# Systematic Functional Prioritization of Protein Post-translational Modifications

Pedro Beltrao, Véronique Albanèse, Lillian R. Kenner, Danielle L. Swaney, Alma Burlingame, Judit Villén, Wendell A. Lim, James S. Fraser, Judith Frydman, Nevan J. Krogan

## Supplemental Figure Legends

Figure S1, related to Figure 2 – **Clustering of lysine modifications and phosphorylation sites**. Phosphoacceptor residues – serine (S), threonine (T) and tyrosine (Y) – were found to be preferentially phosphorylated closer to lysines that were modified, when compared to random (Figure 2, main text). (A) In order to rule out that this was due to a preferential accessibility within unstructured regions of proteins we repeated the analysis excluding all PFAM domain sequences. Solid lines represent observed fraction of phosphorylated residues while the dotted lines represent the average from random samples of a similar number of phosphosites. B) In order to show that the observed clustering of phosphorylation sites with ubiquitylation sites seen in human proteins is conserved in other species we performed a similar analysis in *S. cerevisiae* C) In order to show that the observed clustering of phosphorylation sites with acetylation sites seen in human proteins is conserved in other species we performed a similar analysis in *S. cerevisiae* and *M. musculus*. Acetylation data for these species was obtained from the Global Proteome Machine Database (http://www.thegpm.org). As observed in human, the fraction phosphoacceptor residues that are phosphorylated increases near acetylation sites and is higher than expected by chance when within 15 residues.

Figure S2, related to Figure 3 – **Phosphosites at human interface residues and at proteasome subunit interfaces**. A) We used models of interfaces for human protein-interactions either from xray structures or homology models to define interface residues. Additionally, we used the 3DID database to annotate all human PFAM domain residues that might participate in interactions (PFAM interaction residues). We observed that human phosphorylation sites found at interface residues were found to be more conserved that average sites or sites found within PFAM domains. Additionally, the conservation of the phosphorylation of an interface was twice that of the positional conservation of average phosphosites. The observed conservation of all categories was found to be significantly higher than expected based on a random sampling of an equivalent number of phosphorylation sites with exception for the conservation at the level of the interface. The lack of significance could be, in part, due to the low total number of human interfaces that are currently available for analysis. B and C) Conserved N-terminal

phosphorylation sites in the *S.cerevisiae* alpha-proteasome subunits might regulate proteasome assembly or activity.

Figure S3, related to Figure 5 – **Domain family phosphopeptide enrichment analysis for RRM, PH, Ras, UCH, MSF and Mitochondrial Carrier domains**. All domain instances were aligned to a representative structure sequence and the phosphopeptide positions were transferred by homology. Using a moving window we calculated for each position within the domain family the enrichment over a random distribution of all phosphorylation sites. The significance was accessed by randomly sampling the same number of positions within the domain. The dotted line corresponds, in each case, to a p-value cut-off of 0.005. Positions that had an enrichment value above the cut-off were considered to be significantly enriched for phosphorylation and we denominate these regulatory "hot-spots". Phosphosites from the 11 species found within these regions are provided at ptmfunc.com.

Figure S4, related to Figure 6 – **Phosphorylation hot-spots within the HSP70 domain family**. A) All alanine and phosphomimetic mutants of the Regions 1 and 2 described in the main text were analyzed for growth defects under heat-shock and in a strain deficient in SSA function (ssa1-45Δssa2Δssa3Δssa4). All mutants, with the potential exception of S38D, displayed a growth defect under heat shock conditions that is not observed when the WT SSA1 is expressed or when a control mutation T326A is introduced. B) and C) The association of the indicated SSA1 mutants with polysomes was examined by immunoblot analysis following the fractionation of cell lysates by sedimentation on sucrose gradients. Ribosomal profiles (top) were determined by OD254nm and confirmed by immunoblot analysis of the ribosomal proteins Rpl3p. B) Unlike region 2 mutants, region 1 mutants, including a double phosphomimetic mutant, displayed a wild-type like association with polysomes. C) A phosphomimetic mutation in position S495 is sufficient to disrupt the association of SSA1 with polysomes. The same phenotype is not observed with an alanine mutant.

