Science Signaling
AAAS
www.sciencesignaling.org/cgi/content/full/2/81/ra39/DC1

Supplementary Materials for

Comparative Analysis Reveals Conserved Protein Phosphorylation Networks Implicated in Multiple Diseases

Chris Soon Heng Tan, Bernd Bodenmiller, Adrian Pasculescu, Marko Jovanovic, Michael O. Hengartner, Claus Jørgensen, Gary D. Bader, Ruedi Aebersold, Tony Pawson, Rune Linding*

*To whom correspondence should be addressed. E-mail: linding@icr.ac.uk

Published 28 July 2009, *Sci. Signal.* **2**, ra39 (2009)
DOI: 10.1126/scisignal.2000316

This PDF file includes:

Other Supplementary Material for this manuscript includes the following:
(available at www.sciencesignaling.org/cgi/content/full/2/81/ra39/DC1)

Supplementary Data (ZIP file containing files in cys, tsv, txt, csv, and fa formats)

Supplementary Information for:
Comparative Analysis Reveals Conserved Protein Phosphorylation Networks Implicated in Multiple Diseases

Chris Soon Heng Tan[1,2,*], Bernd Bodenmiller[3,*], Adrian Pasculescu[1], Marko Jovanovic[4], Michael O. Hengartner[4], Claus Jørgensen[1], Gary D. Bader[1,2], Ruedi Aebersold[3,5,6,7], Tony Pawson[1,2] and Rune Linding[8†]

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

[2] Department of Molecular Genetics, University of Toronto, Toronto, Canada

[3] Institute of Molecular Systems Biology, ETH, Zurich, Switzerland

[4] Institute of Molecular Biology, University of Zurich, 8057 Zurich, Switzerland

[5] Institute for Systems Biology, Seattle, WA 98103, USA

[6] Competence Center for Systems Physiology and Metabolic Diseases, ETH Zurich, 8093 Zurich, Switzerland

[7] Faculty of Science, University of Zurich, 8057 Zurich, Switzerland

[8] Cellular & Molecular Logic Team, Section of Cell and Molecular Biology, The Institute of Cancer Research (ICR), SW3 6JB, London, UK

[*] These authors contributed equally to this work

[†] To whom correspondence should be addressed; E-mail: linding@icr.ac.uk

Supplementary Methods

Generation of peptide samples

D. melanogaster samples were generated as follows: Kc167 cells were grown in Schneiders Drosophila medium (Invitrogen) supplemented with 10% fetal calf serum, 100 U penicillin (Invitrogen) and 100 g/ml streptomycin (Invitrogen, Auckland, New Zealand) in an incubator at 25°C. To increase the number of mapped phosphorylation sites, different batches of cells were pooled. Cells were either: 1) grown in rich medium, 2) serum-starved, 3) treated for 30 min with 100 nM Rapamycin (LClabs, Woburn, MA, USA) in rich medium, 4) treated for 30 min with 100 nM insulin (serum starved), or 5) treated for 30 min with 100 nM Calyculin A (rich medium). Then the cells were washed with ice-cold phosphate-buffered saline and resuspended in ice-cold lysis buffer containing 10 mM HEPES, pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM dithiothreitol and a protease inhibitor mix (Roche, Basel, Switzerland). To preserve protein phosphorylation, several phosphatase inhibitors were added to a final concentration of 20 nM calyculin A, 200 nM okadaic acid, 4.8 μm cypermethrin (all bought from Merck KGaA, Darmstadt, Germany), 2 mM vanadate, 10 mM sodium pyrophosphate, 10 mM NaF and 5 mM EDTA. After 10 min incubation on ice, cells were lysed by douncing. Cell debris and nuclei were removed by centrifugation for 10 min at 4°C at 5500 g. Then the cytoplasmic and membrane fraction were separated by ultracentrifugation at 100000 g for 60 min at 4°C. The proteins of the cytosolic fraction (supernatant) were subjected to acetone precipitation. The protein pellets were resolubilized in 3 mM EDTA, 20 mM Tris-HCl, pH 8.3, and 8 M urea. The disulfide bonds of the proteins were reduced with tris (2-carboxyethyl) phosphine at a final concentration of 12.5 mM at 37°C for 1 h. The produced free thiols were alkylated with 40 mM iodoacetamide at room temperature for 1 h. The solution was diluted with 20 mM Tris-HCl (pH 8.3) to a final concentration of 1.0 M urea and digested with sequencing-grade modified trypsin (Promega, Madison, WI) at 20 μg per mg of protein overnight at 37°C. Peptides were desalted on a C18 Sep-Pak cartridge (Waters, Milford, MA) and dried in a speedvac.

The *S. cerevisiae* peptide samples were generated as follows. Wild type (BY7092: can1::STE2prSp his5 lyp1Δ his3Δ leu2Δ ura3Δ met15Δ) were grown to OD ∼0.8 at 30°C in synthetic dened (SD) medium (per liter 1.7 g YNB, 5 g ammonium sulfate, 2% glucose, 0.03 g isoleucine, 0.15 g valine, 0.04 g adenine, 0.02 g arginine, 0.1 g leucine, 0.03 g lysine, 0.02 g methionine, 0.05 g phenylalanine, 0.2 g threonine, 0.02 g histidine, 0.02 g tryptophane, 0.03 g tyrosine, 0.02 g uracil, 0.1 g glutamic acid and 0.1 g aspartic acid). Cells were harvested at 30°C by centrifugation at 850 g and then washed once in SD medium. Finally, they were collected by centrifugation and shock-frozen in liquid nitrogen. Pellets were thawed in ice-cold lysis buffer (20 mM Tris/HCl pH 8.0, 100 mM KCl, 10 mM EDTA, 0.1% NP40, 20 nM calyculin A, 200 nM okadaic acid, 4.8 μM cypermethrin (all obtained from Merck KGaA, Darmstadt, Germany), 2 mM vanadate, 10 mM sodium pyrophosphate and 10 mM NaF) using 1 mL of lysis buffer per gram of yeast. The cells were lysed by glass-bead beating (using acid-washed glass beads), the protein supernatant was precipitated using ice-cold acetone, and the pellet was resuspended in 8 M urea, 20 mM Tris/HCl at pH 8.3. After dilution to $<$ 1.5 M urea with 20 mM Tris/HCl at pH 8.3, proteins were digested using trypsin in a w/w ratio of 1:125 and puried using C18 reverse phase chromatography (Sep-Pack, Waters).

