri.on.ca, bork@embl.de, yaffe@mit.edu or pawson@mshri.on.ca

Kinase family prediction

To maximise the coverage of kinases, we combined two existing methods for predicting the kinase families responsible for the phosphorylation of a particular site: the position specific scoring matrices from Scansite[37] (`http://scansite.mit.edu`) and the artificial neural networks from NetPhosK[21] (`http://www.cbs.dtu.dk/services/NetPhosK`).

For each kinase family, we tested the sequence-based prediction methods available from NetPhosK[21] and Scansite[37]. To benchmark these methods, we applied each method to all phosphorylation sites in the Phospho.ELM database, sorted the predictions by score, and plotted the fraction of known sites identified for the family in question as a function of the total number of sites suggested for that family (Figure S1). Based on the resulting performance curves, we selected the best consensus sequence motif for each protein family. In cases where only one predictor was available it was used.

To speed up calculation, we re-implemented the NetphosK method as a single ANSI-C program. It provides a score between 0 and 1 for each of the following kinase families: PIKK(ATM, DNA-PK), CK-I, CK-II, CaM-II, GSK3, PKA, PKB, PKC, PKG, RSK, CDK, p38 MAPK, EGFR, INSR, and SRC. Similarly, Scansite provides a score between 0 and 1 for the kinase families: Abl, Clk, Erk, DMPK and PDGFR. We predict each site to be a target for all kinase families that receive a NetPhosK score better than 0.5 or predicted at "high confidence" according to Scansite, corresponding to the optimal cutoffs for these methods.

Within a proteome, kinases of each family are identified by sequence similarity searches. 82 representative kinase domain sequences, which have been manually assigned to families, are searched against the proteome using BLASTP[2]. Only hits with an E-value better than 10^{-40} and with at least 50% sequence identity are considered; among these, the hit with the best bit score is used for assigning a kinase family.

Detection of Rad50 phosphorylation

EBV-transformed lymphoblasts (National Institute of General Medical Sciences Human Genetic Mutant Cell Repository, Camden, NJ) from a normal ($ATM^{wt/wt}$) individual (GM02254) and from an individual with Ataxia-Telangectasia (GMO1526, $ATM^{-/-}$) were maintained in RPMI supplemented with 15% FCS (Life Technologies, Inc., Grand Island, NY), 1% L-Glutamine, 1% penicillin and 1% streptomycin. Cells were treated with 10 μM doxorubicin for 8 hours and then harvested. 3x10^7 cells were resuspended in a 10 mL lysis buffer containing 1% Triton X-100, 50 mM Tris pH 7.5, 150 mM sodium chloride, 50 mM β-glycerophosphate, 10 mM sodium pyrophosphate, 30 mM sodium flouride, 1 mM benzamidine, 2 mM EDTA, 2 mM magnesium chloride, 1 mM DTT, 1 mM AEBSF, 1 mM PMSF, 100 μM sodium orthovanadate, 20 μM Leupeptin, 15 μM E-64, 10 μM Pepstatin, 0.8 μM Bestatin, 1 μg/ml Microcystin LR. Cells were lysed by sonication and nucleic acids digested by the addition of 25 units of benzonase (Novagen). Lysates were cleared by centrifugation for 10 minutes at 19,000 g. All lysates were first subjected to a mock immunoprecipitation using 4 μg of mouse IgG (Santa Cruz, sc-2025). Subsequently, Rad50 was immunoprecipitated with 4 μg of a mouse monoclonal antibody (Abcam, ab89). Immunoprecipitations were separated by 8% SDS-PAGE and visualised by immunoblotting. Blots were probed with antibodies to Rad50 (Abcam, ab89), Nbs1 (Abcam, ab7860), Mre11 (Abcam, ab214) and an antibody that recognises the phospho-SQ/TQ motif of the ATM and ATR kinases (Cell Signaling Technology, 2851). Immunoblots were visualised using a goat anti-mouse antibody conjugated to IRDye 700DX (Rockland) and a goat anti-rabbit antibody conjugated to IRDye 800 (Rockland) and the Li-Cor Odyssey Infrared Imaging System. Immunoprecipitation prior to phospho-peptide mapping by mass spectroscopy was performed exactly as described above except that 25x10^7 cells were harvested for immunoprecipitation. The immunoprecipitation was separated by 8% SDS-PAGE and the Rad50 band cut from the gel and submitted to phospho-peptide mapping by mass spectrometry.

Excised gel bands containing Rad50 protein were digested with Trypsin (Promega, Madison, WI) according to the protocol described in[22]. After lyophilisation, tryptic peptides were analysed by liquid chromatography-mass spectrometry consisting of a nanoflow HP1100 HPLC system (Agilent, Palo Alto, CA) and an LTQ mass spectrometer (Thermo Electron, San Jose, CA). Peptides were separated on a custom-made 75 μm I.D. PicoTip column packed with Pursuit 3 μm C-18 beads (Varian, Palo Alto, CA). The column effluent was sprayed directly into the interface of the Mass Spectrometer. The gradient used for separation was 3–60 % of Acetonitrile for 180 minutes. Raw mass spectrometric data were screened against the NCBI Data Base (NCBInr version 20060227) using the Mascot Search Engine (version 2.1, Matrixscience, London, UK). Ser/Thr phosphorylation was set as a variable modification in Mascot Search Settings. Examination of the MS/MS fragmentation spectra from Rad50 confirmed that S635 (predicted ATM site) and T690 (predicted CK2 site) on Rad50 are phosphorylated (Figure S7), in agreement with previous data[6].

Detection of 53BP1 phosphorylation

The human osteosarcoma cell line U2OS was maintained in DMEM, supplemented with 10% fetal calf serum and penicillin/streptomycin. To activate the DNA damage checkpoint, cells were treated with 0.5 μM doxorubicin (Sigma) for 1 hour and subsequently incubated for 16 hours in the presence of 1 μM paclitaxel (Sigma). Mitotic cells were detached using a gentle shake-off after treatment with 1 μM paclitaxel

for 16 hours. If indicated (Figure 6D) those mitotic cells were replated in the presence of 20 μM roscovitine (Sigma). Cells were lysed in buffer containing 50 mM Tris pH 7.4, 1% Triton X100, 5 mM NaF, 20 mM $NaVO_3$ supplemented with Complete protease inhibitor cocktail (Roche). 250 μg of cleared lysate was used for 53BP1 immunoprecipitations using Protein-A sepharose beads (Pharmacia). Fluorescent cell sorting was performed on cells fixed in ice cold 70% ethanol for 1 hour. Subsequently, mitotic cells were stained using anti-Histone H3 (phosho-ser10) and Alexa 647-conjugated anti-rabbit and treated with propidium iodide and RNase. We analysed 10^4 events on a Becton Dickinson FACScalibur and processed data using CellQuest software. Immunoblotting and immunoprecipitations were performed using the following antibodies: Rabbit Anti-cyclin A(H432) and mouse anti-cyclin B1 (GNS1) (SantaCruz Biotechnology), rabbit anti-53BP1 (Novus Biologicals), mouse anti-β-Actin(Sigma), mouse anti-MPM2 and rabbit anti-Histone H3 (phosho-ser10) (Upstate Biotechnology) and Alexa 647-conjugated anti-rabbit Molecular Probes (Invitrogen).

Preparation of BCLAF1 samples

Human embryonic kidney (HEK) 293 cells were cultured in DMEM supplemented with 10% fetal bovine serum. Cells were treated with 20 mM LiCl or KCl (control) for 2 hours and then harvested and resuspended in a lysis buffer containing 1% Triton X-100, 50 mM Tris pH 7.5, 150 mM NaCl, 50 mM β-glycerolphosphate, 10 mM sodium pyrophosphate, 30 mM NaF, 2 mM EDTA, 2 mM $MgCl_2$, 1 mM DTT, 100 uM sodium orthovanadate, phosphatase inhibitor cocktail 1 (Sigma), and mini complete EDTA-free protease inhibitor cocktail (Roche). Lysates were passed through a 22.5 gauge needle and were cleared by centrifugation for 10 minutes at 40,000 g. The supernatant was precleared with GammaBind Plus Sepharose (GE Healthcare) before BCLAF1 was immunoprecipitated with a rabbit anti-BCLAF1 antibody (Bethyl Laboratories, A300-608A). Immunoprecipitates were washed and boiled in reducing gel sample buffer prior to separation on an 8-16% SDS-PAGE gel. Excised gel bands containing BCLAF1 protein were digested with Trypsin (Promega, Madison, WI) according to the protocol described in[22]. The tryptic peptides were analysed by liquid chromatography-mass spectrometry.

Curation of mass spectrometry datasets

By mapping sequences to full-length polypeptides and adjusting residue numbers, we curated eight large-scale phosphoproteome data sets from five publications: 1) two Jurkat cell-line sets consisting of 67 sites in 31 proteins[42] and 195 sites in 105 proteins, respectively[8], 2) a larger set of 952 sites in 435 proteins from HeLa cells[6], 3) data from Rush et al. with 325 sites in 237 human proteins and 171 sites (in 118 mouse proteins)[40], 4) a set of 221 sites in 114 mouse brain proteins and 269 sites in 75 synaptic mouse proteins[3;12]. These data have been deposited in the Phospho.ELM database[15].

