Supplementary Information

Contents:

Table S1. Table of related databases/tools.

Tool	Type	Descriptions
PhosphoSitePlus (1) PhosphoELM (2) PHOSIDA (3) PhosphoPOINT (4)	Database	Phospho-site and regulatory information
DEPOD (5)	Database	Phosphatase activity and substrates
KEA (6)	Prediction tool	Kinase enrichment
KinasePhos2.0 (7)	Prediction tool	Machine learning
NetworKIN (8), (9) KinomeXplorer (10), PhosphoSiteAnalyzer (11)	Prediction tool	Motifs/Position Specific scoring matrices & contextual network information
Scansite 2.0 (12)	Prediction tool	Motifs/Position Specific scoring matrices
GPS 2.0 (13)	Prediction tool	Kinase group hierarchy
HeRS (14)	Prediction tool	Kinase/Transcription factor similarity
PhosphoChain (15)	Prediction tool	Expression profile integration
RegPhos (16)	Prediction tool	Expression and protein localization data integration
Saez-Rodriguez et al (17)	Pathway Modeling	Logic/Boolean based
Mitsos et al (18)	Pathway Modeling	Integer Linear Programming
Zhang et al (19)	Pathway Modeling	Bayesian network inference

Table S2. Complete set of input parameters and their default values for SELPHI

Parameter	Default	Description
Job name	NONE	Name of your job
Input Data	NONE	Tab delimited text or Excel™ files with phospho-proteomics data (Proteins, Peptides, Ratios, optional Intensity and/or Score)
Merge samples	NONE	File with comma separated samples that should be merged for the analysis
Log	NO	Indicate if your data has already been log transformed
Analyze Motifs	NO	Request a motif over-representation analysis
Phospho-Site map	NONE	File with the protein-peptide pairs mapped to the sequence MAPK14_pSQERPTFYR MAPK14_S2
ID map	NONE	Tab delimited text or Excel™ file with the proteins identified mapped to UniprotID or GeneID
FASTA database	NONE	Database searched to identify proteins in set (in FASTA format)
Clustering method	NONE	Choose whether to cluster phospho-peptides using PCA & k-means clustering or mclust
STRING evidence	Overall	Select which STRING score to use as a cutoff for your network overlap: Overall, Experimental and/or Database
GeneMania evidence	0	Select which GeneMania data to use, everything (0) or physical interaction only (1)
Ratio Cutoff	3	Cutoff for fold change ratio to be used
Kinase Cutoff	Same as Ratio	Cutoff for fold change ratio to be used for Kinase/Phosphatase phospho-peptides
Correlation Cutoff	0.8-0.9	Correlation cutoff for resulting network. Default is 0.9 for Pearson, 0.8 for other
Correlation P-value Cutoff	0.05	Correlation significance cutoff
Minimum Samples	3	Minimum samples in which peptides need to appear to be considered
Merge Method	MAX	Choose max or average value when phospho-peptides are merged
Method	Spearman	Choose correlation index among Spearman, Pearson and Kendall tau.
Motif file	NONE	Upload a file with additional motifs to search for
Minimum Paths	5	Minimum number of proteins in Pathway or GO term to consider for enrichment analysis

Table S3. List of output files provided by SELPHI

File	Analysis	Description
.idsnotmapped	Pre-processing	Proteins that SELPHI couldn't match to a UniprotID
.idsmapped	Pre-processing	IDs maps to UniprotID, GeneID,GeneName, KEGG_ID, KEGG pathways, SMART domains, Pfam domains, Transcription Factor families, GO terms, Ensembl Gene and Protein (If ID map was not provided)
.phosmapped	Pre-processing	Maps of phospho-sites (if phospho-site map was not provided)
.sitesnotfound.tsv	Pre-processing	Phospho-sites that were not mapped onto the sequence (tagged 'not found', nf)
.datatable.tsv	Pre-processing	Ratio cutoff filtered table with the sites mapped to gene name and phosphosite location
.pathways.pvalues.tsv	Pathway enrichment	pvalues, odds ratio and corrected pvalue of pathways
.pathmtxoddsratio.tsv .pathmtx0304oddsratio.tsv .pathmtx05oddsratio.tsv	Pathway enrichment	odds ratio of pathways that made the pvalue cutoff (0.05) 0304: signaling pathways, 05: disease pathways
.pathmtx.tsv .pathmtx0304.tsv .pathmtx05.tsv	Pathway enrichment	cumulative log change from input data of pathways that made the pvalue cutoff (0.05) 0304: signaling pathways, 05: disease pathways
.pathdata.pdf .pathdata0304.pdf .pathdata05.pdf	Pathway enrichment	plot of clustered pathway enrichment with the pathmtx values represented in color and the oddsratio in size of cells 0304: signaling pathways, 05: disease pathways
.gotermsbpmtxoddsratio.tsv .gotermsmfmtxoddsratio.tsv .gotermsccmtxoddsratio.tsv	GO terms enrichment	log odds ratio of Biological Process (BP), Molecular Function (MF), Cellular component (CC) GO terms (pvalue <0.05)
.gotermsbpmtx.tsv .gotermsmfmtx.tsv .gotermsccmtx.tsv	GO terms enrichment	cumulative log change from input data of BP,MF,CC GO terms that made the pvalue cutoff (0.05)

.gotermsbp.pdf .gotermsmf.pdf .gotermscc.pdf	GO terms enrichment	plot of clustered XX (BP/MF or CC) GO terms enrichment with the gotermsXXmtx values represented in color and the XX mtxoddsratio in size of cells
.correlations.tsv	Correlation	Phospho-profile correlation pairs: overall, only overlapping data points, correlation p-value
.allinfo.tsv	Correlation & Data integration	Identified associations between kinase/phosphatases-phosphopeptides and annotations for these connections (known phosphosites, motifs, substrate similarity, pathway, genemania co-function)
.stringoverlap.tsv .stringoverlapgenes.tsv	Correlation & Data integration	Known (STRING database >800 or also experimental/database score>800) connections amongst peptides/genes in set and between them and phospho-S/T/Y binding domain containing proteins
.allkinsubtable.tsv .tyrkinsubtable.tsv	Correlation	Correlations Table of All/only tyrosine kinases vs associated phospho-peptides
.allkinsubtable.tsv .tyrkinsubtable.tsv	Correlation	Correlations table of all/only tyrosine phosphatases vs associated phospho-peptides
.allkinvssubcluster.pdf .tyrkinvssubcluster.pdf	Correlation	Plots of clustered all/tyrosine only kinases vs associated phospho-peptides
.allphosvssubcluster.pdf .tyrphosvssubcluster.pdf	Correlation	Plots of clustered all/tyrosine only phosphatases vs associated phospho-peptides
.<number>.cluster	Clustering	Cluster groups identified (one file for each group)
<number>.gotermstable	Clustering	Funcassociate output for goterms enrichment for each cluster group
.gomtxcloddsratio.tsv	Clustering	table of log odds ratio for GO terms enrichment in each cluster (top 10 only)
.gomtxclpvalues.tsv	Clustering	table of pvalues for GO terms enrichment in each cluster (top 10 only)
.clusters	Clustering	The assignment of the peptides to clusters
.clusteredpeptides.pdf	Clustering	All sublusters of the peptides
.gotermsclust.pvalues.tsv	Clustering	Pvalues, log odds ratio and adjusted pvalues (only<0.05) for goterms