For *C. elegans* the wild-type strain N2 (Bristol) was grown on 9-cm nematode growth medium (NGM) agar plates seeded with a lawn of the E. coli strain OP50 or in 100-ml liquid cultures in S-basal buffer in bevelled flasks (with concentrated E. coli NA22 as a food source). Worms were harvested from plates or liquid culture and separated from the bacteria by washing with M9 buffer three times. The worms were harvested by centrifugation at 500 g and shock frozen using liquid nitrogen. Subsequently, the worms were homogenized with glass beads (diameter of 212-300 μm, Sigma-Aldrich, St Louis, MO, USA) in the ratio of 1:1:2 (worms:beads:buffer) in a cell disrupter (FastPrep FP120, Thermo Savant, Qbiogene Inc., Carlsbad, CA, USA) at 4°C three times for 45 s at level 6. The buffer used was 50 mM Tris/HCl, pH 8.3, 5 mM EDTA, 8 M urea. After glass bead beating treatment, 0.125% SDS was added and the homogenate was incubated for 1 h at room temperature to solubilize proteins. Cell debris was removed by centrifugation. The peptides were produced from the proteins in the supernatant as described above. For further details refer to (1–4).

Peptide separation by isoelectric focusing

All peptides were separated according to their isoelectric point. For the *D. melanogaster* this was performed using an free-flow electrophoresis instrument, type prometheus from FFE Weber Inc. (now BD-Diagnostics, PAS) and FFE-Weber reagent basic kit (Prolyte 1, Prolyte 2, Prolyte 3 and Prolyte 4-7 and pI markers) (BD-Diagnostics, NJ, USA). The digested peptides were diluted in separation media containing 8 M Urea, 250 mM Mannitol and 20% ProLyte solution at a concentration of 10 mg/ml. This sample was loaded continuously for 1 h at 1 ml/h. Total collection time was 24 h and the volume of each collected fraction was about 25 to 50 ml. A Thermo Orion needle tip micro pH electrode (Thermo Electron Corporation, Beverly, MA) was used to measure the pH value of each fraction. Peptides from the FFE fractions 18-60 were purified on a C18 Sep-Pak cartridge (Waters Corporation, Milford, MA, USA) (1). For *C. elegans* and *S. cerevisiae* the dried-down

peptide samples (15 mg and 20 mg, respectively) were seperated with an Offgel fractionator and therefore resolubilized to a final concentration of 1 mg/ml in off-gel electrophoresis buffer containing 6.25% glycerol and 1.25% IPG buffer (GE Healthcare). The peptides were separated on pH 3-10 IPG strips (GE Healthcare) with a 3100 OFFGEL fractionator (Agilent) as previously described (4; 5). We performed a 1-hour rehydration at maximum 500 V, 50 mA, and 200 mW followed by the separation at maximum 8000V, 100 mA, and 300 mW until 50 kVh were reached. Following isoelectric focussing, the fractions were concentrated and cleaned up by C18 reversed-phase spin columns according to the manufacturers instructions (Sep-Pack, Waters).

Detailed description of phosphopeptide isolation

Phosphopeptides were isolated using a titanium dioxide resin as follows: 1-3 mg of dried peptides were reconstituted in 280 μl of a washing solution (WS), containing 80% acetonitrile and 3.5% TFA, which is saturated with phtalic acid ($\sim$100 mg phtalic acid per ml). Then 1.25 mg TiO_2 (GL Science, Saitama, Japan) resin was placed into a 1-ml Mobicol spin column (MoBiTec, Göttingen, Germany) and was subsequently washed with 280 μl water, 280 μl methanol, and finally was equilibrated with 280 μl WS for at least 10 minutes. After removal of the WS by centrifugation using 500 x g, the peptide solution was added to the equilibrated TiO_2 in the blocked Mobicol spin column and was incubated for $>$ 30 min with end-over-end rotation. After this step, the peptide solution was removed by centrifugation, and the resin was thoroughly washed two times each with 280 μl of the WS, with a 80% acetonitrile, 0.1% TFA solution, and finally with 0.1% TFA. In the final step, phosphopeptides were eluted from the TiO_2 resin using two times 150 μl of a 0.3 M NH_4OH solution (pH $\sim$10.5). After elution, the pH of the pooled eluents was rapidly adjusted to 2.7 with 10% TFA, and the phosphopeptides were purified with an appropriate reverse-phase column suitable for up to 20 μg peptide. Besides the separated peptides, this procedure was also performed on yeast and worm whole-cell or whole-organism lysates.

Alternatively, phosphopeptides were also isolated with immobilization by metal affinity chromatography (IMAC). In detail, 1-3 mg of peptides were reconstituted in 280 μl of a WS, consisting of 250 mM acetic acid with 30% acetonitrile at pH 2.7. Then 60 μl of uniformly suspended PHOS-Select iron affinity gel (Sigma Aldrich), corresponding to $\sim$30 μl resin, was placed into a 1-ml Mobicol spin column. The resin was equilibrated three times with 280 μl of the WS. After removal of the WS by centrifugation at 500 x g, the peptide solution was added to the equilibrated IMAC resin in the blocked Mobicol spin column. To obtain reproducible results, it is crucial that the pH in all replicate samples is maintained at $\sim$2.5. The affinity gel was then incubated with the peptide solution for 120 min with end-over-end rotation. After the incubation, the liquid was removed by centrifugation and the resin was thoroughly washed two times with 280 μl of the WS, and once with ultra pure water. In the final step, phosphopeptides were eluted once with 150 μl of a 50 mM phosphate buffer (pH 8.9) and once with 150 μl of a 100 mM phosphate buffer (pH 8.9), each time incubating the resin $<$ 3 min with the elution buffer. Both elutes were pooled, the pH were rapidly adjusted to 2.7 using 10% TFA, and the phosphopeptides were purified with an appropriate reverse-phase column. This procedure was performed on separated fly and worm peptide samples and yeast whole-cell lysates.