## Supplemental Tables

Table S1 – Phosphorylation, acetylation and ubiquitylation sites used for the analysis. The data can be accessed via a website (http://ptmfunc.com). For each species we used PFAM annotations to count the number of PTMs that are found within globular protein domains. On average, ~75% of the phosphosites ~40% of acetylation and ~45% of ubiquitylation sites are found outside PFAM domains.

| Species | PTM type | PTM total | Within PFAM domains | % Outside PFAM domain |
|---|---|---|---|---|
| ***H. sapiens*** | Phosphorylation | 31165 | 11726 | 62.4 |
| | Acetylation | 8042 | 4604 | 42.8 |

| | | | | |
|---|---|---|---|---|
| | Ubiquitylation | 22057 | 11079 | 49.7 |
| ***M. musculus*** | Phosphorylation | 24921 | 6825 | 72.6 |
| | Acetylation | 3384 | 2298 | 32.1 |
| ***R. norvegicus*** | Phosphorylation | 1885 | 913 | 51.6 |
| ***X. laevis*** | Phosphorylation | 470 | 149 | 68.3 |
| ***C. elegans*** | Phosphorylation | 6715 | 1074 | 84.0 |
| ***D. melanogaster*** | Phosphorylation | 17535 | 2081 | 88.1 |
| | Acetylation | 1707 | 858 | 49.7 |
| ***S. pombe*** | Phosphorylation | 2540 | 636 | 75.0 |
| ***S. cerevisiae*** | Phosphorylation | 15144 | 3747 | 75.3 |
| | Acetylation | 657 | 433 | 34.1 |
| | Ubiquitylation | 2499 | 1426 | 42.9 |
| ***C. albicans*** | Phosphorylation | 2910 | 532 | 81.7 |
| ***A. thaliana*** | Phosphorylation | 4527 | 648 | 85.7 |
| ***O. sativa*** | Phosphorylation | 3140 | 633 | 79.8 |

Table S2 – **Estimation of false-discovery rate for the *S. cerevisiae* phosphosite dataset after compilation of 12 different reported experiments**. Although individual proteomic experiments report false discovery rates (FDR) of peptide identification on the order of 2% or less, the accumulation of different independent datasets results in an increase of FDR. We collected 12 different phosphoproteomics studies for *S. cerevisiae* and used these to estimate an upper bound for this increase. Assuming a rate of 2% peptide FDR for each individual dataset and that no false-positive site is identified more than once, we estimate that the combined *S. cerevisiae* phosphoproteome has, at most, an FDR of ~4%. This suggests that even for species that have been extensively studied such as *S. cerevisiae* and *H. sapiens*, the fraction of incorrectly identified PTMs is likely to be low. For each study we estimated the total number of false positive phosphosites assuming a false-discovery rate (FDR) of 2%. We assumed, conservatively, that no false-positive phosphosite is observed more than once. Under these assumptions the total expected number of false positives would be equivalent to 4% of the compiled non-redundant *S. cerevisiae* phosphosites (836 out of 20658).

| Pubmed ID | Phosphosites obtained from study | Projected false-positives (assuming 2% FDR) |
|---|---|---|
| 19823750 | 2876 | 57.52 |
| 17563356 | 6489 | 129.78 |
| 19684113 | 4000 | 80 |
| 17287358 | 1154 | 23.08 |
| 19795423 | 3010 | 60.2 |
| 20377248 | 2526 | 50.52 |
| 15665377 | 591 | 11.82 |
| 19547744 | 3435 | 68.7 |

| | | |
|---|---|---|
| 17330950 | 1386 | 27.72 |
| 21177495 | 3540 | 70.8 |
| 21298081 | 6071 | 121.42 |
| 19779198 | 6744 | 134.88 |
| **Sum** | **41822** | **836.44** |
| **Sum non-redundant** | **20658** | **836.44** |