.gotermsclust.pvaluesweb.tsv	Clustering	Top 10 pvalues <0.05 for goterms per sample
.gocldata.pdf	Clustering	Plot of GO terms enrichment in each sub cluster
.clustersviolins.pdf	Clustering	Plot of phosphoprofiles for each cluster (violin format)
.motifstotaloverrep.tsv	Motifs	Over representation of known motifs in overall dataset
<kinase>.peps	Motifs	List of peptides that the kinase is associated with
.controlpeps	Motifs	Background peptides for use as reference
<kinase>.eps	Motifs	Postscript file with weblogos for the motifs
phosphoinfo.attr	Cytoscape	The attributes file for the fold change of phosphorylation (log transformed)
phosphoinfogenes.attr	Cytoscape	The attributes file for the fold change of phosphorylation (log transformed) of your proteins (phosphopeptide info has been removed: log(fold change) has been averaged or kept the absolutely maximum value depending on your indication in the input (Merge Method, default: max)
type.attr	Cytoscape	The attributes file that specifies if a node is a kinase,phosphatase,transcription factor and what kind
kinkin.net kinkingenes.net	Cytoscape	SELPHI-network between kinases shown as peptides/genes
STkinSTkin.net STkinSTkingenes.net	Cytoscape	SELPHI-network between Serine/Threonine kinases shown as peptides/genes
TYRkinSTkin.net TYRkinSTkingenes.net	Cytoscape	SELPHI-network between Tyrosine kinases and Serine/Threonine shown as peptides/genes
TYRkinTYRkin.net TYRkinTYRkingenes.net	Cytoscape	SELPHI-network between Tyrosine kinases shown as peptides/genes
kinphosTF.net kinphosTFgenes.net		SELPHI-network between kinases/phosphatases and transcription factors shown as peptides/genes
Ykinsubstr.net Ykinsubstrgenes.net	Cytoscape	SELPHI-network between Tyrosine kinases and their associated peptides shown as peptides/genes

kinkinphos.net kinkinphosgenes.net	Cytoscape	SELPHI-network amongst kinases and phosphatases shown as peptides/genes
stringoverlap.net stringoverlapgenes.net	Cytoscape	network as described above shown as peptides/genes

Table S4a. GO terms enrichment (Biological Process only) of hits associated with Group 1 kinases of Figure S2. The terms have been curated to remove redundancy.

Log(odds ratio)	**P-value**	**attrib name**
1.23	0.002	actin filament capping
1.08	0.006	mitotic nuclear envelope disassembly
1.06	0.008	nuclear envelope disassembly
0.95	0.011	cellular component disassembly involved in execution phase of apoptosis
0.93	0.006	regulation of protein depolymerization
0.91	<0.001	establishment or maintenance of cell polarity
0.84	0.025	regulation of protein complex disassembly
0.84	0.002	negative regulation of cytoskeleton organization
0.79	<0.001	mRNA transport
0.77	0.031	regulation of microtubule cytoskeleton organization
0.71	<0.001	epidermal growth factor receptor signaling pathway
0.7	<0.001	ERBB signaling pathway
0.67	0.024	Ras protein signal transduction
0.67	<0.001	microtubule cytoskeleton organization
0.64	0.001	nucleocytoplasmic transport
0.63	<0.001	neurotrophin TRK receptor signaling pathway
0.62	<0.001	Fc receptor signaling pathway
0.53	0.006	regulation of Ras GTPase activity
0.53	<0.001	response to growth factor
0.51	0.034	immune response-regulating cell surface receptor signaling pathway
0.46	<0.001	viral process
0.44	0.031	regulation of cell cycle process
0.34	0.031	cell death

Table S4b. GO terms enrichment (Biological Process only) of hits associated with Group 2 kinases of Figure S2.

LOD	P_adj	attrib name
2.03	0.049	regulation of vitamin D receptor signaling pathway
1.02	0.001	Fc-gamma receptor signaling pathway involved in phagocytosis
1.01	0.001	Fc receptor mediated stimulatory signaling pathway
0.61	<0.001	cytoskeleton organization
0.6	<0.001	transmembrane receptor protein tyrosine kinase signaling pathway
0.48	0.007	enzyme linked receptor protein signaling pathway
0.44	0.05	regulation of cell cycle
0.38	<0.001	organelle organization

Table S4c. GO terms enrichment (Biological Process only) of hits associated with Group 3 kinases of Figure S2.

Log(odds ratio)	P-value	attrib name
1.17	0.016	regulation of microtubule polymerization or depolymerization
0.9	0.011	mRNA transport
0.88	0.014	chromatin remodeling
0.84	0.034	RNA transport
0.57	0.006	cytoskeleton organization
0.53	0.014	cell cycle
0.46	<0.001	organelle organization

Table S5. Serine/Threonine kinase associations with transcription factor peptides

TF family	kinase	TF
TF: CSD	MAPK14_Y182	CARHSP1_S52
TF: CSD	MAPK1_Y187	CARHSP1_S52
TF: CSD	MAPK3_T202	CARHSP1_S52
TF: CSD	MAPK3_Y204	CARHSP1_S52
TF: CSD	RPS6KA3_T577	CARHSP1_S52
TF: ETS	CAMKK2_S100	ETV6_T18
TF: ETS	CDK12_S685	ETV6_T18
TF: ETS	CDK13_T1058	ETV3_S159
TF: ETS	CDK17_S180	ETV6_T18
TF: ETS	CLK3_S157	ETV6_T18
TF: ETS	MAP3K2_S153	ETV6_T18
TF: ETS	PRKD2_S198	ETV6_T18
TF: ETS	ROCK2_S1374	ETV6_T18
TF: ETS	TTK_S436	ETV3_S159