Finally, phosphoramidate chemistry (PAC) was used in the case of *D. melanogaster* for phosphopeptide isolation on the peptide samples after isoelectric focusing (1; 2). Phosphopeptides were isolated with phosphoramidate chemistry as follows: 1 mg of dried peptide was reconstituted in 750 μl of methanolic HCl, which was prepared by *slowly* adding 120 μl of acetyl chloride to 750 μl of anhydrous methanol. The methyl esterification was then allowed to proceed at 12°C for 120 min. The solvent was quickly removed in a cool vacuum concentrator and peptide methyl esters were dissolved in 40 μl methanol, 40 μl water, and 40 μl acetonitrile. Then 500 μl of a solution containing 50 mM N-(3-Dimethylaminopropyl)-N'-ethylcarbodiimide (EDC), 100 mM imidazole pH 5.6, 100 mM 2-(N-Morpholino)ethanesulfonic acid (MES) pH 5.6, and 2 M cystamine was added to the peptide solution. The reaction was allowed to proceed at room temperature with vigorous shaking for 8 hours. The solution was then loaded onto an appropriate reverse-phase column and the derivatized peptides were subsequently: First, washed with 0.1% TFA; second, treated with 10 mM TCEP (pH should be adjusted to $\sim$3 using sodium hydroxide (NaOH)) for 8 minutes, in order to produce free thiol groups; third, washed again with 0.1% TFA to remove residual TCEP. Finally, the derivatized peptides were eluted with 80% acetonitrile, 0.1% TFA and the pH was adjusted to 6.0 with phosphate buffer. Then acetonitrile was partially removed in the vacuum concentrator to yield a final concentration of $\sim$30%, and the derivatized phosphopeptides were incubated with 5 mg maleimide functionalized-glass beads for 1 h at pH 6.2 in a Mobicol column. (The beads were synthesized by dissolving 120 μmol hydroxybenzotriazole, 120 μmol of 3-maleimidopropionic acid, and 120 μmol diisopropylcarbodiimide in 1 ml of dry dimethylformamide, completely. After 30 minutes of incubation, 100 mg CPG beads (Proligo Biochemie, Hamburg, Germany) corresponding to 40 μmol free amino groups were added for 90 minutes. After the reaction, beads were washed with dimethylformamide and dried with a vacuum concentrator. Beads were stored dry at 4°C.) The derivatized beads were washed two times sequentially with 300 μl 3 M NaCl, water, methanol, and, finally, with 80% acetonitrile to remove nonspecifically bound peptides. In the last step, the beads were incubated with 5% TFA, 30% acetonitrile for 1 h to

recover the phosphopeptides. The recovered sample was dried in the vacuum concentrator. This procedure was also performed on yeast whole-cell lysates. For further details refer to (1–4).

Mass spectrometry data analysis and sampling depths

The liquid chromatography–tandem mass spectrometry (LC-MS, on a Thermo Fisher Scientic LTQ ORBITRAP XL) analysis and database searches were performed as described in (1). The *S. cerevisiae* and *C. elegans* MS spectra were searched against the SGD (release October 10th 2007) and WormBase (release WS183) databases and the *D. melanogaster* data were searched against the FlyBase database v4.3. In addition to the data stored in the PhoshoPep database, we added in the case of *S. cerevisiae* electron transfer dissociation (ETD) fragmentation data from (6), although these data only constitute 19% of the total dataset. In this study, the *D. melanogaster* phosphopeptide isolates were most extensively analyzed in terms of the total number of LC-MS/MS runs employed and consequently larger coverage was achieved than for the other target species. Finally, the expected sizes of the phosphoproteomes of yeast, worm, and fly strongly differ, simply due to differences in their genome sizes and repertoires of kinases and phosphatases.