Network topology analysis

Cytoscape-2.3.1[43] with the NetworkAnalyzer plugin (`http://med.bioinf.mpi-inf.mpg.de/netanalyzer/index.php`) was used to calculate the topological parameters of three networks: 1) A predicted human phosphorylation network (HPN) which is based on the complete dataset in Phospho.ELM composed of 5189 edges and 1810 nodes (first row of Figure S5). The network was reduced to kinase/substrate interaction level, discarding the site specific information. Thus if a kinase phosphorylates a substrate at multiple sites this would be represented by one edge (predicted interaction) only. 2) An experimentally determined human protein-protein interaction (PPI) network (second row in Figure S5), extracted from STRING, composed of protein pairs with a physical interaction score $\geq$ 0.7, which encompassed 6305 nodes spanning 21828 physical interactions. 3) A human context network (third row in Figure S5) corresponding to all protein pairs in STRING with a functional association score $\geq$ 0.9, in total 6338 nodes 43708 edges. The networks were treated as undirected and can be downloaded as Cytoscape files at `http://networkin.info`.

The degree distribution of the networks predicted by NetworKIN (Figure S5b) approximately follows a power law, indicating scale-free network properties as observed for many other networks. The shortest path distribution was downshifted in our networks with a median of 3 vs. 4 for the PPI and context networks (Figure S5a). The HPN had a slightly lower number of neighbours per node than the PPI network (5.7 versus 6.9) but only half of what was seen in the context network (13.8). The number of substrates per kinase is shown in Figure S4. To measure overall topology we calculated the clustering coefficient ($C(k)$, Figure S5c) which measures the tendency of nodes to form clusters in the network. $C(k)$ decreases as the number of interactions per node increase, which suggests hierarchical organisation of the HPN and PPI networks. The fitted power-law functions were (Figure S5c): $C_{HPN}(k) = 2.1079k^{-1.1516}$ (R^2=0.6881), $C_{PPI}(k) = 0.9125k^{-0.5031}$ (R^2=0.2669) and $C_{context}(k) = 0.4473k^{0.0105}$ (R^2=0.0003). Thus the context network does not show hierarchical organisation. The topological coefficient $TC(k)$ (Figure S5d) is a measure of the extent to which proteins share interaction partners in the network. As this is also a decreasing function of the number of edges it suggests that hub proteins do not have more common neighbours than less connected proteins. Again, a higher similarity between the HPN and PPI network is observed than between the HPN and context network (the function is much more dispersed in the context network).

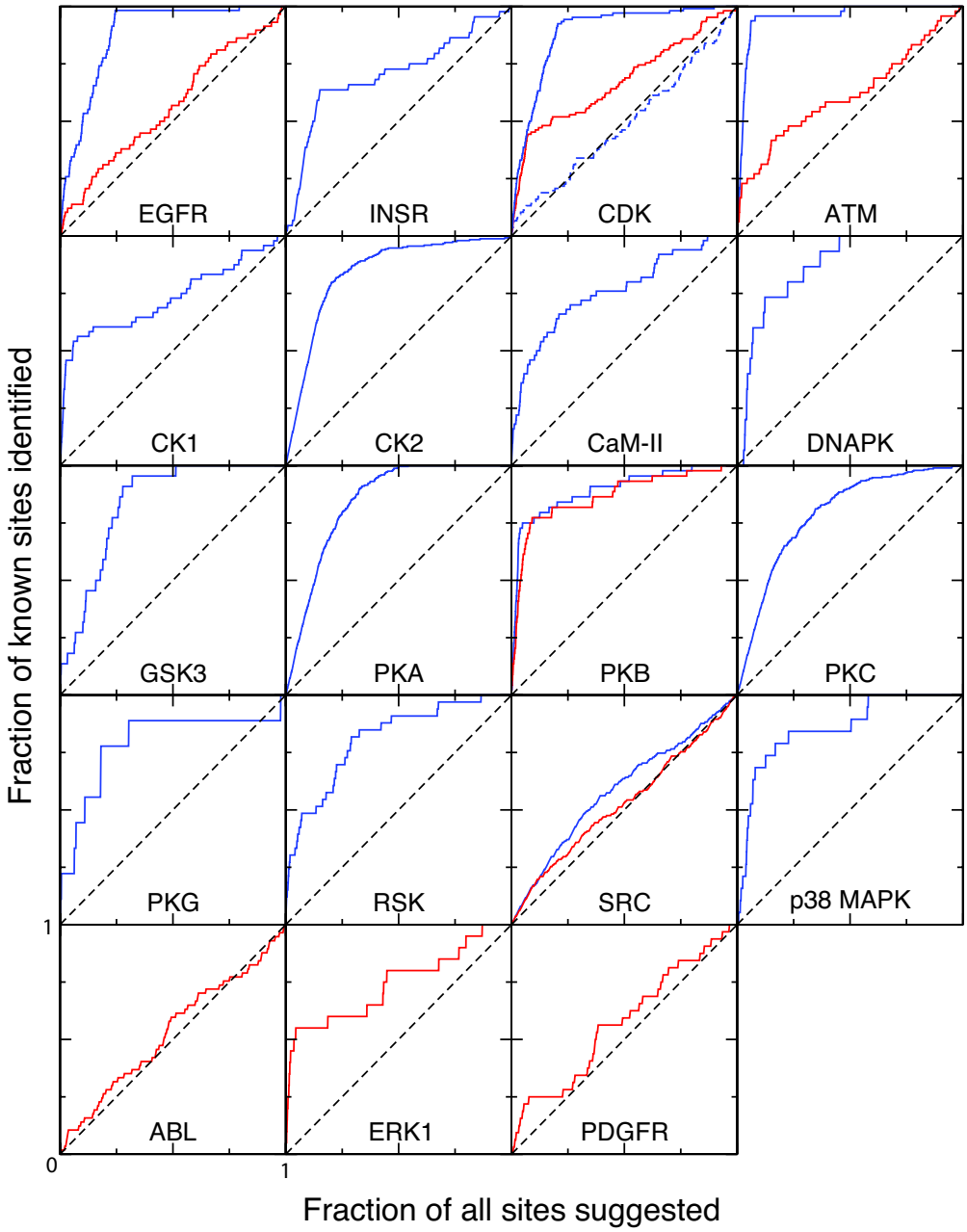


Figure S1: **Performance curves for motif predictors.** To evaluate and compare the ability of NetPhosK (blue) and Scansite (red) to predict the kinase family for known phosphorylation sites, we applied the predictors to sites from the Phospho.ELM database[15]. The resulting predictions were sorted by score, and the fraction of known sites identified was plotted as a function of the fraction of total number of sites suggested for each protein family. An optimal predictor would show a sharp bend in the upper left corner, meaning all correct sites were predicted in as few suggestions as possible. The dashed black lines show the performance that would be obtained through random guessing.

		Motif-based methods		NetworKIN	
	Number of sites (TP+FN)	Total family predictions (TP+FP)	Correct family predictions (TP)	Total family predictions (TP+FP)	Correct family predictions (TP)
CDK	122	337	63	105	49
PKC	26	43	12	15	8
PIKK	88	200	67	74	60
INSR	46	113	31	39	31
Total	282	693	173	233	148

Table S1: **Benchmark statistics for kinase family identification.** Proteins containing the indicated number of experimentally verified CDK (Cyclin-dependent kinase), PKC (Protein kinase C: α, δ, ϵ, μ, θ and ζ isoforms), and PIKK (ATM, ATR or DNA-PK) sites were investigated using purely motif-based prediction methods (NetphosK and Scansite) or the context based NetworKIN algorithm. Prediction accuracy is defined as $TP/(TP + FP)$, where *TP* denotes true positives and *FP* false positives. Thus, this is the fraction of predictions known to be correct. Sensitivity is defined as $TP/(TP + FN)$, where *FN* denotes false negatives. Thus, this is the fraction of known sites that are correctly predicted. Using only consensus motifs we obtained prediction accuracies of 19% (CDK), 28% (PKC), 34% (PIKK) and 27% (INSR) and sensitivities of 52% (CDK), 46% (PKC), 76% (PIKK) and 67% (INSR). By including contextual information the prediction accuracies more than doubled to 47% (CDK), 53% (PKC), 81% (PIKK) and 79% (INSR), with only a small drop in sensitivities to 40% (CDK), 31% (PKC), 68%(PIKK) and 67% (INSR). Notably, the accuracy of NetworKIN predictions is likely to be an underestimate since not all the kinases that target each phosphorylation site in the set of test proteins may currently be known from experiments.

