LETTERS

Global analysis of protein phosphorylation in yeast

Jason Ptacek[1]*, Geeta Devgan[2]*, Gregory Michaud[3]*, Heng Zhu[2], Xiaowei Zhu[2], Joseph Fasolo[2], Hong Guo[3], Ghil Jona[2], Ashton Breitkreutz[4], Richelle Sopko[4], Rhonda R. McCartney[5], Martin C. Schmidt[5], Najma Rachidi[6], Soo-Jung Lee[7], Angie S. Mah[8], Lihao Meng[3], Michael J. R. Stark[6], David F. Stern[7], Claudio De Virgilio[9], Mike Tyers[4], Brenda Andrews[4,10], Mark Gerstein[1], Barry Schweitzer[3], Paul F. Predki[3] & Michael Snyder[1,2]

Protein phosphorylation is estimated to affect 30% of the proteome and is a major regulatory mechanism that controls many basic cellular processes[1–3]. Until recently, our biochemical understanding of protein phosphorylation on a global scale has been extremely limited; only one half of the yeast kinases have known *in vivo* substrates and the phosphorylating kinase is known for less than 160 phosphoproteins. Here we describe, with the use of proteome chip technology[4], the *in vitro* substrates recognized by most yeast protein kinases[5]: we identified over 4,000 phosphorylation events involving 1,325 different proteins. These substrates represent a broad spectrum of different biochemical functions and cellular roles. Distinct sets of substrates were recognized by each protein kinase, including closely related kinases of the protein kinase A family and four cyclin-dependent kinases that vary only in their cyclin subunits. Although many substrates reside in the same cellular compartment or belong to the same functional category as their phosphorylating kinase, many others do not, indicating possible new roles for several kinases. Furthermore, integration of the phosphorylation results with protein–protein interaction[6–10] and transcription factor binding data[11,12] revealed novel regulatory modules. Our phosphorylation results have been assembled into a first-generation phosphorylation map for yeast. Because many yeast proteins and pathways are conserved, these results will provide insights into the mechanisms and roles of protein phosphorylation in many eukaryotes.

To develop a kinase-substrate map for eukaryotes, we determined the substrates recognized by 87 different yeast protein kinases and bovine protein kinase A, by using a yeast proteome array and the scheme depicted in Fig. 1a. A total of 82 unique kinases representing most of the 122 yeast protein kinases[5] were tested; two cyclin-dependent kinases, Pho85 (in complex with Pcl1, Pcl2, Pcl9 and Pho80) and Cdc28 (in complex with Cln2 and Clb5), were also analysed.

Each kinase was incubated separately with two yeast proteome microarrays in the presence of [γ-^{33}P]ATP (Fig. 1b). The microarrays contained about 4,400 proteins spotted in duplicate on the array. The arrays also contained a variety of control proteins including three protein kinases that served both as positive controls and as landmarks for the identification of phosphorylation signals. For each experiment, two slides were also incubated in the absence of a protein kinase serving to identify protein kinases on the array that autophosphorylate. Four protein kinases (Rim15, Dbf2, Hsl1 and Rad53) that contained inactivating mutations in their catalytic domain were used as negative controls and exhibited signals identical to those obtained in the absence of protein kinase. The extent of phosphorylation was measured with algorithms specifically designed to detect positive signals. Proteins that were reproducibly phosphorylated in the presence of active kinase relative to the control slides were scored as positive substrates. All results are accessible at http://networks.gersteinlab.org/phosphorylome/.

Approximately 4,200 phosphorylation events affecting 1,325 proteins were identified from the 87 yeast protein kinase assays. Each kinase recognized between 1 and 256 substrates with an average of 47 substrates per kinase. A distinct set of substrates was phosphorylated by each protein kinase, indicating that every kinase has a unique substrate recognition profile. Most (73%) substrates were recognized by fewer than three kinases, indicating a strong preference of particular kinases for specific substrates (Supplementary Fig. 1). The largest class of proteins phosphorylated by the protein kinases was transcription factors (311 phosphorylations; $P < 10^{-99}$).