Supplementary Figures

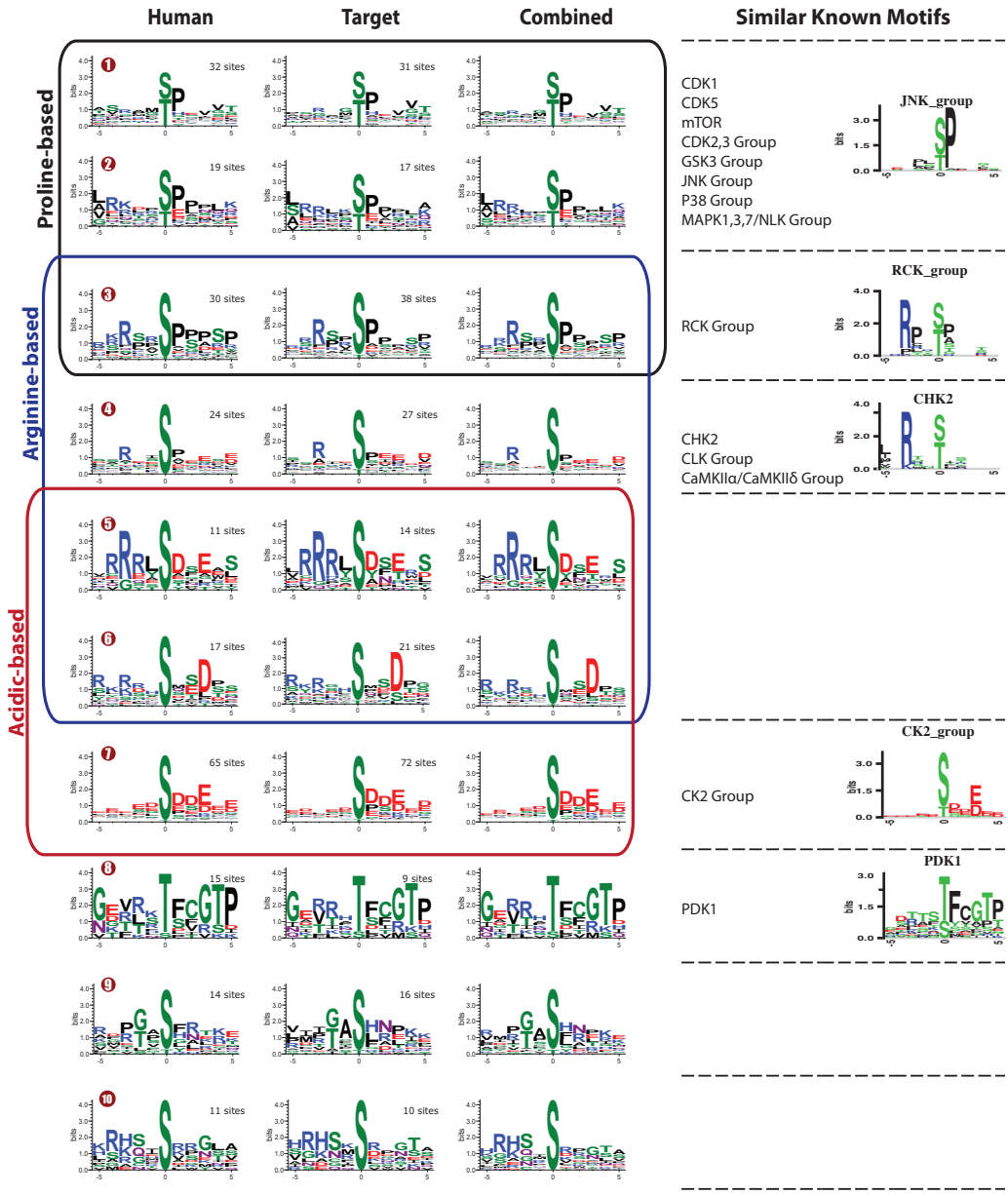

Figure S1: **Clustering analysis of core phosphorylation sites.** For every pair of aligned phosphorylated residues, a consensus sequence of the local alignment from -5 to 5 of the aligned phosphorylated residues is first defined. For example, `..RK.SP..D.` is the consensus pattern of GTRKG*p*SPLKDE aligned to NERKV*p*SPDEDM. Next, a consensus pattern S encoded as a vector set $V = (v_{-5}, v_{-4}, \ldots, v_4, v_5)$ is defined, where vector v_i is a vector of the 20 element coding for number of specific amino acids appearing at position i among the consensus sequences. The similarity between vector set V_x and V_y is computed as the sum of cosine similarities of all corresponding vectors across the two sets, as follows: First, a vector set is encoded for every consensus sequence. Secondly, the similarity between pairs of vector sets are computed, and the most similar pair is then merged into a new vector set by summing up the corresponding vectors across the two old sets. The previous step is iteratively performed, and if the two most similar vector sets at each iteration encodes 10 or more core sites, they are distilled out from the computational analysis. Lastly, core sites in human and target species represented by output vector sets are visualized separately with sequence logo, manually, inspected and grouped as shown. Any kinase consensus motif in NetPhorest (termed kinase group) (7) that resembles a motif discovered above are listed to the right and the corresponding selected motif logos from the NetPhorest atlas is shown. For example, number 8 corresponds to a conserved segment in the activation loop of some serine/threonine kinases and resembles the consensus motif of the ”master” kinase regulator PDK1. As the motifs above are derived from unsupervised clustering, some of them may not correspond to kinase consensus motif. In addition, some phosphorylation motifs could be missed due to weak signal from limited motif instances.

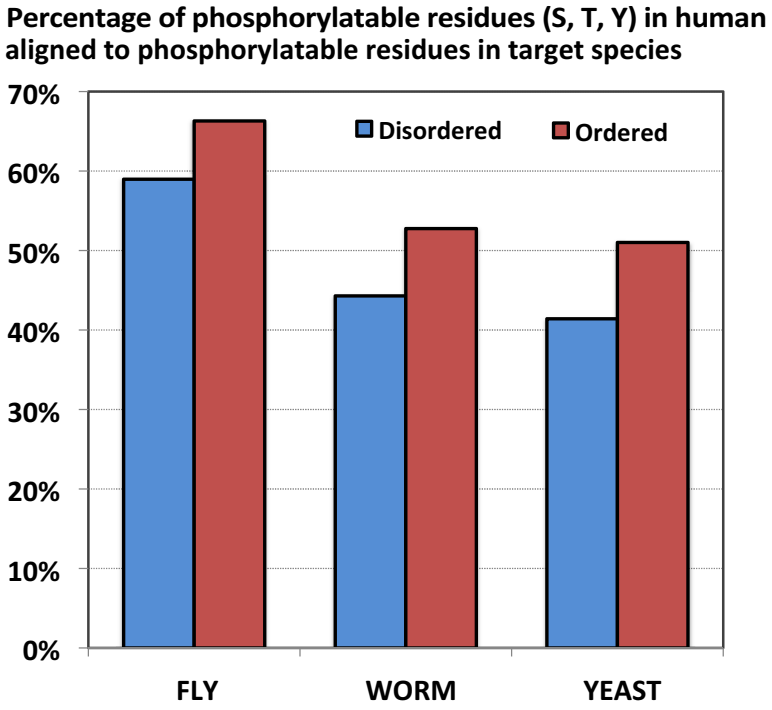

Figure S2: **Phosphorylatable residues in disordered regions are fast evolving.** Intrinsic disordered regions in proteins are generally rapidly evolving and hard to align. However, human phosphorylatable residues (serine, theronine, and tyrosine) in disordered regions that are well-aligned to target species (defined as those with at least 50% identity in -5 to +5 of the central residue) are less conserved than phosphorylatable residues in ordered regions at the same identity threshold. Here, we deemed a phosphorylatable residue in human to be conserved in target species if it is aligned to another phosphorylatable residue. The statistics are computed only from human orthologs in target species that have one-to-one orthologous relationship to avoid relaxation of conservation caused by duplicate genes.