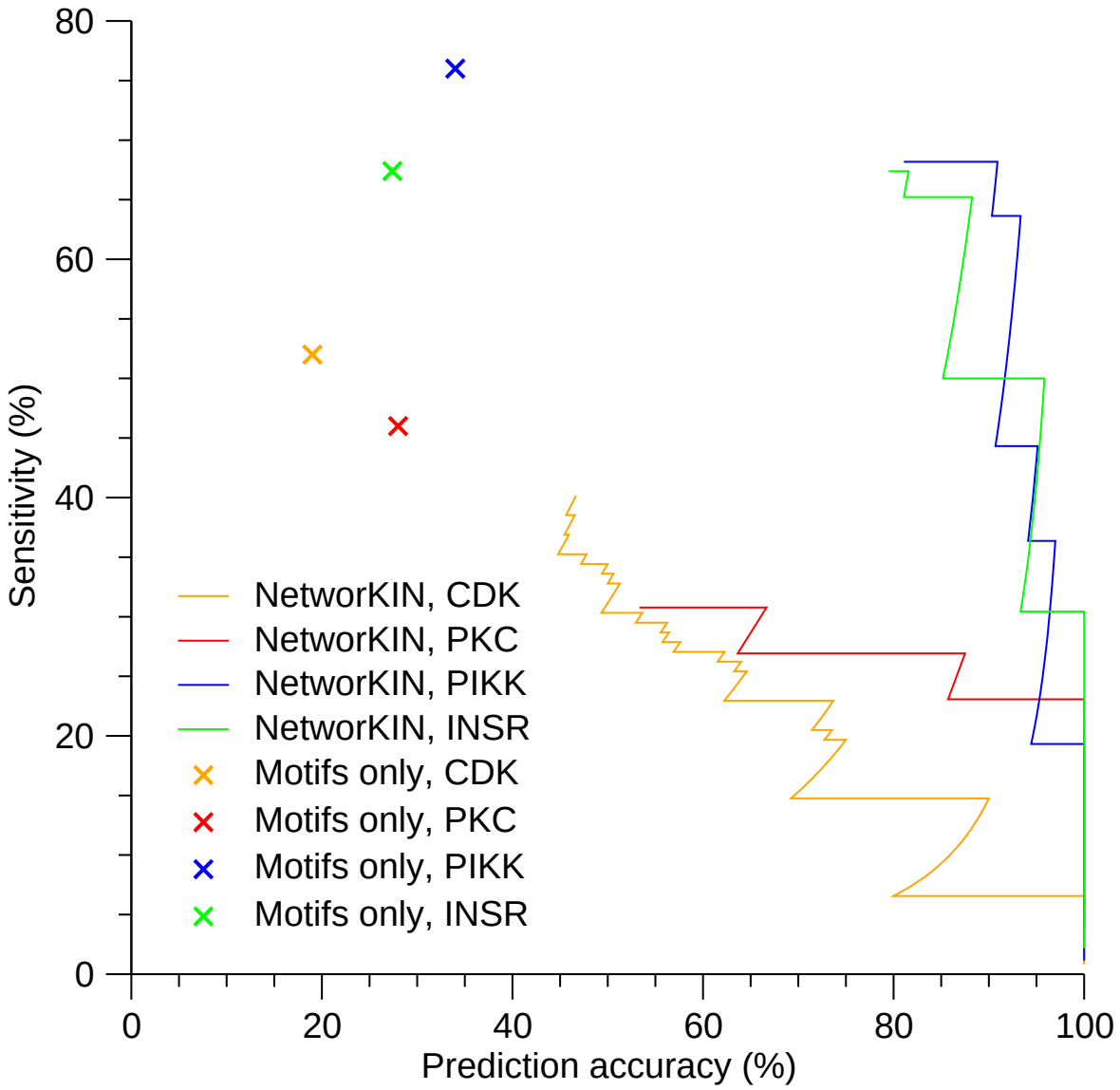


Figure S2: **Performance curves for kinase family identification.** In Figure 2, the performance of NetworKIN was reported when always trusting the prediction with the highest context score for each phosphorylation site. Here we assessed how the performance varies when imposing a lower limit on the context scores. Since the sites used for benchmarking have necessarily been described in the literature, the use of literature mining could in theory lead to overestimation of the accuracy and sensitivity; however, excluding the abstracts that mention the phosphorylation events in question hardly affects the performance (data not shown).

	Predicted	Not predicted	Fraction
Cdc28p substrates	281	59	83%
Other proteins	2522	3818	40%

Table S2: **Comparison of NetworKIN predictions with high-throughput screens for Cdc28p (Cdk1) substrates in *S. cerevisiae*[33;48].**. Since the screens in question only identified Cdc28p substrates but not the actual phosphorylation sites, the NetworKIN method was applied to all serines and threonines in the yeast proteome. Proteins that contain one or more sites, which we predict to be phosphorylated by Cdc28p, were defined as predicted Cdc28p substrates. Comparing these to the set of substrates identified in the two screens, shows that 83% of the known substrates were correctly predicted, corresponding to a 2-fold enrichment of known substrates over random among the predicted substrates. This enrichment is statistically highly significant ($P < 10^{-54}$, Fisher's exact test).

	Entire network	High-confidence associations	High-confidence kinase associations
Genomic context	25.5%	0.7%	0.4%
Experimental data	8.9%	21.5%	30.3%
Curated pathways	12.2%	36.5%	27.4%
Literature mining	54.1%	41.2%	41.9%

Table S3: **Contribution of different evidence types to the context network.** Quantifying the contribution of different evidence types to the context networks is non-trivial because the network is probabilistic. If one simply sums the probability contributions of each type of evidence across all protein pairs in the complete interaction network, text mining provides over half of the total evidence, and genomic context methods provide over a quarter of the evidence. However, this does not show the complete picture since some evidence types give rise to few interactions with high confidence, whereas others give vast numbers of interactions but with low confidence. Since the predictions by NetworKIN mostly rely on interactions of fairly high confidence, we repeated the analysis but considered only interactions with a STRING confidence score of 70%. This increases the importance of experimental evidence and curated pathway databases, decreases the reliance on text mining and almost eliminates genomic context. This also may not accurately reflect the impact of the individual evidence types on NetworKIN, since some type of evidence is more important for some classes of proteins than for others. We thus further restricted the analysis to consider only associations involving at least one of the 112 protein kinases currently included in NetworKIN. Doing so leads to a considerable increase in the estimated importance of experimental data and a corresponding decrease in the importance of manually curated pathway databases. We believe that this last column in the table best reflects the relative importance of each evidence type to the predictions made by NetworKIN.

Kinase	Cytoplasmic substrates	Nuclear substrates	P-value
INSR	**45**	4	$2 \cdot 10^{-12}$
LCK	**28**	0	$3 \cdot 10^{-11}$
CDK2	56	**161**	$3 \cdot 10^{-7}$
IGF1R	**27**	5	$9 \cdot 10^{-7}$
SRC	**16**	0	$9 \cdot 10^{-7}$
EGFR	**15**	0	$3 \cdot 10^{-6}$
LYN	**14**	1	$5 \cdot 10^{-5}$
FYN	**11**	1	$6 \cdot 10^{-4}$
PKCα	**127**	118	$6 \cdot 10^{-4}$
CK1ϵ	43	**104**	$9 \cdot 10^{-4}$
CK2α	96	**183**	$4 \cdot 10^{-3}$
MAPK9	23	**61**	$4 \cdot 10^{-3}$
MAPK10	26	**66**	$5 \cdot 10^{-3}$
HCK	**8**	1	$6 \cdot 10^{-3}$
CDC2	51	**108**	$6 \cdot 10^{-3}$
PKAβ	**67**	59	$6 \cdot 10^{-3}$
ATM	21	**54**	$9 \cdot 10^{-3}$
Total	960	1327	

Table S4: **Kinases predicted to preferentially target cytoplasmic/nuclear substrates.** SwissProt and Phospho.ELM were used to compile two sets of phosphoproteins that are believed to localize exclusively to the cytoplasm and the nucleus, respectively. NetworKIN was used to predict the kinases responsible for the phosphorylation of these proteins, and Fisher's exact test was used to identify 17 kinases that we predict to preferentially target either cytoplasmic or nuclear proteins. For another 18 kinases, no statistically significant difference was observed; because of these additional kinases, the total number of predicted cytoplasmic/nuclear substrates is greater than the sum of each column.