Table S3 - **Overlap of different post-translational modifications within the same proteins does not depend on protein abundance**. We observed that proteins with different lysine modifications (acetylation, ubiquitylation and sumoylation) are also very likely to be phosphorylated. One possible explanation for this would be an experimental identification bias in MS experiment. Since MS experiments preferentially identify highly abundant proteins then the overlap could be due this bias. In order to control for this we used protein abundance estimates for human proteins (http://pax-db.org/). We excluded all human proteins with estimated abundance over half of the median such that there was no significant difference in the abundance levels of phosphorylated versus non-phosphorylated proteins (p-value=0.44 with a KS ranked test). After controlling for abundance we still see a very significant enrichment of phosphoproteins among the lysine modified proteins over random (using a Fisher's exact test).

| Protein Subset | Total | Phosphoproteins (% from total) | Enrichment over random | p-value for enrichment |
|---|---|---|---|---|
| **All proteins** | 3522 | 1526 (43%) | - | - |
| **Ubiquitylated** | 816 | 539 (66%) | 1.5 | $<10^{-40}$ |
| **Acetylated** | 306 | 237 (77%) | 1.8 | $<10^{-33}$ |
| **Sumoylated** | 46 | 43 (93%) | 2.1 | $<10^{-12}$ |

Table S4 – **Domain families selected for phosphopeptide enrichment analysis**. We used PFAM annotations and the compilation of known phosphorylation sites for the 11 species considered in this study to count the total number of phosphorylation sites found for each PFAM domain. Among the domain families with higher number of total phosphosites across all the species we selected 10 that had an available representative structure deposited in the PDB for analysis. For each of these PFAM domains we provide here the total number of domain instances annotated across the 11 species as well as the total number of phosphosites found for each and the representative structure used for the enrichment analysis.

| PFAM id | Domain name | Total phosphosites | Total number of domains | PDB ID of structure used |
|---|---|---|---|---|
| **Pkinase** | Protein kinase | 1273 | 4269 | 1QMZ |
| **Pkinase_Tyr** | Protein tyrosine kinase | 495 | 1427 | 1M14 |

| | | | | |
|---|---|---|---|---|
| **HSP70** | Heat shock proteins, Hsp70 | 313 | 195 | 1YUW |
| **RRM_1** | RNA recognition motif | 253 | 2438 | 3BS9 |
| **UCH** | Ubiquitin carboxyl-terminal hydrolase | 200 | 390 | 3H0X |
| **Ras** | Ras domain | 190 | 821 | 1EK0 |
| **HSP90** | Heat shock proteins, Hsp90 | 145 | 78 | 2IOP |
| **PH** | Pleckstrin homology domain | 144 | 856 | 1MAI |
| **MFS_1** | Major Facilitator Superfamily (MFS) transporters | 120 | 723 | 2GFP |
| **Mito_carr** | Mitochondrial carrier | 119 | 1434 | 2C3E |

Table S5 – **Total 1-to-1 orthoproteins and phosphosites used in species to human comparative analysis**. We counted the total number of human to species 1-to-1 orthologs with an inparanoid score greater than 90%. We also detailed how many of these are phosphoproteins and how many phosphosites in total were used for the comparative studies throughout the paper.

| Species | Total number of orthologs | Orthologs that are phosphoproteins | Number of phosphosites in orthologs |
|---|---|---|---|
| ***M.musculus*** | 15982 | 5923 | 31317 |
| ***R. norvegicus*** | 14940 | 1059 | 2547 |
| ***X.laevis*** | 8261 | 279 | 394 |
| ***C.elegans*** | 4489 | 1445 | 4309 |
| ***D.melanogaster*** | 5346 | 2279 | 9264 |
| ***S.pombe*** | 2225 | 597 | 1382 |
| ***S.cerevisiae*** | 1992 | 1234 | 6413 |
| ***C.albicans*** | 2129 | 526 | 1481 |
| ***A.thaliana*** | 2888 | 476 | 1036 |
| ***O.sativa*** | 2872 | 464 | 989 |