Inspection of the substrate list revealed that at least 14 known *in vivo* substrates of particular kinases were identified (Supplementary Table 1). In addition, each kinase phosphorylated proteins residing in the same cellular location and/or functional category as the kinase. For example, Ark1 (actin-regulating kinase) phosphorylated three substrates involved in late secretory functions, a known role for actin; two of these, Sla1 and Ent2, reside at the cell cortex, the same location as actin.

To determine whether other kinase-substrate pairs represent *in vivo* phosphorylation events, we tested whether the phosphorylation of several candidate substrates depended on the identified kinase *in vivo*. Substrates of six kinases were assayed for loss of phosphorylation by either a reduction or an absence of a signal, a mobility shift, or both in kinase deletion strains relative to wild type (Fig. 1c, i–vi). Differences were observed in 12 cases. Interestingly, in at least five cases we observed that strains lacking the kinase gene had significantly altered levels of the putative substrate (Fig. 1c, iii–vi); in three cases kinase loss increased substrate levels, and in two the substrate levels decreased, indicating that the kinases control protein levels directly or indirectly. Overall, 9.2% of substrates exhibited a reduced phosphorylation, a mobility shift or a markedly altered level of substrate. This is presumably a significant underestimate of the number of *in vivo* substrates for these kinases because many proteins do not exhibit mobility shifts upon phosphorylation, may be modified by redundant kinases or have multiple phosphorylations that mask the loss of one or more phosphorylations. Nonetheless, our *in vivo*

[1]Department of Molecular Biophysics and Biochemistry, and [2]Molecular, Cellular and Developmental Biology, Yale University, New Haven, Connecticut 06511, USA. [3]Invitrogen Corporation, 1600 Faraday Avenue, Carlsbad, California 92008, USA. [4]Department of Medical Genetics and Microbiology, University of Toronto, 1 King's College Circle, Toronto, Ontario M55 1A8, Canada. [5]Department of Molecular Genetics and Biochemistry, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15261, USA. [6]Division of Gene Regulation and Expression, School of Life Sciences, University of Dundee, Dundee DD1 5EH, UK. [7]Department of Pathology, Yale University School of Medicine, PO Box 208023, 310 Cedar Street BML 342, New Haven, Connecticut 06520-8023, USA. [8]Department of Biology, California Institute of Technology, Pasadena, California 91125, USA. [9]Department of Microbiology and Molecular Medicine, CMU, University of Geneva, 1211 Geneva, Switzerland. [10]Banting & Best Department of Medical Research, University of Toronto, Room 4285, Medical Sciences Building, 1 King's College Circle, Toronto, Ontario M55 1A8, Canada.
*These authors contributed equally to this work.

validation results indicate that the proteome microarray approach has identified many genuine substrates of protein kinases and that phosphorylation can markedly affect protein levels.

We also examined substrate profiles in closely related kinases, a common feature of eukaryotes. Yeast protein kinase A homologues Tpk1 and Tpk3 are 84% identical in amino-acid sequence and 67% and 76% identical to Tpk2, respectively. Strains lacking all three are nonviable, whereas those containing any one of the three Tpks propagate, indicating that each is genetically redundant for cell growth[13]. To determine whether the Tpk kinases are functionally redundant biochemically, each Tpk was directly tested for reactivity with substrates by using proteome arrays prepared and probed at the same time. As shown in Fig. 2, Tpk1, Tpk2 and Tpk3 recognized 256, 29 and 79 substrates, respectively; however, only 8 were recognized by all three kinases and 39 were recognized by two of the three. The vast majority (87.7%) were recognized by only one of the Tpks, indicating that each kinase has a unique substrate specificity; 86.6% of the bovine cAMP-dependent protein kinase (PKA) targets were also substrates of Tpk1. In comparison, two slides probed by the same Tpk showed a greater than 90% substrate overlap. Thus, the closely related Tpks have distinct substrate specificities. These results are consistent with the observation that Tpk1 and Tpk3, although redundant with Tpk2 for cell growth, have different roles in pseudohyphal growth[14].