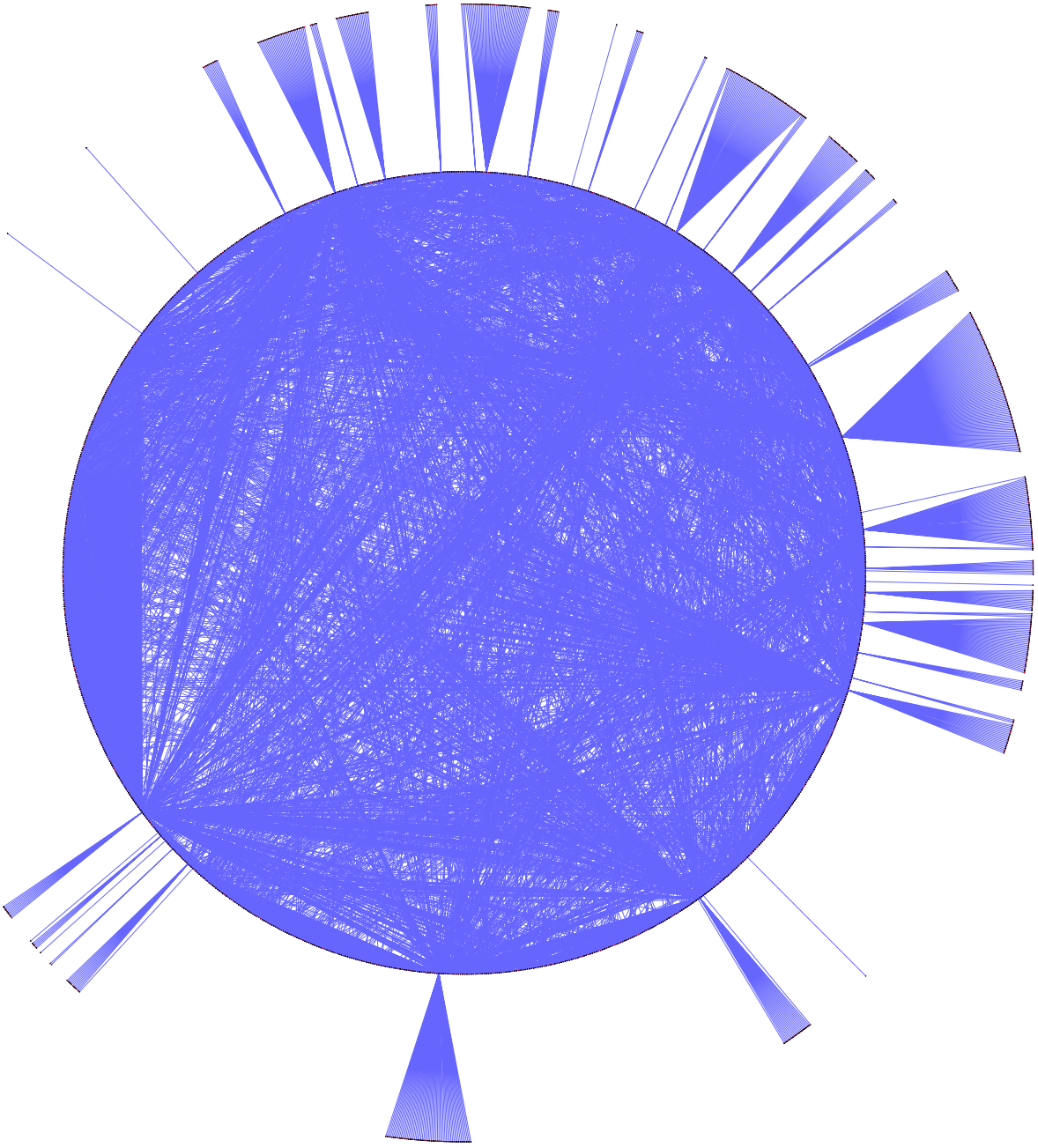


Figure S3: **Map of the human phosphorylation network.** Circular network graph of the complete HPN (best viewed by zooming in a PDF viewer). Blue circles and red boxes denote kinases and substrates respectively. An edge corresponds to a predicted kinase–substrate interaction (non site-specific). The network is composed of 4723 edges and 1777 nodes (of which 68 are kinases). The typical shortest path between any pair of proteins in the network is 3 and each protein has an average of 5.7 neighbours.

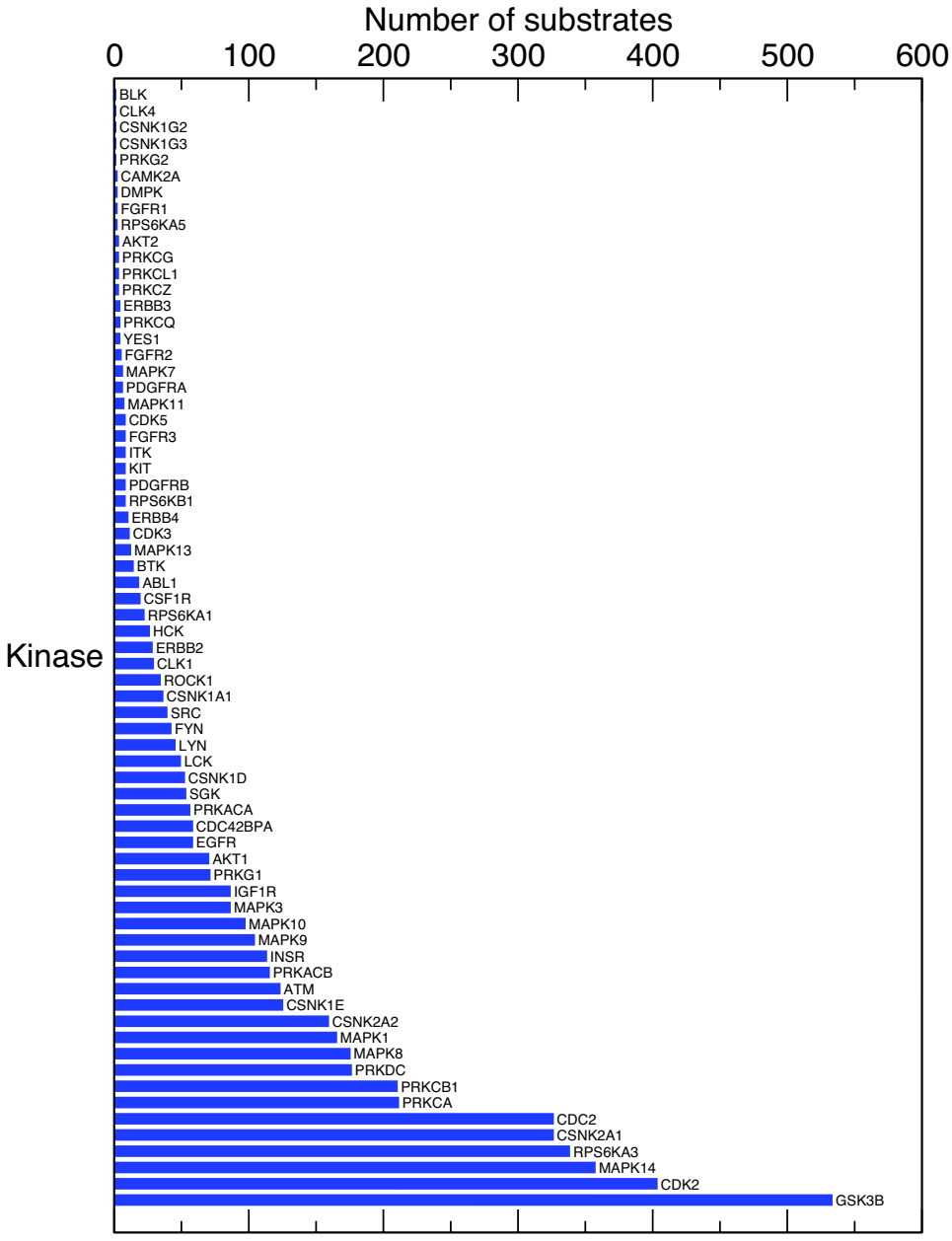


Figure S4: **Substrate distribution for human kinases.** The number of predicted substrates for the 68 kinases we have predictions for in the human phosphoproteome. The median number of substrates per kinase is 12 and the average is 70.