# Supplemental Experimental Procedures

**Post-translational modification sites, protein disorder, genomes and ortholog assignments**

From each publication we obtained PTM sites that had high localization probability and obtained, whenever possible, counts for the number of times the PTM, modified-peptide or mass-spectra was observed in the experiment. All of the sites compiled are provided in a searchable website. Protein sequences, protein identifiers and ortholog assignments were obtained from the Inparanoid database (inparanoid.sbc.su.se, version 7). We note that for genes having multiple splicing isoforms, Inparanoid retains only the longest form. For the

comparative analysis we considered only 1-to-1 orthologs with Inparanoid confidence scores greater than 90%. One-to-many ortholog assignments were discarded since it has been shown that gene duplication influences the evolution of protein phosphorylation (Amoutzias et al., 2010). Protein disorder predictions were obtained using disEMBL (Linding et al., 2003). The total number of human to species ortholog pairs used in this studied are listed in Table S5.

**Immunoprecipitation, mass spectrometry and ubiquitylation site identification**

*S. cerevisiae* Sub592 (containing a HisTag modified ubiquitin) and Sub62 were grown separately in YPD and harvested during mid-log phase (OD600 ~1.0). Harvested cells were resuspended in lysis buffer containing 8 M urea, 300 mM NaCl, 50 mM Tris pH 8.2, 50 mM NaF, 50 mM sodium β-glycerophosphate, 10 mM sodium pyrophosphate, 1mM sodium orthovanadate, and 1 tablet mini protease inhibitor (Roche Diagnostics, Indianapolis, IN) and lysed by bead beating. Protein extract from Sub592 cells (~40mg) was enriched for ubiquitylated proteins via HisTag purification on a $Co^{2+}$ sepharose resin (Talon Superflow Resin, Clontech Laboratories, Inc., Mountain View, CA). All samples were reduced, alkylated, and diluted 5-fold with 50 mM Tris, pH 8.2. Sub62 proteins and half of the ubiquitin enriched Sub592 protein were digested overnight with trypsin, while the remaining half was supplemented with 10 mM $CaCl_2$ and digested with ArgC. After enzymatic digestion, all three samples were desalted and enriched for diGly containing peptides using a polyclonal antibody as previously described (Cell Signaling, Technology, Danvers, MA) (Kim et al., 2011). Peptides were separated over a linear gradient from ~8%-30% acetonitrile in 0.125% formic acid and injected into an Orbitrap Velos mass spectrometer. We analyzed each sample once using CAD MS/MS spectra collected in the linear ion trap (120min analysis) and once with HCD MS/MS collected in the orbitrap (180min analysis). Raw files were searched with Sequest against the target-decoy (Elias and Gygi, 2007) *S. cerevisiae* protein sequence database (downloaded from SGD Jan 2011). Search parameters included a static modification on cysteine residues (57.02146 Da) and variable modifications of methionine oxidation (15.99491 Da) and diGly on lysine (114.04293 Da). Peptide spectral matches were filtered to a 1% false-discovery rate at the peptide and protein level and diGly sites were localized using a version of the Ascore algorithm that can accept any post-translational modification (Ascore > 13) (Beausoleil et al., 2006).

**Alignments and conservation metrics**

Protein sequence alignments were done with muscle version 3.6 (Edgar, 2004) using standard alignments options. We defined two different conservation metrics: PTM acceptor residue conservation and site (or state) conservation. We consider that the acceptor residue (lysine for ubiquitylation and acetylation and serine, threonine or tyrosine for phosphorylation) is conserved when the aligned residue, opposite to the PTM site under analysis, is identical. We consider that the site (or state) is conserved when we find experimental evidence supporting

the conservation of the PTM site under analysis. To account for errors in the PTM positional assignments we define a PTM site to be conserved if the aligned peptide is also modified within a window of +/-2 alignment positions. In order to estimate a random expectation (i.e. null model) for the conservation of acceptor residues and PTM sites from species A to species B, we randomly shuffled the PTM sites of each protein in species A. When determining the fraction of phosphoacceptor residues phosphorylated as a function of the distance to a modified lysine we define the distance as the number of amino-acids to the closest modified lysine residue. For each fixed distance N from a modified lysine we counted the total number of phosphoacceptor residues (S/T/Y) and phosphorylation sites within a window of 10 amino-acids centered on N.