TF: ETS	ULK1_S623	ETV3_S159
TF: HMG	CDK13_T1058	HMGXB4_S197
TF: HMG	MAP2K2_T394	TOX4_T175
TF: HMG	MARK2_S631	HMGXB4_S197
TF: HMG	TTK_S436	HMGXB4_S197
TF: HMGI/HMGY	CDK18_S117	HMGA1_S44
TF: HMGI/HMGY	CDK18_S117	HMGA1_T42
TF: HMGI/HMGY	CDK18_S117	HMGA1_T53
TF: HMGI/HMGY	CHEK2_S303	HMGA1_S44
TF: HMGI/HMGY	CHEK2_S303	HMGA1_T42
TF: HMGI/HMGY	CHEK2_S303	HMGA1_T53
TF: HMGI/HMGY	CHEK2_S422	HMGA1_S44
TF: HMGI/HMGY	CHEK2_S422	HMGA1_T42
TF: HMGI/HMGY	CHEK2_S422	HMGA1_T53
TF: HMGI/HMGY	CHEK2_T421	HMGA1_S44
TF: HMGI/HMGY	CHEK2_T421	HMGA1_T42
TF: HMGI/HMGY	CHEK2_T421	HMGA1_T53
TF: HMGI/HMGY	MARK3_S469	HMGA1_T53
TF: HMGI/HMGY	STK10_T952	HMGA1_S44
TF: HMGI/HMGY	STK10_T952	HMGA1_T53
TF: HSF	CAMKK2_S100	HSF1_S363
TF: HSF	CDK12_T692	HSF1_S363
TF: HSF	CDK16_S227	HSF1_S363
TF: HSF	CLK3_S157	HSF1_S363
TF: HSF	SIK3_S673	HSF1_S363
TF: MBD	CDK16_S227	MECP2_S228
TF: MBD	CDK18_S14	MBD3_S56
TF: MBD	CIT_S2035	BAZ2A_S1397
TF: MBD	MELK_S529	MBD3_S56
TF: MBD	SIK3_S673	MECP2_S228
TF: MYB	CDK12_T692	SMARCC1_S330
TF: MYB	CDK16_S217	SMARCC1_S330
TF: MYB	CDK18_S117	TERF2_S323
TF: MYB	CHEK2_S303	TERF2_S323
TF family	kinase	TF
TF: MYB	CHEK2_S422	TERF2_S323
TF: MYB	CHEK2_T421	TERF2_S323
TF: MYB	MELK_S529	MYSM1_S110
TF: P53	CDK18_S117	TP53_S315
TF: Retinoicacidreceptor	CDK18_S117	SF1_S214
TF: SRF	CAMKK2_S100	MEF2D_S251
TF: SRF	CDK16_S227	MEF2A_S255

TF: SRF	CLK3_S157	MEF2D_S251
TF: SRF	PRKCD_S304	MEF2D_S180
TF: SRF	PRKCD_S304	MEF2D_S231
TF: SRF	PRKCD_S304	MEF2D_S251
TF: SRF	ROCK2_S1374	MEF2D_S180
TF: SRF	ROCK2_S1374	MEF2D_S231
TF: SRF	ROCK2_S1374	MEF2D_S251
TF: SRF	SIK3_S673	MEF2A_S255
TF: STAT	CDK18_S117	STAT3_S727
TF: STAT	MAPK1_Y187	STAT3_S727
TF: STAT	MAPK3_Y204	STAT3_S727
TF: STAT	RPS6KA3_T577	STAT3_S727
TF: TF_bZIP	CAMKK2_S100	JUN_S63
TF: TF_bZIP	CDK12_T692	JUN_S63
TF: TF_bZIP	CDK16_S227	JUN_S63
TF: TF_bZIP	PAK1_S204	JUN_S63
TF: TF_bZIP	SIK3_S673	JUN_S63
TF: ZBTB	CDK12_T692	ZBTB7A_S525
TF: ZBTB	CDK16_S217	ZBTB7A_S525
TF: ZBTB	CLK3_S157	ZBTB7A_S525
TF: ZBTB	MAP3K2_S153	ZBTB7A_S525
TF: ZBTB	MAPK3_T202	ZNF295_S435
TF: ZBTB	MAPK3_Y204	ZNF295_S435
TF: ZBTB	TNIK_S707	ZNF295_S411
TF: bHLH	CDK13_T1058	TFE3_S556
TF: bHLH	MAP2K2_T394	TFE3_S556
TF: bHLH	MARK2_S631	TCF3_S379
TF: bHLH	MARK2_S631	TFE3_S556
TF: bHLH	TTK_S436	TFE3_S556
TF: zf-C2H2	AKT1_S122	ZNF217_S407
TF: zf-C2H2	CAMKK2_S100	ZNF444_S232
TF: zf-C2H2	CDC42BPG_S1482	ZNF800_S336
TF: zf-C2H2	CDK12_T692	WIZ_S294
TF: zf-C2H2	CDK12_T692	ZFP91_S82
TF: zf-C2H2	CDK12_T692	ZNF444_S232
TF: zf-C2H2	CDK16_S217	WIZ_S294
TF: zf-C2H2	CDK16_S217	ZNF644_S1189
TF: zf-C2H2	CDK16_S227	ZNF444_S232
TF: zf-C2H2	CLK3_S157	ZNF444_S232
TF: zf-C2H2	MARK2_S631	WIZ_S289
TF: zf-C2H2	PAK1_S204	ZFP91_S82
TF: zf-C2H2	PAK1_S204	ZNF444_S232

TF family	kinase	TF
TF: zf-C2H2	PAK2_S197	ZNF768_S83
TF: zf-C2H2	PAK2_S197	ZNF800_S336
TF: zf-C2H2	SIK3_S673	ZFP91_S82
TF: zf-C2H2	SIK3_S673	ZNF444_S232
TF: zf-C2H2	STK39_T354	ZKSCAN1_S208
TF: zf-C2H2	TNIK_S678	ZKSCAN1_S208
TF: zf-C2H2	TNIK_S707	ZKSCAN1_S208
TF: zf-C2H2	TNIK_S707	ZNF217_S407
TF: zf-C2H2	TTK_S436	WIZ_S289
TF: zf-GATA	CDK18_S117	MTA1_S576
TF: zf-GATA	CHEK2_S303	MTA1_S576
TF: zf-GATA	CHEK2_S422	MTA1_S576
TF: zf-GATA	CHEK2_T421	MTA1_S576
TF: zf-GATA	MARK3_S469	MTA1_S576
TF: zf-GATA	STK10_T952	MTA1_S576
TF: zf-NF-X1	CAMKK2_S100	NFX1_S50
TF: zf-NF-X1	CDK12_S685	NFX1_S50
TF: zf-NF-X1	CDK17_S180	NFX1_S50
TF: zf-NF-X1	CLK3_S157	NFX1_S50
TF: zf-NF-X1	MAP3K2_S153	NFX1_S50
TF: zf-NF-X1	ROCK2_S1374	NFX1_S50