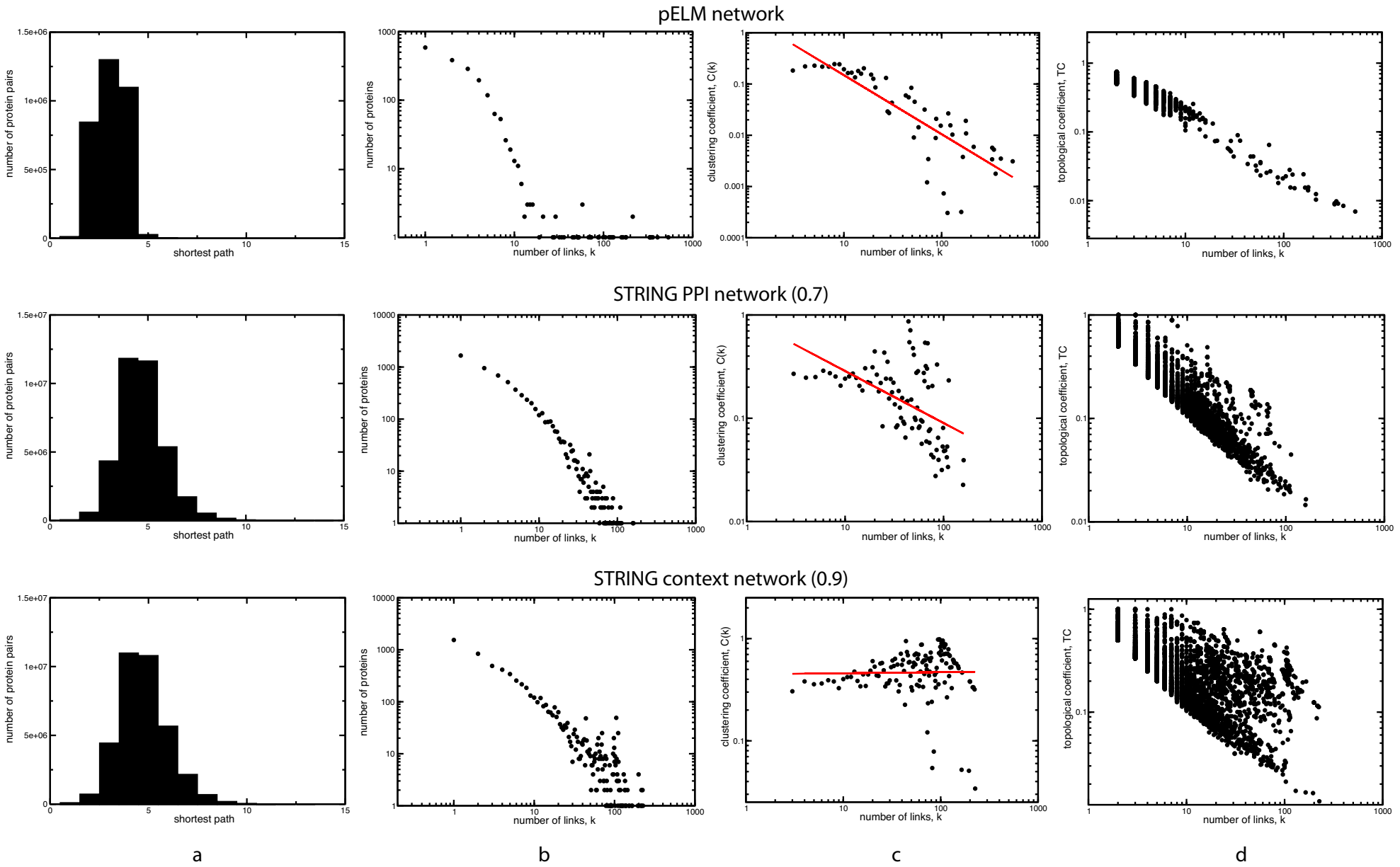


Figure S5: **Topology plots.** Four topological measures were determined (by columns): **a,** The distribution of the shortest path between pairs of proteins. **b,** The node degree distribution, where the degree of a node x is the number of links incident to x. **c,** The average clustering coefficient, $C(k)$, of all nodes x with $k \geq 3$ neighbours is plotted against the number of neighbours. Where the $C(k) = 2n/k_x(k_x\text{-}1)$, with n as the number of links connecting the k_x neighbours of node x to each other. $C(k)$ is a measure of the tendency of proteins in a network to form clusters or groups. We fitted the $C(k)$ data to power-law functions indicated by red lines. **d,** The topological coefficient, $TC(k)$, was calculated for every protein in the network and plotted against the number of links, where $TC_x(k)$ = average$(J(x,j)/k_x)$, where $J(x,j)$ denotes the number of nodes to which both x and j are linked, k_x is the number of links of node x). $TC(k)$ is a relative measure of the extent to which proteins share interaction partners with other proteins.

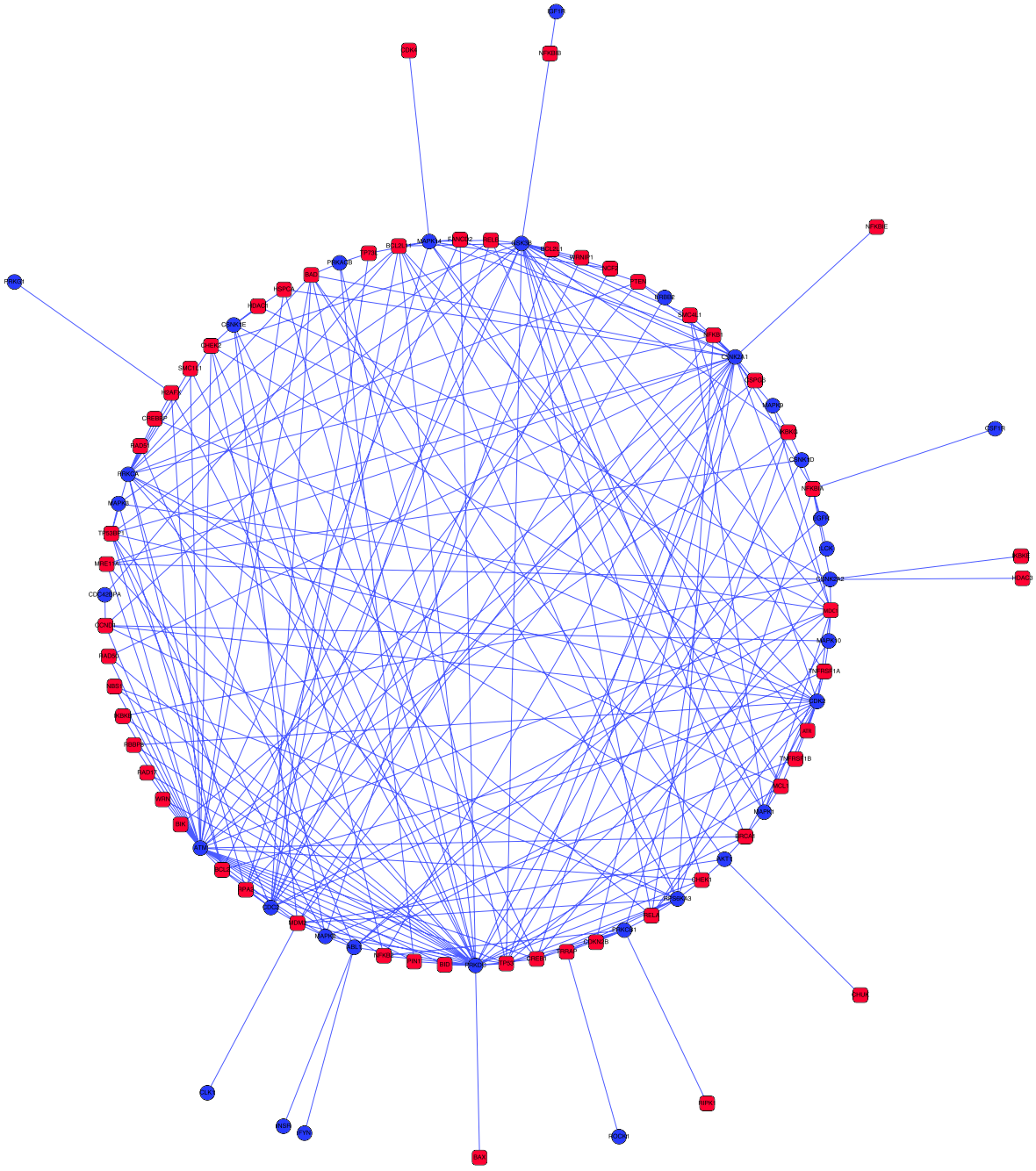
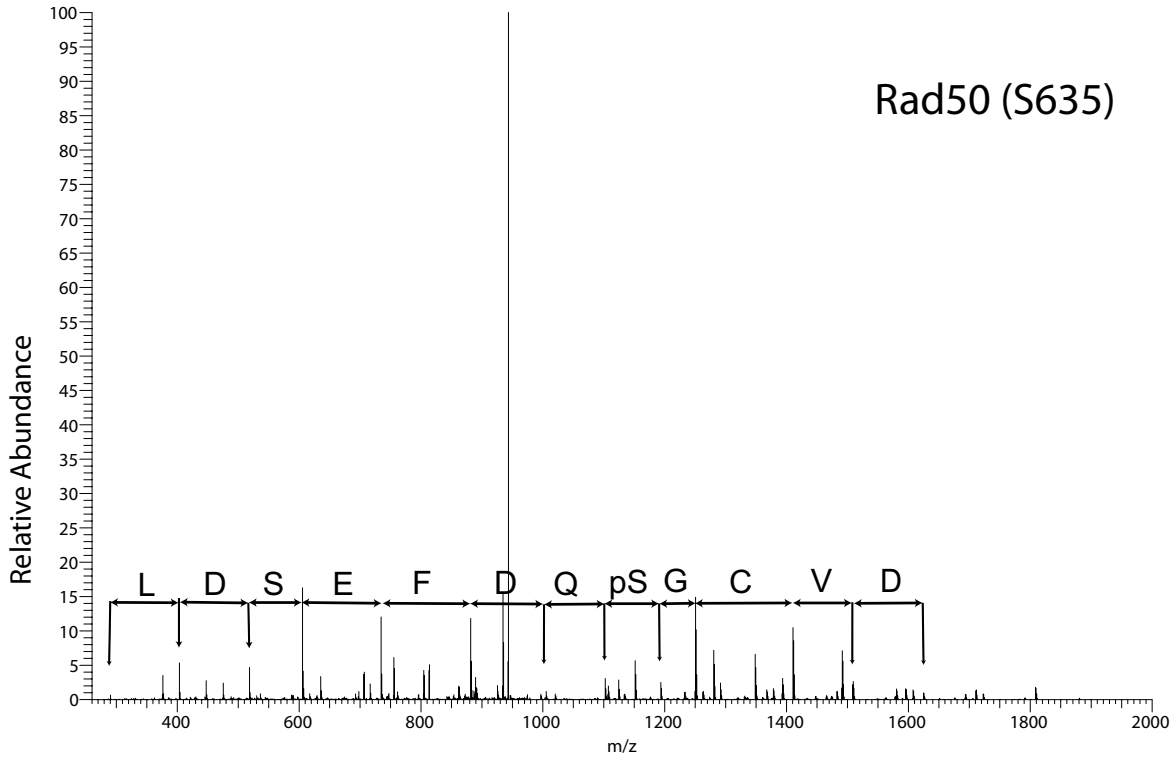
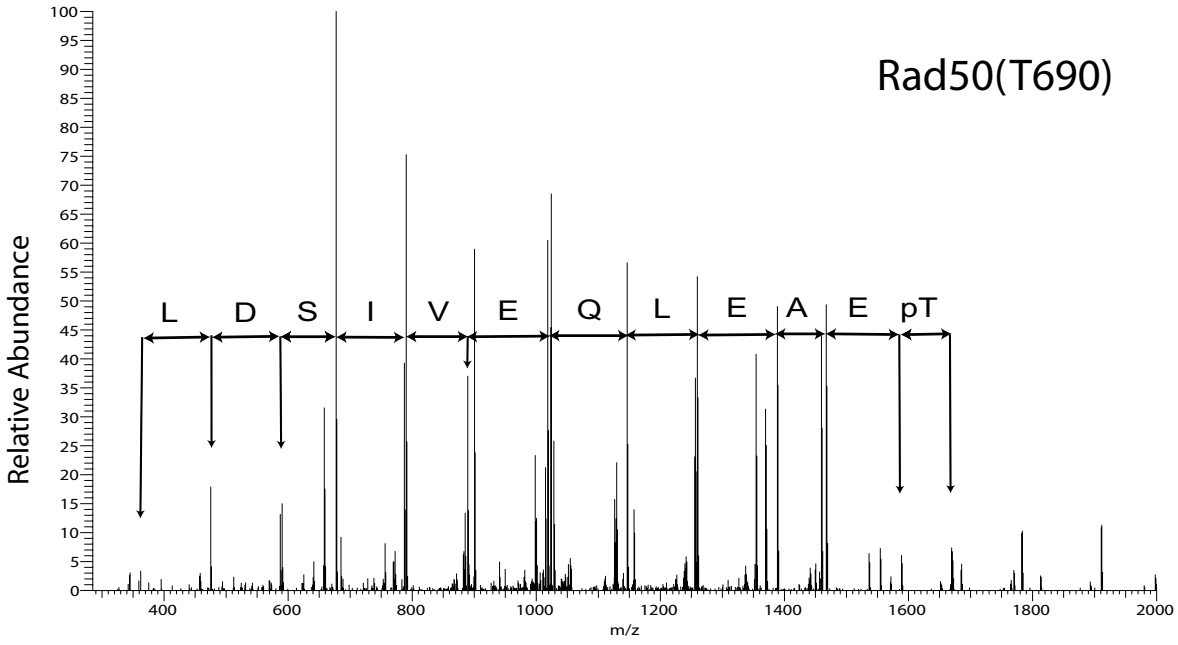