**Phosphosite kinase preference and phosphosite similarity**

We used the kinase recognition matrices from the study by Mok and colleagues (Mok et al., 2010) to define a set of *S.cerevisiae* phosphosites that match a kinase substrate preference. We assumed that sites that had a matrix score above 2 standard deviations away from the mean score for any given kinase matched that kinase's substrate recognition preference. Using these kinase recognition matrices we also devised a phosphosite similarity metric. For each phosphosite, we calculated the predicted z-scored normalized recognition score for the 63 kinases studied by Mok and colleagues. These 63 scores constitute a vector that describes how likely it is that different types of kinases will phosphorylate a given site. We assume that the correlation score for a pair of vectors will be proportional to the similarity of the corresponding phosphosites. We benchmarked this approach by comparing the similarity of random pairs of phosphosites with pairs that are known to be phosphorylated by the same kinase based on the Phosphogrid database (www.phosphogrid.org) (Stark et al., 2010) (Figure 4B).

**Protein interface models and interface residue assignments**

We used interface models from xray structures and homology models for *S.cerevisiae* and *H.sapiens* and complemented these with docking solutions for *S.cerevisiae* from the work of Mosca and colleagues (Mosca et al., 2009). Homology models for protein-pairs that are known to interact were obtained from the GWIDD database (Kundrotas et al., 2010) with a cut-off of 25% sequence identity between the modeled proteins and template structures. Interface residues for these structure models were assigned using PSAIA (Mihel et al., 2008). Additional interface residues were predicted by homology using the 3DID database of inter-domain contacts for PFAM domains (Stein et al., 2011). 3DID maintains a list of inter-domain interface contact residues. We used HMMER (www.hmmer.org, version 3.0) to annotate the 11 proteomes with PFAM domains and used the 3DID annotations to predict PFAM domain residues that are more likely to participate in protein-protein interactions. For the 3DID annotations (Global Interface file from 13 of March 2011) we considered only PFAM domain

residues that were annotated in 3DID to participate in interactions with, at least, 3 domains or linear motifs.

**Protein complementation assays**

Skp1 and Met30 were fused to fragments F1 and F2, respectively, of a split cytosine deaminase by gap-repair cloning. Point mutants were constructed using PCR with site directed oligonucleotides. Protein complementation was assayed as previously described (Ear and Michnick, 2009; Michnick et al., 2010). Briefly, a direct protein-protein interaction between proteins fused to the two fragments of the split cytosine deaminase allows it to fold and become enzymatically active. The folded cytosine deaminase converts cytosine to uracil, permitting growth on –Ura media, but also converts 5-FC to 5-FUTP, which kills cells on media containing 5-FC. In contrast, when the cytosine deaminase fragments are fused to proteins that do not interact, growth is no longer supported on –Ura media, but the strain is viable on media containing 5-FC. To assay the interaction, 300ng of Met30 and 300ng of Skp1 (WT, S162A, or S162D) plasmids were transformed into 25uL of log phase cells using the LiAc/PEG transformation procedure. Equal volumes were plated on all three media types and transformants were grown for 3 days (-His-Leu, +5-FC) or 5 days (-Ura).

**SKP1-Met30 Co-IP**

Yeast cells expressing endogenous Myc-tagged Met30 (Aghajan et al., 2010) were transformed with a plasmid expressing a Flag-tagged SKP1 S162A or S162D under control of the Gal promoter and selected for Leucine auxotrophy. Cells were grown in –Leu SC liquid media containing 2% raffinose and 2% galactose and harvested at an OD of 1.0. Cell pellets where resuspended in buffer containing 20mM HEPES pH 7.5, 10% glycerol, 200mM NaCl, 1mM EDTA, 1mM DTT, 1mM NaF, and protease inhibitors (Sigma). Pellets were broken by bead beating with glass beads 5 X 1 min with 2 minutes on ice between beatings. Whole cell lysate was centrifuged at 14,000 rpm for 20 minutes, and supernatant was collected and immunopurified on Anti-Flag Magnetic beads (Sigma), which were washed and incubated according to manufacturer's instructions. Co-immunopurified proteins where washed and resuspended in 3M SDS loading Buffer, run on an SDS-PAGE gel and transferred to a PDVF membrane. The membrane was probed with either primary Anti-c-myc produced in chickens (Invitrogen) with Anti-Chicken HRP conjugate (Promega) secondary or with Anti-Flag produced in mouse (Sigma) primary with Goat anti Mouse HRP Light Chain Specific (Jackson ImmunoResearch) secondary. Immunoblots were processed using ECL detection.