Table S6. List of peptides associated with MAPK1

Phospho-site	Corre-lation	Modified peptide	Aligned sequence
NEDD4L_S448	0.984	SLpSSPTVTLSAPLEGAK	XXXSLSSPTVT
CLASP2_S1113	0.998	NTGNGTQSSMGpSPLTRPTPR	QSSMGSPLTRP
PFKFB2_S466	0.973	RRPpSAASLMLPC	XXRRPSAASLM
SH3KBP1_S230	0.999	pSIEVENDFLPVEK	XXXXXSIEVEN
NDRG1_T366	0.984	SHpTSEGAHLDITPNSGAAGNSAGPK	XXXSHTSEGAH
EHBP1L1_S1273	0.958	AHGpSFSHVR	XXAHGSFSHVR
C10orf47_S215	0.996	MAGNEALSPTpSPFR	ALSPTSPFRXX
TNKS1BP1_S691	0.962	WLDDLLApSPPPSGGGAR	DDLLASPPPSG
SENP2_S333	0.992	LGpSGSNGLLR	XXXLGSGSNGL
AHNAK_S3426	0.951	VSMPDVELNLKpSPK	ELNLKSPKXXX
PANK2_S169	0.954	ASpSASVPAVGASAEGTRR	XXXASSASVPA
STAT3_S727	0.992	FICVTPTTCSNTIDLPMpSPR	IDLPMSPRXXX
DOCK7_S2098	0.986	AVLPVTCHRDpSFSR	TCHRDSFSRXX
MLLT4_S1718	0.985	LFpSQGQDVSNKVK	XXXLFSQGQDV
SLC9A1_S703	0.983	IGpSDPLAYEPK	XXXIGSDPLAY
TRIM47_S588	0.982	RGGIPApSPIDPFQSR	GGIPASPIDPF

Phospho-site	Corre-lation	Modified peptide	Aligned sequence
MLPH_S337	0.953	ASpSESQGLGAGVR	XXXASSESQGL
CAST_S230	0.968	ELLAKPIGPDDAIDALSSDFTCGpSPTAAGK	DFTCGSPTAAG
KIAA0284_S1179	0.985	AGpSFTGTSDPEAAPAR	XXXAGSFTGTS
ZFR_S1054	0.951	RRDpSDGVDGFEAEGK	XXRRDSDGVDG
NUP153_S614	0.964	EGSVLDILKpSPGFASPK	LDILKSPGFAS
EGFR_T693	0.987	ELVEPLpTPSGEAPNQALLR	LVEPLTPSGEA
SVIL_S547	0.965	pSLSDFTGPPQLQALK	XXXXXSLSDFT
EHBP1L1_S310	0.955	LRKGpSDALRPPVPQGEDEVPK	XLRKGSDALRP
ARHGEF2_S886	0.994	pSLPAGDALYLSFNPPQPSR	XXXXXSLPAGD
DOCK7_S1352	0.969	MNpSLTFKK	XXXMNSLTFKK
DOCK7_S180	0.989	pSMSIDDTPR	XXXXXSMSIDR
AHNAK_S2397	0.981	ISMPDLDLHLKpSPK	DLHLKSPKXXX
SRRM2_S2272	0.977	TPAAAAAMoxNLApSPR	OXNLASPRXXX
C10orf47_S212	0.982	MAGNEALpSPTSPFR	GNEALSPTSPF
CARHSP1_S52	0.967	TRTFpSATVR	XTRTFSATVRX
C10orf47_S43	0.983	SRpSFTLDDESLK	XXXSRSFTLDD
SPECC1L_S832	0.972	RSpSTSSEPTPTVK	XXXRSSTSSEP
RAPH1_S610	0.982	MESMNRPYTSLVPPLpSPQPK	LVPPLSPQPKX
NUP98_S623	0.992	NLNNSNLFSPVNRDSENLApSPSEYPENGER	SENLASPSEYP
KANK2_S548	0.958	ERVPpSVAEAPQLRPAGTAAAK	XERVPSVAEAP
RICTOR_S1302	0.959	RAQpSLKAPSIATIK	XXRAQSLKAPS
AHNAK_S135	0.965	LKpSEDGVEGDLGETQSR	XXXLKSEDGVE
AHNAK_S5110	0.986	FKAEAPLPpSPK	EAPLPSPKXXX
DOCK1_S1704	0.964	RNpSKHQEIFEK	XXXRNSKHQEI
TPR_S2155	0.957	TDGFAEAIHpSPQVAGVPR	AEAIHSPQVAG
GTSE1_S186	0.995	LLApSSPALPSSGAQAR	XXLLASSPALP
DLG5_S1666	0.953	RLpSMSEVKDDNSATK	XXXRLSMSEVK
CDCA5_S75	0.997	RIVAHAVEVPAVQpSPR	VPAVQSPRXXX
PALM2-AKAP2_S951	0.993	TLpSMIEEEIR	XXXTLSMIEEE
SMG7_S735	0.998	AVPALGKpSPPHHSGFQQYQQADASK	PALGKSPPHHS
KIF16B_S662	0.965	pSFHIENK	XXXXXSFHIEK
METTL1_S27	0.975	AHpSNPMADHTLR	XXXAHSNPMAD
RANBP2_T1396	0.987	ELVGPPLAETVFpTPKTpSPENVQDR	AETVFTPKTSP
SPTBN1_T2328	0.961	AQpTLPTSVVTITSESSPGKR	XXXAQTLPTSV
CRTC2_S433	0.967	VPLpSPLSLLAGPADAR	XXVPLSPLSLL
DAP_S3	0.998	SpSPPEGKLETK	XXXXSSPPEGK
PLEC_S4276	0.990	SSpSVGSSSSYPISPAVSR	XXXSSSVGSSS
SART1_S448	0.996	RVpSEVEEEKEPVPQPLPSDDTR	XXXRVSEVEEE
FAM63A_S489	0.969	VLpSLQGR	XXXVLSLQGRX