Figure S6: **Map of the DDR+ subnetwork.** Circular network graph of the complete DDR+ subnetwork (best viewed by zooming in a PDF viewer). The subnetwork is composed of selected proteins from DNA damage response, apoptosis and NFκB signalling and spans 84 nodes and 235 edges. Blue circles and red boxes denote kinases and substrates respectively. An edge corresponds to a predicted kinase–substrate interaction. The typical shortest path between any pair of proteins in the network is 3 and the average number of neighbours is 5.6, topology-wise this network is similar to the HPN (data not shown).



a



b

Figure S7: **MS/MS spectra for two phosphorylated peptides from human Rad50. a,** Ion product spectrum of the peptide LFD-VCG**pS**QDFESDLDR derived from human Rad50. Overlaid on top of the spectrum is the Y ion sequence showing a phospho-serine at S635 (the predicted ATM site). **b,** Ion product spectrum of the peptide VFQ**pT**EAELQEVISDLQSK from human Rad50. The Y ion sequence showing phospho-threonine at T690 (the predicted CK2AK2 site) is mapped above the spectrum.

Table S5: **Annotated ATM predictions.** ATM kinase predictions annotated with supportive evidence from the literature. In total 98 proteins were predicted ATM targets (on 132 sites). We annotated 113 of these predictions and found 45 substrates were novel targets. We provide a complete list of all predictions for further experimentation on the supplemental website (http://networkin.mshri.on.ca) and as supplemental data.

Substrate	HUGO	Residue	Motif score	Context score	Function	Known kinases
Abl	ABL1	446	0.65	0.999	DNA damage response	
Abl	ABL1	465	0.65	0.999	DNA damage response	ATM [5]
APC5	ANAPC5	217	0.63	0.948	Cell cycle control	
BCL-2	BCL2	24	0.51	0.998	Apoptosis	
BID	BID	78	0.60	0.997	Apoptosis	ATM [23;56]
BAD	BAD	167	0.59	0.997	Apoptosis	
BRCA1	BRCA1	1387	0.65	0.999	DNA damage response	ATM [52]
BRCA1	BRCA1	1423	0.59	0.999	DNA damage response	ATM [13]
BRCA1	BRCA1	1457	0.67	0.999	DNA damage response	ATM [13]
BRCA1	BRCA1	1466	0.60	0.999	DNA damage response	ATR [47]
BRCA1	BRCA1	1524	0.58	0.999	DNA damage response	ATM [13]
Cdc25A	CDC25A	279	0.51	0.998	Cell cycle control	Chk1 [45]
Chk1	CHEK1	317	0.51	0.999	DNA damage response	ATM [17]
Chk1	CHEK1	345	0.57	0.999	DNA damage response	ATR [32;55]
Chk2	CHEK2	19	0.55	0.999	DNA damage response	ATM [34]
Chk2	CHEK2	26	0.61	0.999	DNA damage response	ATM [34]
Chk2	CHEK2	28	0.66	0.999	DNA damage response	ATM [34]
Chk2	CHEK2	33	0.57	0.999	DNA damage response	ATM [34]
Chk2	CHEK2	35	0.62	0.999	DNA damage response	ATM [34]
Chk2	CHEK2	68	0.61	0.999	DNA damage response	ATM [1;34;36]
CREB	CREB1	111	0.59	0.992	Transcription factor	ATM [44]
CREB	CREB1	121	0.55	0.992	Transcription factor	ATM [44]
ATF-2	CREBP1	472	0.54	0.987	Transcription factor	ATM [7]
ATF-2	CREBP1	480	0.55	0.987	Transcription factor	ATM [7]
SMC3	CSPG6	1065	0.48	0.997	DNA damage response	
SMC3	CSPG6	1067	0.57	0.997	DNA damage response	
SMC3	CSPG6	1074	0.47	0.997	DNA damage response	
Beta-Catenin	CTNNB1	112	0.52	0.994	Transcription factor	
E2F1	E2F1	31	0.67	0.997	Cell cycle control	ATM [31]
PHAS-1	EIF4EBP1	111	0.62	0.996	Translation	ATM [53]
DAL-1	EPB41L3	88	0.54	0.940	Apoptosis	
DAL-1	EPB41L3	96	0.53	0.940	Apoptosis	
DAL-1	EPB41L3	460	0.57	0.940	Apoptosis	
DAL-1	EPB41L3	486	0.57	0.940	Apoptosis	
ERBB2IP	ERBB2IP	1271	0.63	0.995	Cell adhesion	
EPN2	EPN2	173	0.606	0.629	Endocytosis	
FANCD2	FANCD2	222	0.68	0.999	DNA damage response	ATM [46]
FANCD2	FANCD2	1257	0.65	0.999	DNA damage response	
FANCD2	FANCD2	1404	0.57	0.999	DNA damage response	
γH2AX	H2AFX	139	0.67	0.999	DNA damage response	ATM [9]
HMGA1	HMGA1	98	0.63	0.835	Transcription factor	
HSP90	HSPCA	6	0.62	0.998	Heat shock protein	
IKKβ	IKBKB	682	0.60	0.995	Apoptosis	
IKKγ/NEMO	IKBKG	85	0.51	0.995	Apoptosis	ATM [51]

Substrate	HUGO	Residue	Motif score	Context score	Function	Known kinases
beta-2 Integrin	ITGB2	745	0.51	0.958	Cell adhesion	
Jun	JUN	249	0.65	0.998	Transcription factor	
EKLF	KLF1	23	0.61	0.963	Transcription factor	
DNA Ligase IV	LIG4	132	0.48	0.986	DNA damage response	
HSL	LIPE	660	0.60	0.994	Lipase	
MARCKS	MARCKS	313	0.52	0.943	Cytoskeleton	
Pur-1/MAZ	MAZ	469	0.67	0.927	Transcription factor	
MBP	MBP	141	0.52	0.996	Myelin basic protein	
MDC1	MDC1	988	0.58	0.900	DNA damage response	ATM[19]
MDM2	MDM2	262	0.49	0.999	Apoptosis	
MDM2	MDM2	395	0.48	0.999	Apoptosis	ATM[35]
MDM2	MDM2	407	0.50	0.999	Apoptosis	
Mre11	MRE11A	264	0.58	0.999	DNA damage response	ATM[25]
Nbs1	NBS1	278	0.66	0.999	DNA damage repsonse	ATM[14]
Nbs1	NBS1	343	0.67	0.999	DNA damage repsonse	ATM[18;29]
Nbs1	NBS1	397	0.57	0.999	DNA damage repsonse	ATM[50]
NCF1	NCF1	320	0.48	0.994	Neutrophil oxidase factor	
NFAT5	NFAT5	1197	0.56	0.963	Transcription factor	
NFAT5	NFAT5	1247	0.52	0.963	Transcription factor	
NFAT5	NFAT5	1367	0.52	0.963	Transcription factor	
Nrf2	NFE2L2	40	0.55	0.995	Transcription factor	
NF-kappaB2	NFKB2	869	0.52	0.971	Apoptosis	
NR3C1	NR3C1	508	0.61	0.997	Transcription factor	
PIN1	PIN1	65	0.51	0.997	Apoptosis	
PPARGC1A	PPARGC1A	263	0.61	0.994	Transcription factor	
PKCmu	PRKCM	249	0.58	0.946	Protein kinase	
DNA-PK	PRKDC	2056	0.51	0.996	DNA damage response	
DNA-PK	PRKDC	2612	0.53	0.996	DNA damage response	
DNA-PK	PRKDC	2638	0.51	0.996	DNA damage response	
DNA-PK	PRKDC	2647	0.56	0.996	DNA damage response	
PRL	PRL	163	0.49	0.996	Cytokine	
PS2	PSEN2	330	0.51	0.993	Cell cycle control	
Rad17	RAD17	646	0.65	0.996	DNA damage repsonse	ATM[4]
Rad17	RAD17	656	0.58	0.996	DNA damage response	ATM[39]
Rad50	RAD50	635	0.55	0.999	DNA damage response	
Rad9	RAD9A	272	0.65	0.996	Cell cycle control	ATM[11]
CtIP	RBBP8	664	0.65	0.997	DNA damage response	ATM[28]
CtIP	RBBP8	745	0.62	0.997	DNA damage response	ATM[28]
RFC1	RFC1	190	0.47	0.997	DNA damage response	
RIM1	RIMS1	218	0.51	0.951	Synaptic membrane reg.	
RIM1	RIMS1	1677	0.54	0.951	Synaptic membrane reg.	
RPS6	RPS6	247	0.58	0.975	Translation	p90RSK[38]
SF3B1	SF3B1	344	0.65	0.984	RNA splicing	
SFRS16	SFRS16	180	0.51	0.901	RNA splicing	
SMC1	SMC1L1	957	0.66	0.999	DNA damage response	ATM[26]
SMC1	SMC1L1	966	0.67	0.999	DNA damage response	ATM[54]
SRF	SRF	435	0.56	0.996	Transcription factor	
SRF	SRF	446	0.48	0.996	Transcription factor	
SRRM	SRRM1	465	0.48	0.970	Nuclear matrix protein	
StAR	STARD1	100	0.53	0.995	Transcription factor	
TEFB	TCEB3	225	0.51	0.994	Transcription factor	
HNF1	TCF1	249	0.64	0.995	Transcription factor	Mirk[30]