**PTM hot-spots for domain families**

Ten domain families that are heavily phosphorylated in most species studied were selected for analysis (Table S8). For each domain family, we selected a representative structure

and transferred the phosphosites, occurring in all instances of this domain across the 11 species, using protein sequence alignments. We then used a sliding window of 10 amino acids and random sampling to identify regions, within each domain family, that are enriched for phosphorylation. Any 10 amino-acid peptide with a significant (p-value<0.005) enrichment of phosphorylation when compared to random was defined a potential regulatory hot-spot. The same analysis was performed using lysine acetylation data for the protein kinase domain.

**DNA, yeast strains**

The SSA1 ORF was amplified by PCR and then cloned into a CEN gateway plasmid containing a GPD promoter. All the mutants were then generated by Quick Change mutagenesis according to the manufacturer's instructions All the plasmids were confirmed by sequencing. *S.cerevisiae* strains were as follows : the ssa1 temperature sensitive strain (mat alpha leu2 trp1 ura3 ade2 his3 lys2, ssa1-45BKD, ssa2::LEU2, ssa3::TRP1, ssa4::LYS2) and was a gift from Betty Craig. The Δssa1::KanMX4 Δssa2::NAT was generated by direct replacement of the SSA2 coding region with the NatMX4 cassette in the single deletion strain Δssa1::KanMX4. The correct insertion of the marker was confirmed by PCR from genomic DNA. Standard yeast protocols were used for yeast transformations.

**Drop test assay**

Cells were grown overnight in selective medium and then diluted to OD600nm O.4. Cells were grown for another 3 hours and then diluted to OD600nm of 0.1. This sample was then subjected to 10-fold serial dilutions. 10µl of each dilution was then spotted onto –URA plates and allowed to grow at 30ºC, 33ºC and 37ºC for 2 days.

**Lysate preparation and ribosome fractionation**

200 ml yeast in exponential growth was treated with 100µg/ml of cycloheximide, harvested, washed with cold water, resuspended in 1ml of Buffer A (20 mM Hepes pH=7.5, 50 mM KCl, 10 mM MgCl2, 1% TritonX-100, 1mM DTT, 100µg/ml cycloheximide and protease inhibitor cocktail) and frozen as drops in liquid nitrogen. The cells were then ground using a grinder (MM301, Retsch). 0.5 ml of lysis buffer A was added to the powder and the lysates were clarified by centrifugation at 14000g for 10 minutes. 20 OD of lysate was loaded on a 12 ml 7%-47% sucrose gradient in buffer A without Triton-X-100 and centrifuged in a SW41 rotor (Beckamn Coulter) for 150 minutes at 39000 rpm at 4ºC. Fractions were collected using a UA/6 detector (ISCO, Inc). The fractions were TCA precipitated and separated by SDS-PAGE and subjected to immunoblot analysis. The Rpl3 antibody was a gift from Jonathan Warner. The monoclonal HA antibody was from Covance.

**Luciferase in vivo refolding assay**

The ssa1-45 ts cells were transformed with firefly luciferase, which is labile above 37 °C, and a plasmid driving the expression of the wild type SSA1 or the phosphomutants. After growth at 30ºC, the cells were shifted to 44ºC for 1 hour, which causes the heat-induced denaturation of luciferase. Cycloheximide was added to 10mg/ml 15 minutes before the end of the heat shock to prevent further expression of luciferase. Cells were then transferred to 30ºC to recover. At different time point during the recovery, aliquots were taken, centifuged and frozen in liquid nitrogen. The luciferase acitivity was measured using the Promega luciferase kit. Recovery is expressed as a percentage of the activity before heat shock treatments.