Phospho-site	Corre-lation	Modified peptide	Aligned sequence
FAM122B_S134	0.974	RIDFTPVpSPAPpSPTR	DFTPVSPAPSP
PPFIA1_T1159	0.963	GLAAGSAEpTLPANFR	AGSAETLPANF
TNS3_S1149	0.978	ASEAApSPLPDSPGDKLVIVK	ASEAASPLPDS
FAM122A_S62	0.963	RNpSTTFPSR	XXXRNSTTFPS
DENND4C_S1042	0.975	RSpSLPLDHGSPAQENPESEK	XXXRSSLPLDH
RBM14_S618	0.959	RLpSESQLSFR	XXXRLSESQLS
ADD1_S12	0.983	AAVVTpSPPPTTAPHKER	AAVVTSPPPTT
LARP1_T299	0.999	KFDGVEGPRpTPK	VEGPRTPKXXX
MACF1_S1376	0.976	MLpSSSDAITQEFMDLR	XXXMLSSSDAI
FAM129B_S696	0.952	AAPEASpSPPApSPLQHLLPGK	APEASSPPASP
LARP1_T572	0.955	ILIVTQpTPHYMR	LIVTQTPHYMR
ACIN1_S561	0.954	RApSHTLLPSHR	XXXRASHTLLP
CAPZB_S2	0.952	pSDQQLDCALDLMR	XXXXXSDQQLD
ADAM17_T735	0.967	IIKPFPAPQpTPGR	FPAPQTPGRXX
NDRG1_T328	0.955	SRpTApSGSSVTSLDGTR	XXXSRTASGSS
STX12_S139	0.952	ARAGpSRLSAEER	XARAGSRLSAE
STMN1_S25	0.994	RApSGQAFELILpSPR	XXXRASGQAFE
PLEKHM1_S435	0.971	LVVSSPTpSPK	VSSPTSPKXXX
PIK3C2A_S259	0.957	VSNLQVpSPK	SNLQVSPKXXX
CDC42EP4_S174	0.990	RNGAAGPHpSPDPLLDEQAFGDLTDLPVVPK	AAGPHSPDPLL
TBC1D4_T642	0.965	AHpTFSHPPSSTK	XXXAHTFSHPP
GIGYF2_S30_S26	1.000	ALSSGGSITpSPPLpSPALPK	GGSITSPPLSP
SYTL4_S289	0.996	SVIDLRPEDVVHESGpSLGDR	VHESGSLGDRX
TRIM33_S1119	0.950	LKpSDERPVHIK	XXXLKSDERPV
PFKFB2_S483	0.970	NYpSVGSRPLKPLSPLR	XXXNYSVGSRP
AKAP1_S151	0.972	SIPLECPLSpSPK	ECPLSSPKXXX
NCBP1_S22	0.971	KTpSDANETEDHLESLICK	XXXKTSDANET
UBAP1_S210	0.987	VLpSPPHIK	XXXVLSPPHIK
RPS6KA3_T577	0.952	AENGLLMpTPCYTANFVAPEVLKR	NGLLMTPCYTA
SHC1_S139	0.983	HGpSFVNKPTR	XXXHGSFVNKP
ARHGEF12_T703	0.973	QVGETSAPGDTLDGpTPR	DTLDGTPRXXX
AHNAK_S511	0.977	ISMQDVDLSLGpSPK	DLSLGSPKXXX
BOD1L1_S1077	0.971	RGpSLSQEMAKGEEK	XXXRGSLSQEM
FAM21A_S288	0.959	pSRPTpSFADELAAR	XXXXXSRPTSF
SVIL_T657	0.981	FRpTQPITSAER	XXXFRTQPITS
AHNAK_S5762	0.961	ASLGSLEGEAEAEApSSPKGK	AEAEASSPKGK
ZMYND8_S445	0.958	RIpSLSDMPR	XXXRISLSDMP
NUP153_S516	0.972	VQMTpSPSSTGSPMFK	XVQMTSPSSTG
CD97_S831	0.978	ALRApSESGI	XALRASESGIX

Phospho-site	Corre-lation	Modified peptide	Aligned sequence
AHNAK_T4100	0.989	ADIDVpSGPKVDIDTPDIDIHGPEGK	ADIDVSGPKVD
PANK2_S189	0.991	LGpSYSGPTSVSR	XXXLGSYSGPT
CD44_S697	0.981	LVINSGNGAVEDRKPpSGLNGEASK	EDRKPSGLNGE
AHNAK_T4564	0.966	VGIDpTPDIDIHGPEGK	XVGIDTPDIDI
PHLDB1_S1017	0.977	SALLTQNGTGpSLPR	QNGTGSLPRXX
SEC22B_T140	0.955	NLGpSINTELQDVQR	XXNLGSINTEL
CDK13_T1058	0.959	TNpTPQGVLPSSQLK	XXXTNTPQGVL
CABLES1_T415	0.985	RNpTIDSTSSFSQFR	XXXRNTIDSTS
PPP1R12A_S507	0.953	LApSTSDIEEK	XXXLASTSDIE
MYO9B_S1354	0.985	RTpSFSTSDVSK	XXXRTSFSTSD
GIT1_S605	0.965	HGpSGADpSDYENTQSGDPLLGLEGKR	XXXHGSGADSD
NEK9_T333	0.978	SSpTVTEAPIAVVTSR	XXXSSTVTEAP
MAP2K2_T394	0.962	LNQPGpTPTRTAV	LNQPGTPTRTA

Figure S1. Biological Process GO term enrichment analysis for phospho-peptides changed in the EGFR case study(20).