Substrate	HUGO	Residue	Motif score	Context score	Function	Known kinases
TRF1	TERF1	219	0.67	0.999	Cell cycle control	ATM[27]
THRAP3	THRAP3	211	0.47	0.972	Transcription factor	
THRAP3	THRAP3	784	0.55	0.972	Transcription factor	
p53	TP53	9	0.58	0.999	Apoptosis	
p53	TP53	15	0.88	0.999	Apoptosis	ATM[10;24]
p53	TP53	20	0.51	0.999	Apoptosis	Chk2[20] ATM[15]
p53	TP53	46	0.60	0.999	Apoptosis	ATM[41] HIPK2[16]
53BP1	TP53BP1	6	0.61	0.998	DNA damage response	ATM[49]
53BP1	TP53BP1	25	0.59	0.998	DNA damage response	ATM[49]
53BP1	TP53BP1	29	0.56	0.998	DNA damage response	ATM[49]
53BP1	TP53BP1	831	0.53	0.998	DNA damage response	
TPR	TPR	2036	0.60	0.997	Nuclear import	
TRX1	TREX1	68	0.60	0.998	DNA damage response	
TSC2	TSC2	1379	0.50	0.994	Apoptosis	
VIM	VIM	458	0.54	0.957	Cytoskeleton	
WRN	WRN	1141	0.68	0.998	DNA damage response	ATM[25]
WRN	WRN	1292	0.63	0.998	DNA damage response	

References

[1] Ahn, J. Y., Schwarz, J. K., Piwnica-Worms, H., and Canman, C. E. (2000). Threonine 68 phosphorylation by ataxia telangiectasia mutated is required for efficient activation of Chk2 in response to ionizing radiation. Cancer Res *60*, 5934–5936.

[2] Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. J Mol Biol *215*, 403–410.

[3] Ballif, B. A., Villén, J., Beausoleil, S. A., Schwartz, D., and Gygi, S. P. (2004). Phosphoproteomic analysis of the developing mouse brain. Mol Cell Proteomics *3*, 1093–101.

[4] Bao, S., Tibbetts, R. S., Brumbaugh, K. M., Fang, Y., Richardson, D. A., Ali, A., Chen, S. M., Abraham, R. T., and Wang, X. F. (2001). ATR/ATM-mediated phosphorylation of human Rad17 is required for genotoxic stress responses. Nature *411*, 969–974.

[5] Baskaran, R., Wood, L. D., Whitaker, L. L., Canman, C. E., Morgan, S. E., Xu, Y., Barlow, C., Baltimore, D., Wynshaw-Boris, A., Kastan, M. B., et al. (1997). Ataxia telangiectasia mutant protein activates c-Abl tyrosine kinase in response to ionizing radiation. Nature *387*, 516–519.

[6] Beausoleil, S. A., Jedrychowski, M., Schwartz, D., Elias, J. E., Villén, J., Li, J., Cohn, M. A., Cantley, L. C., and Gygi, S. P. (2004). Large-scale characterization of HeLa cell nuclear phosphoproteins. Proc Natl Acad Sci U S A *101*, 12130–5.

[7] Bhoumik, A., Takahashi, S., Breitweiser, W., Shiloh, Y., Jones, N., and Ronai, Z. (2005). ATM-dependent phosphorylation of ATF2 is required for the DNA damage response. Mol Cell *18*, 577–587.

[8] Brill, L. M., Salomon, A. R., Ficarro, S. B., Mukherji, M., Stettler-Gill, M., and Peters, E. C. (2004). Robust phosphoproteomic profiling of tyrosine phosphorylation sites from human T cells using immobilized metal affinity chromatography and tandem mass spectrometry. Anal Chem *76*, 2763–72.

[9] Burma, S., Chen, B. P., Murphy, M., Kurimasa, A., and Chen, D. J. (2001). ATM phosphorylates histone H2AX in response to DNA double-strand breaks. J Biol Chem *276*, 42462–42467.

[10] Canman, C. E., Lim, D. S., Cimprich, K. A., Taya, Y., Tamai, K., Sakaguchi, K., Appella, E., Kastan, M. B., and Siliciano, J. D. (1998). Activation of the ATM kinase by ionizing radiation and phosphorylation of p53. Science *281*, 1677–1679.

[11] Chen, M. J., Lin, Y. T., Lieberman, H. B., Chen, G., and Lee, E. Y. (2001). ATM-dependent phosphorylation of human Rad9 is required for ionizing radiation-induced checkpoint activation. J Biol Chem *276*, 16580–16586.

[12] Collins, M. O., Yu, L., Coba, M. P., Husi, H., Campuzano, I., Blackstock, W. P., Choudhary, J. S., and Grant, S. G. N. (2005). Proteomic analysis of in vivo phosphorylated synaptic proteins. J Biol Chem *280*, 5972–5982.

[13] Cortez, D., Wang, Y., Qin, J., and Elledge, S. J. (1999). Requirement of ATM-dependent phosphorylation of brca1 in the DNA damage response to double-strand breaks. Science *286*, 1162–1166.

[14] D'Amours, D. and Jackson, S. P. (2002). The Mre11 complex: at the crossroads of dna repair and checkpoint signalling. Nat Rev Mol Cell Biol *3*, 317–327.

[15] Diella, F., Cameron, S., Gemünd, C., Linding, R., Via, A., Kuster, B., Sicheritz-Pontén, T., Blom, N., and Gibson, T. J. (2004). Phospho.ELM: a database of experimentally verified phosphorylation sites in eukaryotic proteins. BMC Bioinformatics *5*, 79.

[16] D'Orazi, G., Cecchinelli, B., Bruno, T., Manni, I., Higashimoto, Y., Saito, S., Gostissa, M., Coen, S., Marchetti, A., Sal, G. D., et al. (2002). Homeodomain-interacting protein kinase-2 phosphorylates p53 at Ser 46 and mediates apoptosis. Nat Cell Biol *4*, 11–19.

[17] Gatei, M., Sloper, K., Sørensen, C., Syljuäsen, R., Falck, J., Hobson, K., Savage, K., Lukas, J., Zhou, B.-B., Bartek, J., et al. (2003). Ataxia-telangiectasia-mutated (ATM) and NBS1-dependent phosphorylation of Chk1 on Ser-317 in response to ionizing radiation. J Biol Chem *278*, 14806–14811.

[18] Gatei, M., Young, D., Cerosaletti, K. M., Desai-Mehta, A., Spring, K., Kozlov, S., Lavin, M. F., Gatti, R. A., Concannon, P., and Khanna, K. (2000). ATM-dependent phosphorylation of nibrin in response to radiation exposure. Nat Genet *25*, 115–119.

[19] Goldberg, M., Stucki, M., Falck, J., D'Amours, D., Rahman, D., Pappin, D., Bartek, J., and Jackson, S. P. (2003). MDC1 is required for the intra-S-phase DNA damage checkpoint. Nature *421*, 952–956.

[20] Hirao, A., Kong, Y. Y., Matsuoka, S., Wakeham, A., Ruland, J., Yoshida, H., Liu, D., Elledge, S. J., and Mak, T. W. (2000). DNA damage-induced activation of p53 by the checkpoint kinase Chk2. Science *287*, 1824–1827.

[21] Hjerrild, M., Stensballe, A., Rasmussen, T. E., Kofoed, C. B., Blom, N., Sicheritz-Pontén, T., Larsen, M. R., Brunak, S., Jensen, O. N., and Gammeltoft, S. (2004). Identification of phosphorylation sites in protein kinase A substrates using artificial neural networks and mass spectrometry. J Proteome Res *3*, 426–33.

[22] Houthaeve, T., Gausepohl, H., Mann, M., and Ashman, K. (1995). Automation of micro-preparation and enzymatic cleavage of gel electrophoretically separated proteins. FEBS Lett *376*, 91–94.