**Microscopy and aggregation assay**

The ssa1-45 ts strain was transformed with the different SSA1 mutants as well as with the Gal-Ubc9-2-GFP construct. Cells were grown overnight at 30ºC and then diluted to OD600nm 0.3 and induced with 2% galactose for 6 hours. Cells were then shifted to 37ºC for 30 minutes to induce the misfolding of Ubc9-2-GFP. The formation of Ubc9-2-GFP puncta was then examined by fluorescence microscopy using a microscope (Axiovert; Carl Zeiss, Inc) equipped with a 100x NA 1.3 oil immersion objective lens and a digital camera controlled with AxioVision software. Images were then prepared using Photoshop.

# Supplemental References


Aghajan, M., Jonai, N., Flick, K., Fu, F., Luo, M., Cai, X., Ouni, I., Pierce, N., Tang, X., Lomenick, B.*, et al.* (2010). Chemical genetics screen for enhancers of rapamycin identifies a specific inhibitor of an SCF family E3 ubiquitin ligase. Nat Biotechnol *28*, 738-742.

Amoutzias, G.D., He, Y., Gordon, J., Mossialos, D., Oliver, S.G., and Van de Peer, Y. (2010). Posttranslational regulation impacts the fate of duplicated genes. Proc Natl Acad Sci U S A *107*, 2967-2971.

Beausoleil, S.A., Villen, J., Gerber, S.A., Rush, J., and Gygi, S.P. (2006). A probability-based approach for high-throughput protein phosphorylation analysis and site localization. Nat Biotechnol *24*, 1285-1292.

Ear, P.H., and Michnick, S.W. (2009). A general life-death selection strategy for dissecting protein functions. Nat Methods *6*, 813-816.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res *32*, 1792-1797.

Elias, J.E., and Gygi, S.P. (2007). Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nat Methods *4*, 207-214.

Kundrotas, P.J., Zhu, Z., and Vakser, I.A. (2010). GWIDD: Genome-wide protein docking database. Nucleic Acids Res *38*, D513-517.

Linding, R., Jensen, L.J., Diella, F., Bork, P., Gibson, T.J., and Russell, R.B. (2003). Protein disorder prediction: implications for structural proteomics. Structure *11*, 1453-1459.

Michnick, S.W., Ear, P.H., Landry, C., Malleshaiah, M.K., and Messier, V. (2010). A toolkit of protein-fragment complementation assays for studying and dissecting large-scale and dynamic protein-protein interactions in living cells. Methods Enzymol *470*, 335-368.

Mihel, J., Sikic, M., Tomic, S., Jeren, B., and Vlahovicek, K. (2008). PSAIA - protein structure and interaction analyzer. BMC Struct Biol *8*, 21.
Mok, J., Kim, P.M., Lam, H.Y., Piccirillo, S., Zhou, X., Jeschke, G.R., Sheridan, D.L., Parker, S.A., Desai, V., Jwa, M*., et al.* (2010). Deciphering protein kinase specificity through large-scale analysis of yeast phosphorylation site motifs. Sci Signal *3*, ra12.
Mosca, R., Pons, C., Fernandez-Recio, J., and Aloy, P. (2009). Pushing structural information into the yeast interactome by high-throughput protein docking experiments. PLoS Comput Biol *5*, e1000490.
Stark, C., Su, T.C., Breitkreutz, A., Lourenco, P., Dahabieh, M., Breitkreutz, B.J., Tyers, M., and Sadowski, I. (2010). PhosphoGRID: a database of experimentally verified in vivo protein phosphorylation sites from the budding yeast Saccharomyces cerevisiae. Database (Oxford) *2010*, bap026.
Stein, A., Ceol, A., and Aloy, P. (2011). 3did: identification and classification of domain-based interactions of known three-dimensional structure. Nucleic Acids Res *39*, D718-723.