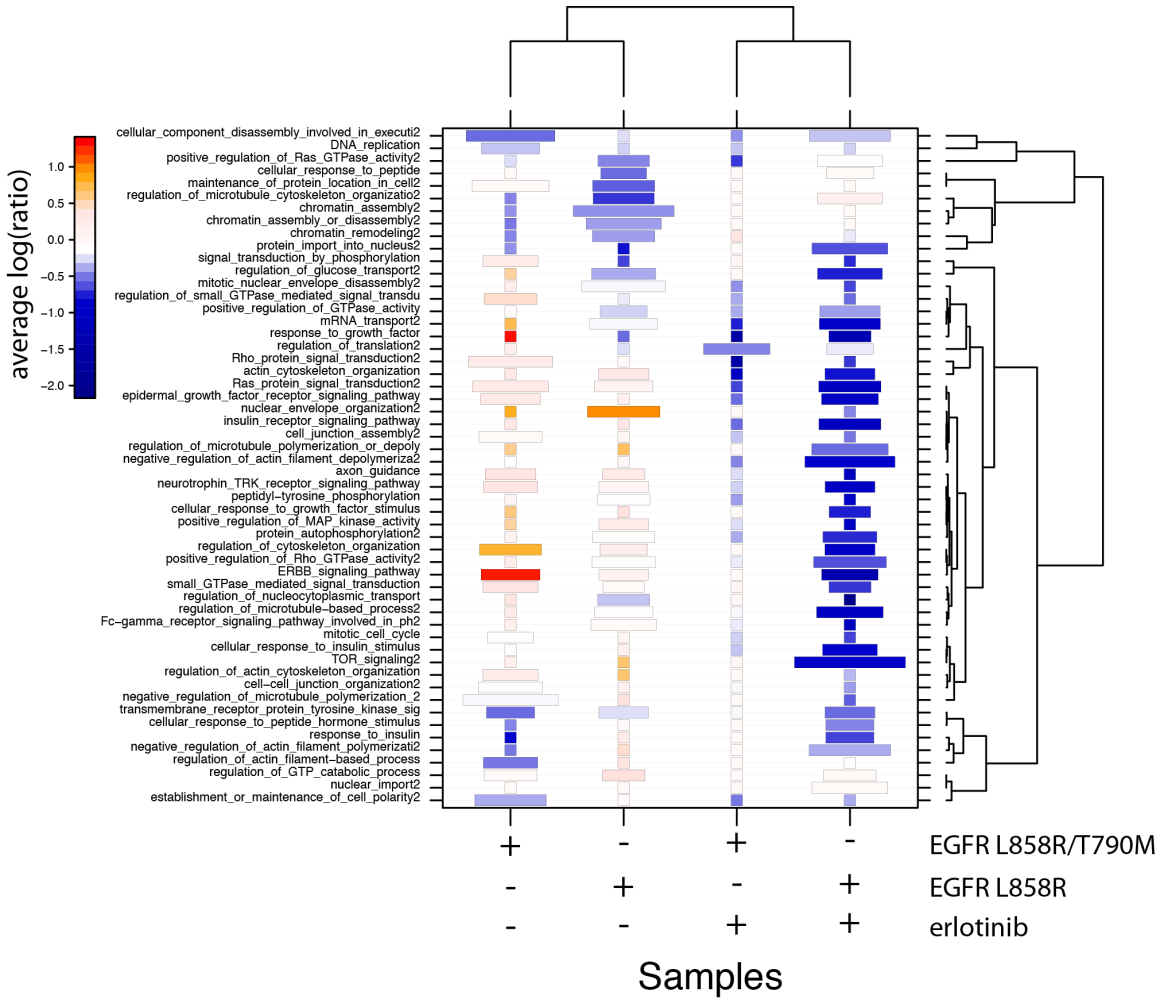


Figure S2. SELPHI-extracted network amongst kinases in the EGFR case study.

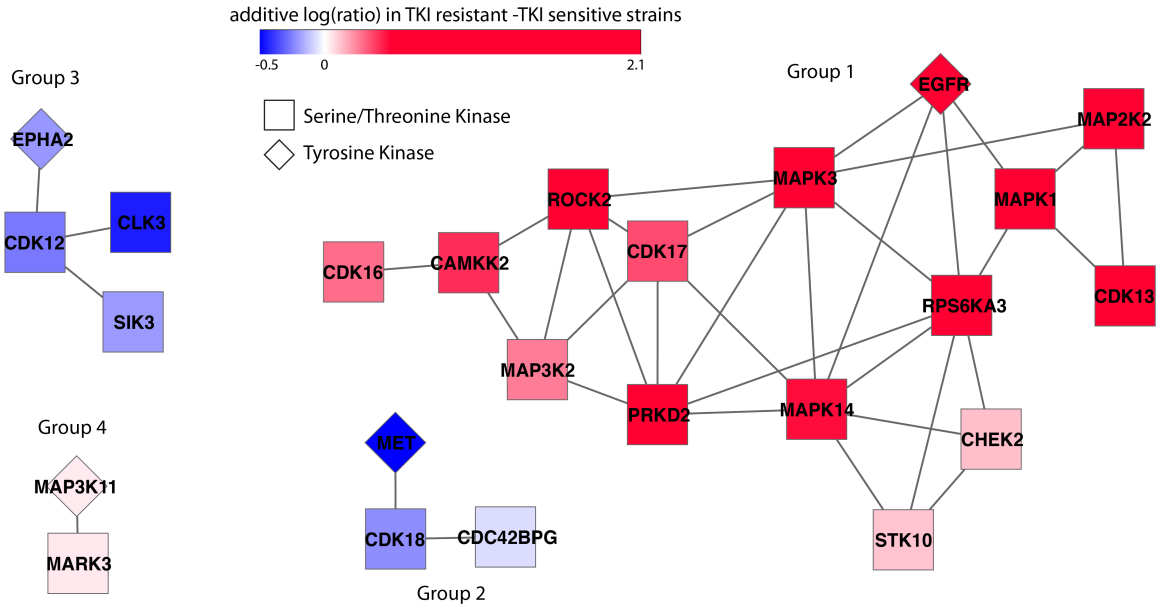
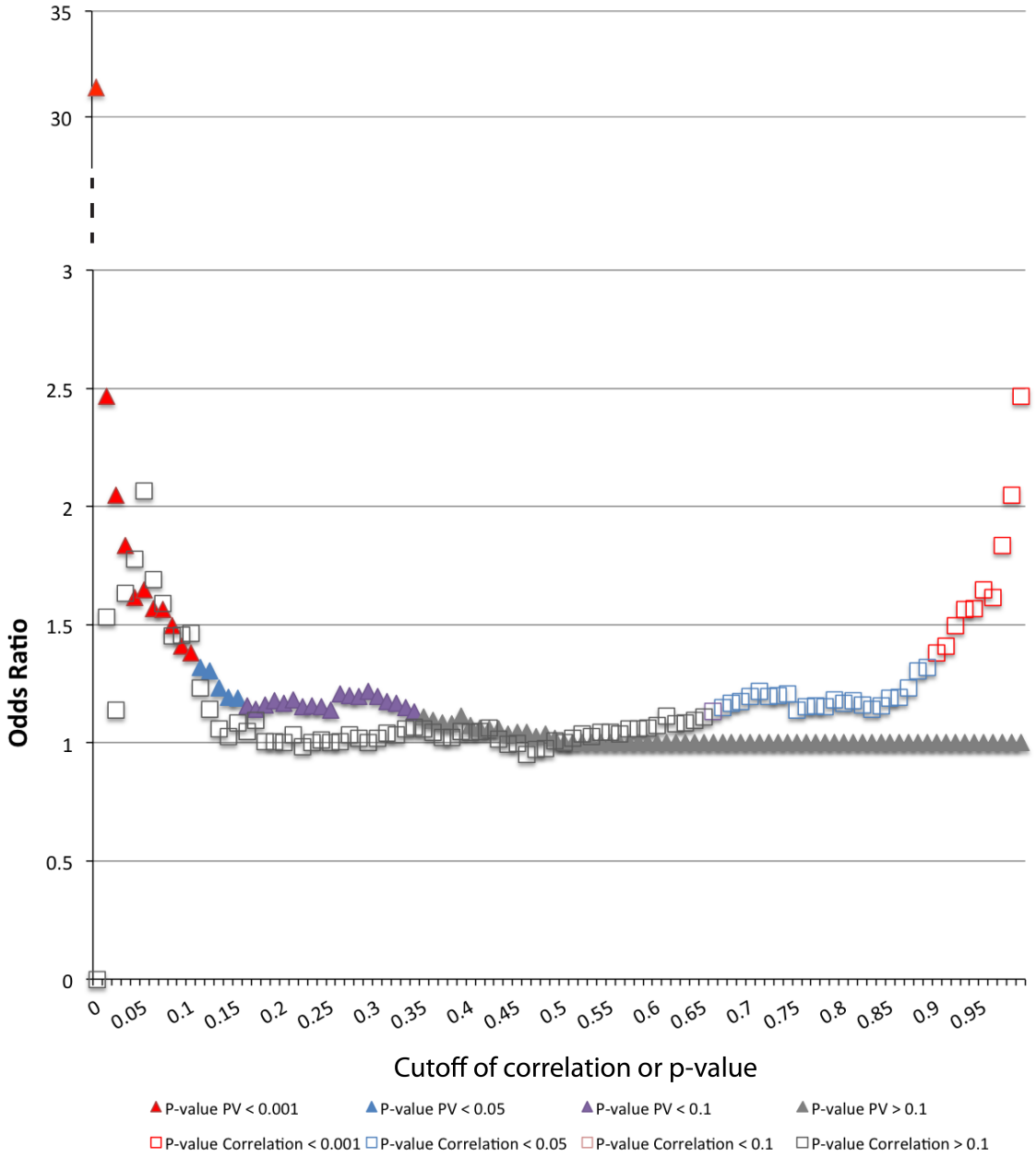