In addition to analysing the Tpks, we examined whether a protein kinase complexed with different regulators recognizes similar or distinct sets of substrates. The substrates recognized by the cyclin-dependent kinase Pho85, purified by itself or complexed with either Pcl1, Pcl2, Pcl9 or Pho80 cyclins, were determined in parallel. Pho85 purified alone recognized only 12 targets, indicating that it has weak activity, as expected. However, between 60 and 255 substrates were observed in the presence of different cyclins. Pho85 complexed with different cyclins exhibited various degrees of overlap in the substrates phosphorylated *in vitro* (Fig. 2). Nearly half (29/60) of the Pho85–Pcl2 substrates overlap with the 89 substrates of Pho85–Pcl9, and most (48/89) of the Pho85–Pcl9 substrates overlap with the 255 substrates of Pho85–Pho80. Pho85–Pcl1 also shares a high degree (43.2%) of substrate preference with Pho85–Pho80, but exhibits very little overlap with Pho85–Pcl9.

These different results indicate that the amino-acid differences of the Tpks and Pho85 cyclins have a considerable influence in substrate recognition. These studies further provide a molecular explanation of why eukaryotic cells have multiple protein kinases with a high degree of sequence similarity: each has different biochemical propensities for particular substrates.

The substrates phosphorylated by the different kinases were also searched for common sequence motifs[15]. Consensus motifs were identified for 11 kinases; these are similar to the sequence motifs determined for kinase orthologues in other species (Table 1).