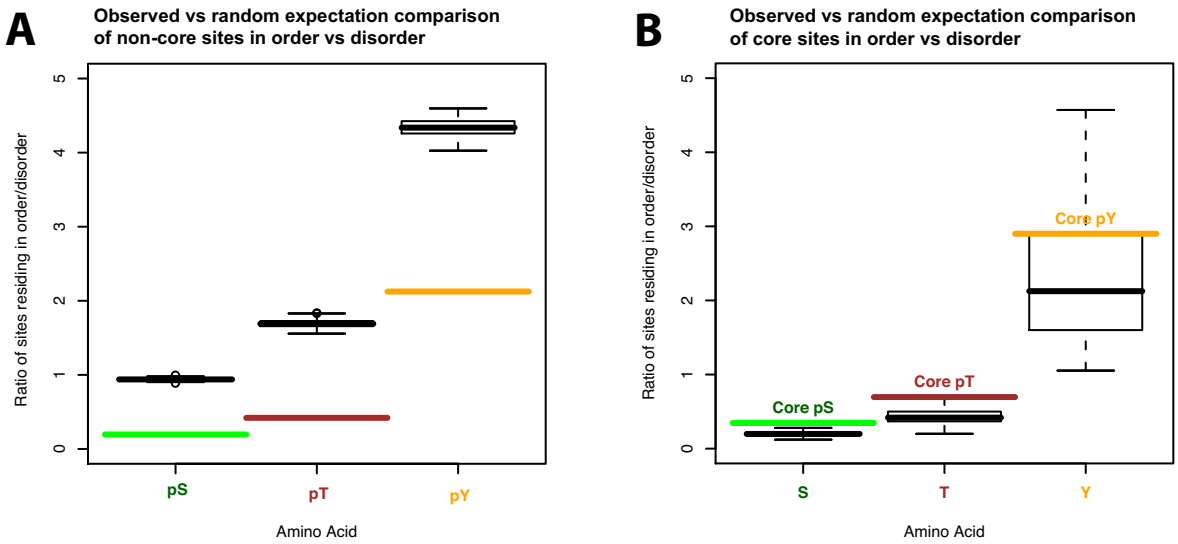

Figure S3: **Phosphorylation site disorder analysis. A,** Known human phosphorylation sites occur more often in disordered regions than would be expected by random chance. The boxplots (in black) display the distributions of in-order/in-disorder ratios for 100 randomly picked sets of serines, threonines, and tyrosines from human phosphoproteins. Phosphosphorylated serines, threonines, and tyrosines might fall by random chance into these sets as well. The size of each set equals that of the respective observed number of phosphorylated residues. The colored bars denote the observed in-order/in-disorder ratio of phosphorylated residues. **B**, Core sites occur more often in ordered or structured regions than do other phosphorylation sites. To assess the significance of this observation, we sampled the distribution of in-order/in-disorder ratio with sets of randomly picked phosphorylation sites from core site proteins. The distributions of the random sampling are shown as boxplots (in black) and the colored bars denote the in-order/in-disorder ratio of core sites. Separate analyses were performed for phosphoserine, -threonine, and -tyrosine.

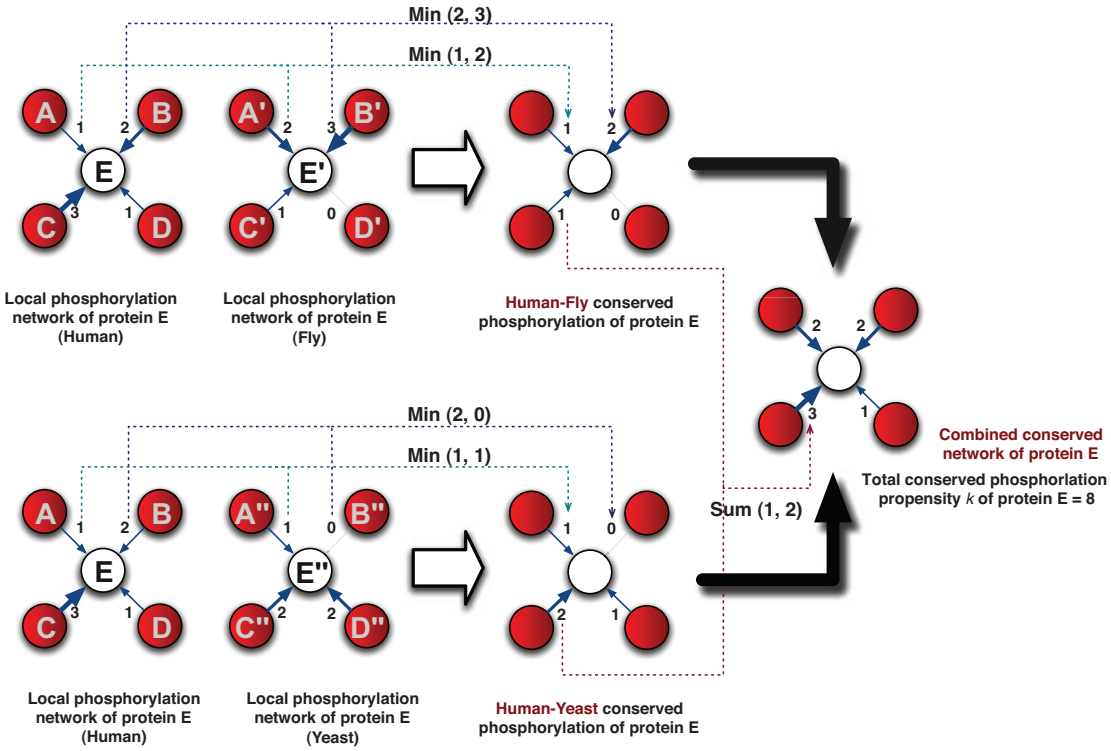

Figure S4: **Schematic diagram of how the conserved phosphorylation propensity *k* of each human substrate is computed.** Inferred conserved human kinase-substrate relationships are quantified based on the number of sites predicted to be phosphorylated by an orthologous kinase on an orthologous substrate in human and target species. The conserved phosphorylation propensity of a substrate is computed from its conserved human kinase-substrate relationships as *k* which is the weighted sum of all its incoming edges. Kinases are represented as red-colored nodes. For simplicity, the illustration is shown for two out of the three target species used. The numbers located at the end of each arrow indicate the number of phosphorylation sites predicted [by NetworKIN (8; 9) using NetPhorest (7) as motif engine] to be targeted by the corresponding kinase.