Figure S3. Demonstration of significant enrichment of known or predicted kinase-substrate associations in higher correlation and lower correlation p-value values for the case study dataset.



Supplementary Note 1.
SELPH-Convert Tool

As an accompanying tool we have developed SELPH-Convert. The tool takes as input the report (tab-delimited or comma separated) from any phospho-proteomics analysis tool, or files from publications that describe the data and converts them to SELPHI-useable format. The user is at minimum expected to define the columns in that file that contain the Proteins, Peptides, Phospho-site location (if the peptides didn't have annotations about the phosphorylated residues) and the columns containing the phospho-proteomics data. If the user wishes, they can indicate which columns describe the Intensity and Score of the peptides. To generate the 'phosphomap' files, SELPH-Convert requires the definition of the column that has at least the start location of the peptide or the global sequence location of the first phosphosite appearing on the peptide. Finally to generate the 'idmap' SELPH-Convert requires the definition of the species, and allows the indication of columns (beyond the Protein column) that could help with mapping, such as protein description or Gene Name columns. The resulting '.dat' files can be given as input to SELPHI and if the 'phosmap' and 'idmaps' were successfully generated they can also be used. Alternatively, the user can provide the database used to search for their phospho-peptides and SELPHI will try to generate these files.

Supplementary Note 2.
Correlation and p-value calculations

If the user provides a very small number of data points e.g. 3-5, the statistical power of SELPHI to calculate individual p-values for the correlated peptide pairs is limited, as is the quality of the overall results. SELPHI nevertheless still proceeds with the exploratory analysis so that the user can get an overview of their dataset and be pointed to hypotheses worthy of further experiments. The value of SELPHI even in these 'low-power' settings is supported by our example with only four data points. Here we still see a marked enrichment of true kinase/substrate associations despite low statistical power (Main text and Figure S3). If preferred, the user can set the p-value cutoff to 1 so that significance testing will not be considered at all in an exploratory analysis.

Supplementary Note 3.
Discussion of serine/threonine kinase SELPHI network

Figure S2 shows the correlation-based network among the kinases in the EGFR case study. We calculate a 'change' value by adding the log(ratios) of all the phospho-peptides found for each gene in each condition and color the nodes according to the difference of the 'change' value for each gene between the strains expressing the TKI resistant $EGFR^{L858R/T790M}$ gene, and the TKI sensitive $EGFR^{L858R}$ gene, in the presence of erlotinib.

Four subgroups, potentially representing signaling branches, are apparent. Group one is mostly colored red, indicating that in the resistant strain these proteins are either unaffected or increase their phosphorylation, whilst in the sensitive strain the phosphorylation is reduced. Many kinases exhibit involvement in MAP kinase signaling. We also find ROCK2, which regulates the cytoskeleton and cell polarity. The major GO terms enriched (21) with the downstream proteins associated with these kinases (Table S4a) involve receptor tyrosine kinase signaling, nuclear envelope disassembly, apoptosis and cytoskeletal organization.

Group 2 starts from the tyrosine kinase MET and seems to be much less phosphorylated in the resistant strain after treatment with erlotinib. MET is known to be up-regulated and to contribute to resistance of the sensitive strain in the presence of erlotinib, therefore we observe this difference because its phosphorylation is increased in the sensitive strain and reduced or unchanged (relative to no treatment) in the resistant strain. CDK18 and CDC42BPG kinases may be responsible for the activation of downstream pathways to compensate for the erlotinib treatment. Indeed we find cytoskeletal rearrangement and cell cycle regulation GO terms to be enriched amongst the downstream targets of these kinases (Table S4b).

Group 3 starts from the tyrosine kinase EPHA2 and shows similar results as Group 2. In addition to cytoskeletal rearrangement and cell cycle regulation the downstream associated proteins are also enriched for RNA transport (Table S4c), a function that has also been seen in other studies (22).Group 3 starts from the tyrosine kinase EPHA2 and shows results similar to those of Group 2. In addition to cytoskeletal rearrangement and cell cycle regulation the downstream associated proteins are also enriched for RNA transport (Table S4c), a function that has also been seen in other studies (22).