[23] Kamer, I., Sarig, R., Zaltsman, Y., Niv, H., Oberkovitz, G., Regev, L., Haimovich, G., Lerenthal, Y., Marcellus, R. C., and Gross, A. (2005). Proapoptotic BID is an ATM effector in the DNA-damage response. Cell *122*, 593–603.

[24] Khanna, K. K., Keating, K. E., Kozlov, S., Scott, S., Gatei, M., Hobson, K., Taya, Y., Gabrielli, B., Chan, D., Lees-Miller, S. P., et al. (1998). ATM associates with and phosphorylates p53: mapping the region of interaction. Nat Genet *20*, 398–400.

[25] Kim, S. T., Lim, D. S., Canman, C. E., and Kastan, M. B. (1999). Substrate specificities and identification of putative substrates of ATM kinase family members. J Biol Chem *274*, 37538–37543.

[26] Kim, S.-T., Xu, B., and Kastan, M. B. (2002). Involvement of the cohesin protein, Smc1, in Atm-dependent and independent responses to DNA damage. Genes Dev *16*, 560–570.

[27] Kishi, S., Zhou, X. Z., Ziv, Y., Khoo, C., Hill, D. E., Shiloh, Y., and Lu, K. P. (2001). Telomeric protein Pin2/TRF1 as an important ATM target in response to double strand DNA breaks. J Biol Chem *276*, 29282–29291.

[28] Li, S., Ting, N. S., Zheng, L., Chen, P. L., Ziv, Y., Shiloh, Y., Lee, E. Y., and Lee, W. H. (2000). Functional link of BRCA1 and ataxia telangiectasia gene product in DNA damage response. Nature *406*, 210–215.

[29] Lim, D. S., Kim, S. T., Xu, B., Maser, R. S., Lin, J., Petrini, J. H., and Kastan, M. B. (2000). ATM phosphorylates p95/nbs1 in an S-phase checkpoint pathway. Nature *404*, 613–617.

[30] Lim, S., Jin, K., and Friedman, E. (2002). Mirk protein kinase is activated by mkk3 and functions as a transcriptional activator of hnf1alpha. J Biol Chem *277*, 25040–25046.

[31] Lin, W. C., Lin, F. T., and Nevins, J. R. (2001). Selective induction of E2F1 in response to DNA damage, mediated by ATM-dependent phosphorylation. Genes Dev *15*, 1833–1844.

[32] Liu, Q., Guntuku, S., Cui, X. S., Matsuoka, S., Cortez, D., Tamai, K., Luo, G., Carattini-Rivera, S., DeMayo, F., Bradley, A., et al. (2000). Chk1 is an essential kinase that is regulated by Atr and required for the G(2)/M DNA damage checkpoint. Genes Dev *14*, 1448–1459.

[33] Loog, M. and Morgan, D. O. (2005). Cyclin specificity in the phosphorylation of cyclin-dependent kinase substrates. Nature *434*, 104–108.

[34] Matsuoka, S., Rotman, G., Ogawa, A., Shiloh, Y., Tamai, K., and Elledge, S. J. (2000). Ataxia telangiectasia-mutated phosphorylates Chk2 in vivo and in vitro. Proc Natl Acad Sci U S A *97*, 10389–10394.

[35] Maya, R., Balass, M., Kim, S. T., Shkedy, D., Leal, J. F., Shifman, O., Moas, M., Buschmann, T., Ronai, Z., Shiloh, Y., et al. (2001). ATM-dependent phosphorylation of Mdm2 on serine 395: role in p53 activation by DNA damage. Genes Dev *15*, 1067–1077.

[36] Melchionna, R., Chen, X. B., Blasina, A., and McGowan, C. H. (2000). Threonine 68 is required for radiation-induced phosphorylation and activation of Cds1. Nat Cell Biol *2*, 762–765.

[37] Obenauer, J. C., Cantley, L. C., and Yaffe, M. B. (2003). Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs. Nucleic Acids Res *31*, 3635–41.

[38] Panta, G. R., Kaur, S., Cavin, L. G., Corts, M. L., Mercurio, F., Lothstein, L., Sweatman, T. W., Israel, M., and Arsura, M. (2004). ATM and the catalytic subunit of DNA-dependent protein kinase activate NF-kappaB through a common MEK/extracellular signal-regulated kinase/p90(rsk) signaling pathway in response to distinct forms of DNA damage. Mol Cell Biol *24*, 1823–1835.

[39] Post, S., Weng, Y. C., Cimprich, K., Chen, L. B., Xu, Y., and Lee, E. Y. (2001). Phosphorylation of serines 635 and 645 of human Rad17 is cell cycle regulated and is required for G(1)/S checkpoint activation in response to DNA damage. Proc Natl Acad Sci U S A *98*, 13102–13107.

[40] Rush, J., Moritz, A., Lee, K. A., Guo, A., Goss, V. L., Spek, E. J., Zhang, H., Zha, X.-M., Polakiewicz, R. D., and Comb, M. J. (2005). Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. Nat Biotechnol *23*, 94–101.

[41] Saito, S., Goodarzi, A. A., Higashimoto, Y., Noda, Y., Lees-Miller, S. P., Appella, E., and Anderson, C. W. (2002). ATM mediates phosphorylation at multiple p53 sites, including Ser(46), in response to ionizing radiation. J Biol Chem *277*, 12491–12494.

[42] Salomon, A. R., Ficarro, S. B., Brill, L. M., Brinker, A., Phung, Q. T., Ericson, C., Sauer, K., Brock, A., Horn, D. M., Schultz, P. G., et al. (2003). Profiling of tyrosine phosphorylation pathways in human cells using mass spectrometry. Proc Natl Acad Sci U S A *100*, 443–8.

[43] Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res *13*, 2498–2504.

[44] Shi, Y., Venkataraman, S. L., Dodson, G. E., Mabb, A. M., LeBlanc, S., and Tibbetts, R. S. (2004). Direct regulation of CREB transcriptional activity by ATM in response to genotoxic stress. Proc Natl Acad Sci U S A *101*, 5898–5903.

[45] Srensen, C. S., Syljusen, R. G., Falck, J., Schroeder, T., Rnnstrand, L., Khanna, K. K., Zhou, B.-B., Bartek, J., and Lukas, J. (2003). Chk1 regulates the S phase checkpoint by coupling the physiological turnover and ionizing radiation-induced accelerated proteolysis of Cdc25A. Cancer Cell *3*, 247–258.

[46] Taniguchi, T., Garcia-Higuera, I., Xu, B., Andreassen, P. R., Gregory, R. C., Kim, S.-T., Lane, W. S., Kastan, M. B., and D'Andrea, A. D. (2002). Convergence of the fanconi anemia and ataxia telangiectasia signaling pathways. Cell *109*, 459–472.

[47] Tibbetts, R. S., Cortez, D., Brumbaugh, K. M., Scully, R., Livingston, D., Elledge, S. J., and Abraham, R. T. (2000). Functional interactions between BRCA1 and the checkpoint kinase ATR during genotoxic stress. Genes Dev *14*, 2989–3002.

[48] Übersax, J. A., Woodbury, E. L., Quang, P. N., Paraz, M., Blethrow, J. D., Shah, K., Shokat, K. M., and Morgan, D. O. (2003). Targets of the cyclin-dependent kinase cdk1. Nature *425*, 859–864.

[49] Ward, I. M., Minn, K., Jorda, K. G., and Chen, J. (2003). Accumulation of checkpoint protein 53BP1 at DNA breaks involves its binding to phosphorylated histone H2AX. J Biol Chem *278*, 19579–19582.

[50] Wu, X., Ranganathan, V., Weisman, D. S., Heine, W. F., Ciccone, D. N., O'Neill, T. B., Crick, K. E., Pierce, K. A., Lane, W. S., Rathbun, G., et al. (2000). ATM phosphorylation of Nijmegen breakage syndrome protein is required in a DNA damage response. Nature *405*, 477–482.

[51] Wu, Z.-H., Shi, Y., Tibbetts, R. S., and Miyamoto, S. (2006). Molecular linkage between the kinase ATM and NF-kappaB signaling in response to genotoxic stimuli. Science *311*, 1141–1146.

[52] Xu, B., O'Donnell, A. H., Kim, S.-T., and Kastan, M. B. (2002). Phosphorylation of serine 1387 in Brca1 is specifically required for the Atm-mediated S-phase checkpoint after ionizing irradiation. Cancer Res *62*, 4588–4591.

[53] Yang, D. Q. and Kastan, M. B. (2000). Participation of ATM in insulin signalling through phosphorylation of eIF-4E-binding protein 1. Nat Cell Biol *2*, 893–898.

[54] Yazdi, P. T., Wang, Y., Zhao, S., Patel, N., Lee, E. Y.-H. P., and Qin, J. (2002). SMC1 is a downstream effector in the ATM/NBS1 branch of the human S-phase checkpoint. Genes Dev *16*, 571–582.

[55] Zhao, H. and Piwnica-Worms, H. (2001). ATR-mediated checkpoint pathways regulate phosphorylation and activation of human Chk1. Mol Cell Biol *21*, 4129–4139.

[56] Zinkel, S. S., Hurov, K. E., Ong, C., Abtahi, F. M., Gross, A., and Korsmeyer, S. J. (2005). A role for proapoptotic BID in the DNA-damage response. Cell *122*, 579–591.