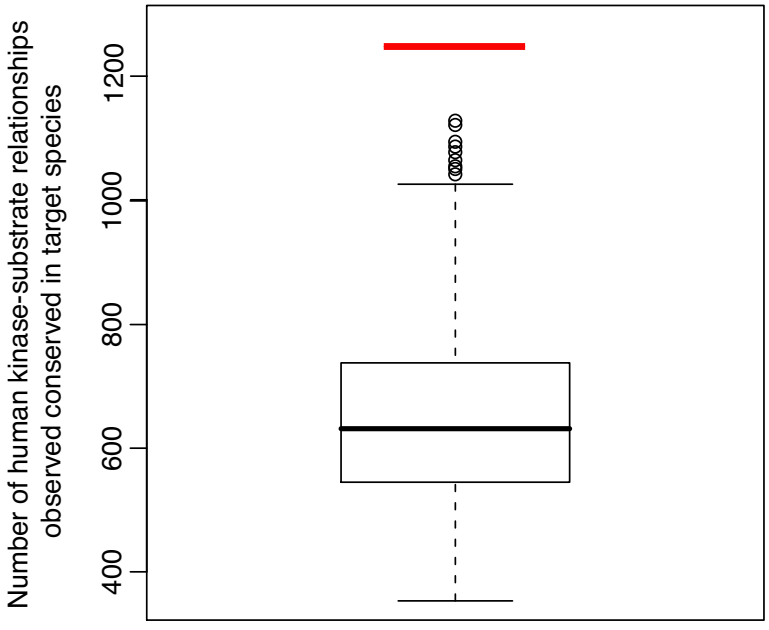

Figure S5: **Boxplot of the number human-substrate relations observed that were conserved in target species from randomized trials.** Red bar denotes the actual number of observed conserved human kinase-substrate relationships.

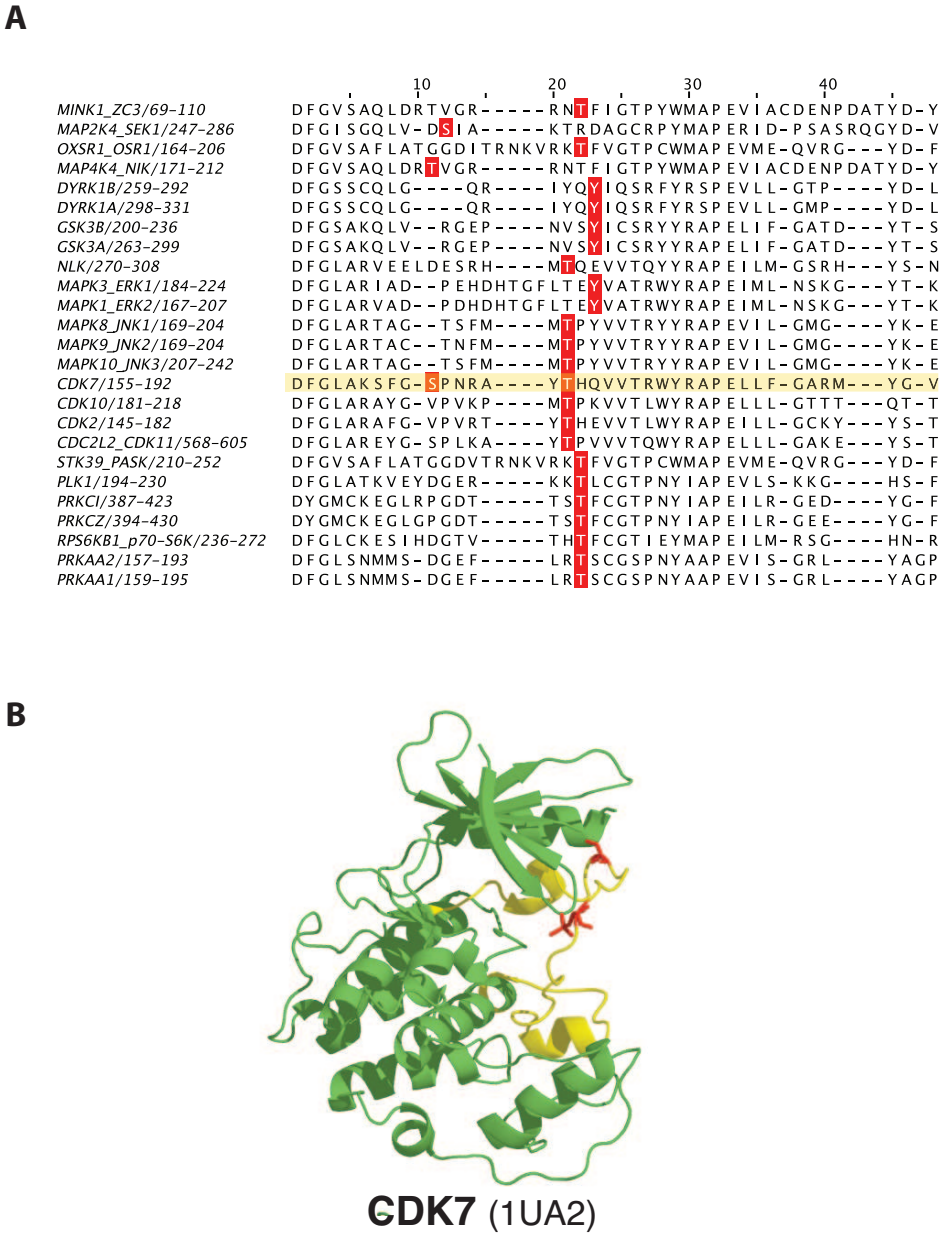

Figure S6: **Core sites observed in activation loops of protein kinases. A,** A multiple sequence alignment of activation loop regions in protein kinases with core sites marked in red. Gene symbols, kinase names, and the sequence coordinates for each activation loop is shown on the left. **B,** The corresponding region from panel A is highlighted (yellow) on the protein structure of CDK7.

Supplementary Tables

Table S1: **Statistical significance of core sites observed between human and yeast**. The number of core sites expected by random chance is computed from the observed sequence alignments. The statistical significance is assessed with χ^2 test in *R* statistical software.

Residue	Expected	Observed	Odds Ratio	*p*-value
pS	9.54	46	4.82	1.0e-07
pT	0.73	2	2.74	-
pY	1.48	5	3.38	-

Table S2: **Statistical significance of core sites observed between human and fly**. The number of core sites expected by random chance is computed from the observed sequence alignments. The statistical significance is assessed with χ^2 test in *R* statistical software.

Residue	Expected	Observed	Odds Ratio	*p*-value
pS	55.53	325	5.85	2.2e-16
pT	6.43	64	9.95	6.9e-12
pY	11.23	33	2.94	1.1e-3

Table S3: **Statistical significance of core sites observed between human and worm**. The number of core sites expected by random chance is computed from the observed sequence alignments. The statistical significance is assessed with χ^2 test in *R* statistical software.