Supplementary Note 4.
Results interpretation

SELPHI networks provide potential associations of kinases and phosphatases with a specific subset of identified phospho-peptides. These associations are enriched for kinase-substrate relationships, and therefore are a great place to look for such relationships, however a large fraction of associations may represent peptides simply co-changing because of upstream effects. It can be considered as a representation of the data similar to peptide clustering, providing groups of potentially co-functioning peptides, but focusing mainly on likely ‘drivers’ of the signaling process, i.e. kinases and phosphatases, making it easier to form hypotheses about the flow of signaling and to design follow up experiments to uncover the mechanistic details behind it.

References

1. Hornbeck, P.V., Zhang, B., Murray, B., Kornhauser, J.M., Latham, V. and Skrzypek, E. (2014) PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res*.
2. Dinkel, H., Chica, C., Via, A., Gould, C.M., Jensen, L.J., Gibson, T.J. and Diella, F. (2011) Phospho.ELM: a database of phosphorylation sites--update 2011. *Nucleic Acids Res*, **39**, D261-267.
3. Gnad, F., Gunawardena, J. and Mann, M. (2011) PHOSIDA 2011: the posttranslational modification database. *Nucleic Acids Res*, **39**, D253-260.
4. Yang, C.Y., Chang, C.H., Yu, Y.L., Lin, T.C., Lee, S.A., Yen, C.C., Yang, J.M., Lai, J.M., Hong, Y.R., Tseng, T.L. *et al.* (2008) PhosphoPOINT: a comprehensive human kinase interactome and phospho-protein database. *Bioinformatics*, **24**, i14-20.
5. Duan, G., Li, X. and Kohn, M. (2015) The human DEPhOsphorylation database DEPOD: a 2015 update. *Nucleic Acids Res*, **43**, D531-535.
6. Lachmann, A. and Ma'ayan, A. (2009) KEA: kinase enrichment analysis. *Bioinformatics*, **25**, 684-686.
7. Wong, Y.H., Lee, T.Y., Liang, H.K., Huang, C.M., Wang, T.Y., Yang, Y.H., Chu, C.H., Huang, H.D., Ko, M.T. and Hwang, J.K. (2007) KinasePhos 2.0: a web server for identifying protein kinase-specific phosphorylation sites based on sequences and coupling patterns. *Nucleic Acids Res*, **35**, W588-594.
8. Linding, R., Jensen, L.J., Pasculescu, A., Olhovsky, M., Colwill, K., Bork, P., Yaffe, M.B. and Pawson, T. (2008) NetworKIN: a resource for exploring cellular phosphorylation networks. *Nucleic Acids Res*, **36**, D695-699.
9. Linding, R., Jensen, L.J., Ostheimer, G.J., van Vugt, M.A., Jorgensen, C., Miron, I.M., Diella, F., Colwill, K., Taylor, L., Elder, K. *et al.* (2007) Systematic discovery of in vivo phosphorylation networks. *Cell*, **129**, 1415-1426.
10. Horn, H., Schoof, E.M., Kim, J., Robin, X., Miller, M.L., Diella, F., Palma, A., Cesareni, G., Jensen, L.J. and Linding, R. (2014) KinomeXplorer: an integrated platform for kinome biology studies. *Nat Methods*, **11**, 603-604.
11. Bennetzen, M.V., Cox, J., Mann, M. and Andersen, J.S. (2012) PhosphoSiteAnalyzer: a bioinformatic platform for deciphering phospho proteomes using kinase predictions retrieved from NetworKIN. *J Proteome Res*, **11**, 3480-3486.
12. 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**, 3635-3641.
13. Xue, Y., Ren, J., Gao, X., Jin, C., Wen, L. and Yao, X. (2008) GPS 2.0, a tool to predict kinase-specific phosphorylation sites in hierarchy. *Mol Cell Proteomics*, **7**, 1598-1608.
14. Wang, L., Hou, L., Qian, M. and Deng, M. (2012) Integrating phosphorylation network with transcriptional network reveals novel functional relationships. *PLoS One*, **7**, e33160.

15. Chen, W.M., Danziger, S.A., Chiang, J.H. and Aitchison, J.D. (2013) PhosphoChain: a novel algorithm to predict kinase and phosphatase networks from high-throughput expression data. *Bioinformatics*, **29**, 2435-2444.
16. Huang, K.Y., Wu, H.Y., Chen, Y.J., Lu, C.T., Su, M.G., Hsieh, Y.C., Tsai, C.M., Lin, K.I., Huang, H.D. and Lee, T.Y. (2014) RegPhos 2.0: an updated resource to explore protein kinase-substrate phosphorylation networks in mammals. *Database (Oxford)*, **2014**, bau034.
17. Saez-Rodriguez, J., Alexopoulos, L.G., Epperlein, J., Samaga, R., Lauffenburger, D.A., Klamt, S. and Sorger, P.K. (2009) Discrete logic modelling as a means to link protein signalling networks with functional analysis of mammalian signal transduction. *Mol Syst Biol*, **5**, 331.
18. Mitsos, A., Melas, I.N., Siminelakis, P., Chairakaki, A.D., Saez-Rodriguez, J. and Alexopoulos, L.G. (2009) Identifying drug effects via pathway alterations using an integer linear programming optimization formulation on phosphoproteomic data. *PLoS Comput Biol*, **5**, e1000591.
19. Zhang, Y., Kweon, H.K., Shively, C., Kumar, A. and Andrews, P.C. (2013) Towards systematic discovery of signaling networks in budding yeast filamentous growth stress response using interventional phosphorylation data. *PLoS Comput Biol*, **9**, e1003077.
20. Zhang, X., Belkina, N., Jacob, H.K., Maity, T., Biswas, R., Venugopalan, A., Shaw, P.G., Kim, M.S., Chaerkady, R., Pandey, A. *et al.* (2014) Identifying novel targets of oncogenic EGF receptor signaling in lung cancer through global phosphoproteomics. *Proteomics*.
21. Berriz, G.F., Beaver, J.E., Cenik, C., Tasan, M. and Roth, F.P. (2009) Next generation software for functional trend analysis. *Bioinformatics*, **25**, 3043-3044.
22. Omerovic, J., Hammond, D.E., Prior, I.A. and Clague, M.J. (2012) Global snapshot of the influence of endocytosis upon EGF receptor signaling output. *J Proteome Res*, **11**, 5157-5166.