Residue	Expected	Observed	Odds Ratio	*p*-value
pS	14.9	116	7.79	2.2e-16
pT	0.89	7	7.87	-
pY	2.99	13	4.35	1.2e-2

Table S4: **Core sites identified in components of beta-catenin destruction complex.**

Gene Symbol (HGNC)	Ensembl Protein ID	Residue	Position	Target Species	Source Type
GSK3B	ENSP00000264235	Y	216	WORM FLY YEAST	HTP LTP
GSK3B	ENSP00000264235	S	9	FLY	HTP LTP
APC	ENSP00000257430	S	1279	FLY	HTP LTP
AXIN1	ENSP00000262320	T	79	FLY	HTP
CTNNB1	ENSP00000344456	S	675	FLY	HTP LTP

Table S5: **Core sites identified in components of the clathrin coat of coated pits.** The LTCL1 subunit of clarthin coat does not contain any core sites according to our approach.

Gene Symbol (HGNC)	Ensembl Protein ID	Residue	Position	Target Species	Source Type
CLTA	ENSP00000242285	Y	94	WORM	HTP
CLTC	ENSP00000269122	Y	634	FLY	HTP
AP2M1	ENSP00000292807	T	156	FLY	HTP LTP
CLTB	ENSP00000309415	Y	87	WORM	HTP

Table S6: **Core sites identified in 22 cancer-associated genes.** Phosphorylation sites annotated with HTP and not LTP evidence are considered recently dicovered in this study.

Gene Symbol (HGNC)	Ensembl Protein ID	Residue	Position	Target species	Source Type
APC	ENSP00000257430	S	1279	FLY	LTP
CCDC6	ENSP00000263102	S	367	FLY	HTP
CCDC6	ENSP00000263102	S	240	FLY	HTP
CCDC6	ENSP00000263102	S	323	FLY	HTP
CDK2	ENSP00000266970	T	160	FLY	HTP LTP
CDK2	ENSP00000266970	Y	15	YEAST	HTP LTP
CLTC	ENSP00000269122	Y	634	FLY	HTP
CREB1	ENSP00000236996	S	117	WORM	LTP
CTNNB1	ENSP00000344456	S	675	FLY	HTP LTP
DEK	ENSP00000244776	S	230	FLY	HTP
DEK	ENSP00000244776	S	231	FLY	HTP
DEK	ENSP00000244776	S	306	FLY	HTP
DEK	ENSP00000244776	S	232	FLY	HTP
FIP1L1	ENSP00000351383	S	85	FLY	HTP
FOXO1	ENSP00000368880	S	319	WORM	LTP
FOXO3	ENSP00000339527	S	315	WORM	LTP
HSP90AA2	ENSP00000216281	S	231	FLY	HTP LTP
HSP90AA2	ENSP00000216281	Y	627	FLY	HTP
HSP90AB1	ENSP00000360709	Y	619	FLY	HTP
HSP90AB1	ENSP00000360709	S	226	FLY	HTP LTP
MAP2K4	ENSP00000262445	S	257	FLY	HTP LTP
MLLT4	ENSP00000345834	S	1090	FLY	HTP
FOX04	ENSP00000363377	S	262	WORM	LTP
MSH6	ENSP00000234420	S	14	FLY	HTP
RPS10	ENSP00000347271	T	101	YEAST	HTP
SEPT6	ENSP00000341524	T	418	FLY	HTP
SMAD2	ENSP00000262160	T	8	FLY	HTP LTP
SUZ12	ENSP00000316578	S	583	FLY	HTP
TCF12	ENSP00000267811	S	559	FLY	HTP
VCP	ENSP00000367954	S	748	YEAST	LTP

Table S7: **Correlation of cancer-associated genes with conserved phosphorylation propensity *k* computed for individual target species.**

	Fly + Worm + Yeast			Fly			Worm			Yeast		
k	Cancer Genes	Total Genes	%	Cancer Genes	Total Genes	%	Cancer Genes	Total Genes	%	Cancer Genes	Total Genes	%
≥ 1	47	759	6.2	44	631	7.0	3	96	3.1	6	215	2.8
≥ 2	25	450	5.6	21	344	6.1	1	49	2.0	3	101	3.0
≥ 3	15	271	5.5	15	209	7.2	0	17	0.0	1	41	2.4
≥ 4	12	192	6.2	12	138	8.7	0	14	0.0	1	25	3.6
≥ 5	11	143	7.7	10	93	10.8	0	6	0.0	1	15	6.7
≥ 6	10	103	9.7	8	70	11.4	0	3	0.0	1	9	11.1
≥ 7	8	73	11.0	7	47	14.9	0	3	0.0	0	7	0.0
≥ 8	4	54	7.4	4	31	12.9	0	3	0.0	0	5	0.0

Table S8: **Correlation of OMIM genes with conserved phosphorylation propensity *k* computed for different target species.**

	Fly + Worm + Yeast			Fly			Worm			Yeast		
k	Cancer Genes	Total Genes	%	Cancer Genes	Total Genes	%	Cancer Genes	Total Genes	%	Cancer Genes	Total Genes	%
≥ 1	99	759	13.0	85	631	13.5	8	96	8.3	27	215	12.6
≥ 2	55	450	12.2	42	344	12.2	5	49	10.2	14	101	13.9
≥ 3	34	271	12.5	26	209	12.4	1	17	5.9	7	41	17.1
≥ 4	24	192	12.5	20	138	14.5	1	14	7.1	4	25	16.0
≥ 5	23	143	16.1	19	93	20.4	0	6	0.0	3	15	20.0
≥ 6	18	103	17.5	16	70	22.9	0	3	0.0	3	9	33.3
≥ 7	15	73	20.5	14	47	29.8	0	3	0.0	2	7	28.6
≥ 8	13	54	24.1	10	31	32.3	0	3	0.0	1	5	20.0
≥ 9	11	41	26.8	5	20	25.0	0	2	0.0	0	4	0.0
≥ 10	10	35	28.6	4	16	25.0	0	2	0.0	0	4	0.0
≥ 11	8	27	29.6	4	11	36.4	0	2	0.0	0	1	0.0
≥ 12	7	24	29.2	4	11	36.4	0	1	0.0	0	1	0.0
≥ 13	7	22	31.8	4	9	44.4	0	1	0.0	0	0	0.0
≥ 14	7	20	35.0	4	6	66.7	0	1	0.0	0	0	0.0
≥ 15	4	16	25.0	3	5	60.0	0	1	0.0	0	0	0.0

Supplementary References

[1] B. Bodenmiller, J. Malmstrom, B. Gerrits, D. Campbell, H. Lam, A. Schmidt, O. Rinner, L. N. Mueller, P. T. Shannon, P. G. Pedrioli, C. Panse, H.-K. Lee, R. Schlapbach, R. Aebersold, Phosphopep–a phosphoproteome resource for systems biology research in drosophila kc167 cells. *Mol Syst Biol* **3**, 139 (2007).

[2] B. Bodenmiller, L. N. Mueller, M. Mueller, B. Domon, R. Aebersold, Reproducible isolation of distinct, overlapping segments of the phosphoproteome. *Nat Methods* **4**, 231–237 (2007).

[3] B. Bodenmiller, L. N. Mueller, P. G. A. Pedrioli, D. Pflieger, M. A. Jnger, J. K. Eng, R. Aebersold, W. A. Tao, An integrated chemical, mass spectrometric and computational strategy for (quantitative) phosphoproteomics: application to drosophila melanogaster kc167 cells. *Mol Biosyst* **3**, 275–286 (2007).

[4] B. Bodenmiller, D. Campbell, B. Gerrits, H. Lam, M. Jovanovic, P. Picotti, R. Schlapbach, R. Aebersold, Phosphopep–a database of protein phosphorylation sites in model organisms. *Nat Biotechnol* **26**, 1339–1340 (2008).

[5] M. Heller, M. Ye, P. E. Michel, P. Morier, D. Stalder, M. A. Jünger, R. Aebersold, F. Reymond, J. S. Rossier, Added value for tandem mass spectrometry shotgun proteomics data validation through isoelectric focusing of peptides. *J Proteome Res* **4**, 2273–2282 (2005).

[6] A. Chi, C. Huttenhower, L. Y. Geer, J. J. Coon, J. E. P. Syka, D. L. Bai, J. Shabanowitz, D. J. Burke, O. G. Troyanskaya, D. F. Hunt, Analysis of phosphorylation sites on proteins from saccharomyces cerevisiae by electron transfer dissociation (ETD) mass spectrometry. *Proc Natl Acad Sci U S A* **104**, 2193–2198 (2007).

[7] M. L. Miller, L. J. Jensen, F. Diella, C. Jørgensen, M. Tinti, L. Li, M. Hsiung, S. A. Parker, J. Bordeaux, T. Sicheritz-Ponten, M. Olhovsky, A. Pasculescu, J. Alexander, S. Knapp, N. Blom, P. Bork, S. Li, G. Cesareni, T. Pawson, B. E. Turk, M. B. Yaffe, S. Brunak, R. Linding, Linear motif atlas for phosphorylation-dependent signaling. *Sci Signal* **1**, ra2 (2008).

[8] R. Linding, L. J. Jensen, G. J. Ostheimer, M. A. van Vugt, C. Jørgensen, I. M. Miron, F. Diella, K. Colwill, L. Taylor, K. Elder, P. Metalnikov, V. Nguyen, A. Pasculescu, J. Jin, J. G. Park, L. D. Samson, J. R. Woodgett, R. B. Russell, P. Bork, M. B. Yaffe, T. Pawson, Systematic discovery of in vivo phosphorylation networks. *Cell* **129**, 1415–1426 (2007).

[9] R. Linding, J. L. J., A. Pasculescu, M. Olhovsky, K. Colwill, P. Bork, M. B. Yaffe, T. Pawson, NetworKIN: a resource for exploring cellular phosphorylation networks. *Nucleic Acids Res.* **36**, D695–D699 (2008).

Supplementary Data File Descriptions

The supplementary data files are provided as a 17 Mb Zip archive and when unzipped, the data are organized into 8 subdirectories as described below.

Yeast/

YeastSeq.fa: Yeast sequences used.
YeastSite.csv: Experimentally detected yeast phosphosites.
YeastSiteNetworKIN.tsv: NetworKIN + NetPhorest predictions of kinases phosphorylating the sites.

Worm/

WormSeq.fa: Worm sequences used.
WormSite.csv: Experimentally detected worm phosphosites.
WormSiteNetworKIN.tsv: NetworKIN + NetPhorest predictions of kinases phosphorylating the sites.

Fly/

FlySeq.fa: Fly sequences used.
FlySite.csv: Experimentally detected fly phosphosites.
FlySiteNetworKIN.tsv: NetworKIN + NetPhorest predictions of kinases phosphorylating the sites.

Human/

HumanSite.csv: Human phosphosites assembled from the PhosphoSite and Phospho.ELM databases.
HumanSiteNetworKIN.tsv: NetworKIN + NetPhorest prediction of kinases phosphorylating the sites

CoreSites/

CoreSite.csv: core sites detected. Included are NetworKIN and NetPhorest predictions.
CoreSiteinDomain.csv: Core site occurring within computationally inferred domains using HMM profiles from the PFAM and SMART databases.
CoreSiteinMultipleSpecies.csv: List of human core sites occurring in more than one target species.
ClusterAnalysis.txt: List of motif clusters identified

CoreSites/GOAnalysis/: Bingo (Cytoscape plugin) outputs for Gene Ontology function enrichment analysis of core site proteins.

CoreNet/

CoreNet.txt: Inferred human kinase-substrate relationships conserved in target species.
CoreNetWithoutCoreSite.txt: Inferred human kinase-substrate relationships conserved in target species when kinase-substrate relationships involving core sites are excluded.

CoreNet/GOAnalysis/: Bingo (Cytoscape plugin) outputs for Gene Ontology function enrichment analysis of core net proteins.

MSA/

Multiple sequence alignments used in core site detection.

DiseaseGeneAnalysis/

DiseaseGene.csv: List of core site and core net genes with known mutations associated with cancer and other diseases (OMIM). Included also is computed conserved phosphorylation propensity (k) of these genes.