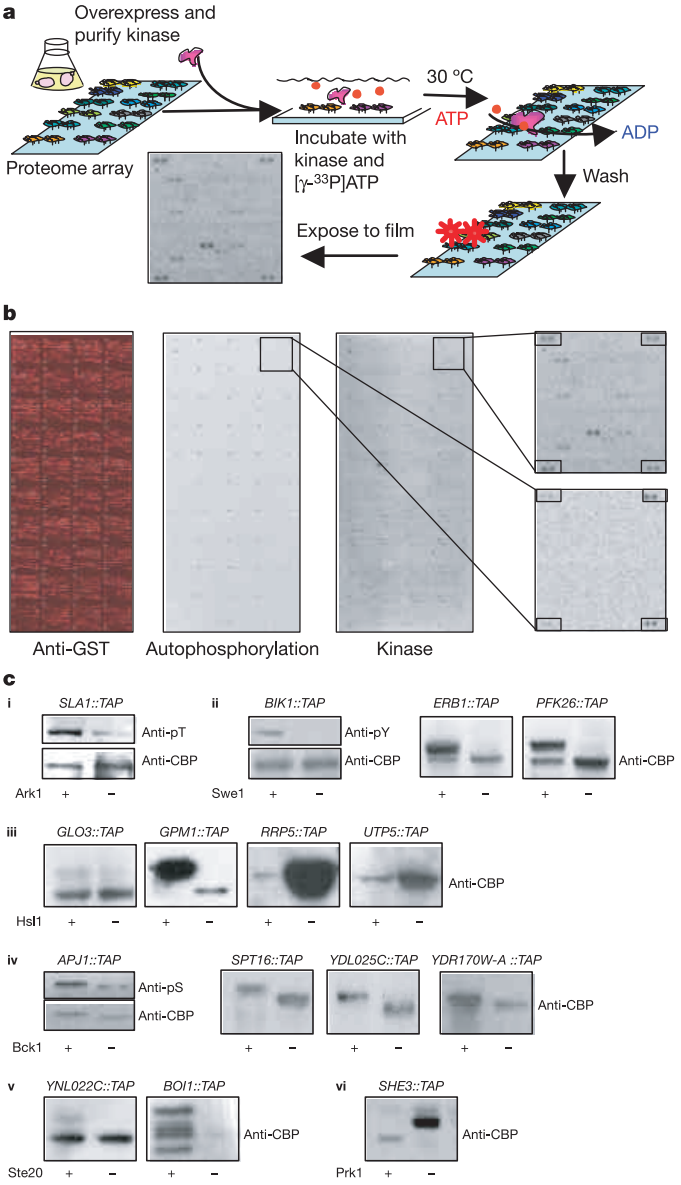

Figure 1 | Identification of kinase substrates using protein chips followed by *in vivo* validation. **a**, Overall scheme to identify kinase substrates. Each kinase was overexpressed, purified and assayed on protein chips containing about 4,400 proteins spotted in duplicate. **b**, Kinase assays on protein chips. Two protein chips were used for every kinase assayed. In addition, two protein chips were probed in the absence of kinase to identify proteins on the chip that autophosphorylate. Commercial kinases were spotted at many defined locations, shown in the four corners of the two boxes on the right; these kinases autophosphorylated in our assay and served as landmarks for the identification of phosphorylation signals. The slide on the left is a representative slide probed with anti-GST antibodies indicating the amount of fusion protein present on the proteome slide. **c**, *In vivo* validation of targets identified on the proteome microarray: *ARK1* (i), *SWE1* (ii), *HSL1* (iii), *BCK1* (iv), *STE20* (v) and *PRK1* (vi) were deleted from the TAP-tagged strains indicated. From the kinase-deleted strains, the tagged proteins were purified and their phosphorylation status compared with wild-type tagged proteins. Immunoblots were probed with anti-phosphothreonine antibody (i, top panel), anti-phosphotyrosine antibody (ii, top panel) or anti-phosphoserine antibody (iv, top panel). In addition, protein isoforms and protein levels were monitored with anti-CBP antibody (i–vi).

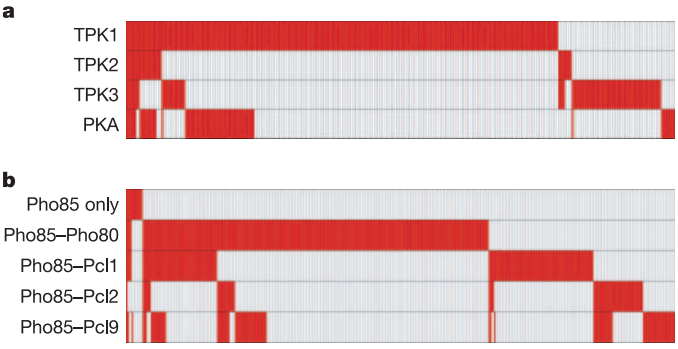

Figure 2 | Comparison of substrates targeted by related kinases. **a**, Comparison of the substrate of the different Tpks. Each Tpk kinase has a unique substrate recognition profile; 86.6% of PKA targets are also Tpk1 substrates. **b**, Comparison of the substrates recognized by the different Pho85–cyclin complexes. Pho85, when assayed in the presence of different cofactors, displays different specificities, indicating that the cyclin subunits have a significant impact on substrate recognition.

Table 1 | Summary of motif results

Kinase	Pratt pattern	*P*	Sites (proteins)	Hits/total substrates	Total protein with sites	Published pattern/site
ARK1	L-x(4)-T-x-[GL]-x-[ST]	4.29×10^{-7}	7 (6)	6/8	209	[LI]-x(2)-Q-x-T-G
CDC28	T-P	2.33×10^{-6}	98 (38)	39/43	2,336	[ST]-P
CKA1	T-x(2)-D	1.04×10^{-3}	42 (18)	18/19	2,487	[ST]-x(2)-[DE]
CMK2	R-x(2)-[ST]-x-[ST]	5.42×10^{-3}	7 (7)	7/9	1,274	R-x(2)-[ST]
DUN1	[ST]-x(3)-S-S	8.63×10^{-5}	33 (15)	15/18	1,530	GSSAS*AS*AS*SLEM (SML1 site)
HRR25	S-x(2)-S	4.75×10^{-2}	58 (14)	14/14	3,288	S-x(2)-[ST]
PKA	R-[KR]-x-S	5.57×10^{-40}	96 (55)	55/56	759	R-[KR]-x-S
PRK1	L-x(4)-[ST]	2.24×10^{-1}	259 (40)	40/41	3,810	[LI]-x(2)-Q-x-T-G
RIM11	S-x(3)-[ST]-x(5)-S-x-[ST]	5.25×10^{-10}	25 (13)	13/17	454	S-x(3)-S
SKY1	D-x(5)-S	8.28×10^{-6}	153 (39)	39/40	2,840	YRTRDAPRERS*PTR (NPL3 site)
YCK2	S-x(2)-D	1.26×10^{-11}	715 (197)	197/220	2,907	S-x(2)-[DE]

Consensus sites were found for 11 protein kinases. These are similar to those reported in the literature. Also summarized are the number of consensus sites in the target proteins (sites (proteins)), the fraction of identified substrates with a consensus site (hits/total substrates) and the total number of proteins on the array with consensus sites (total protein with sites).

Although many of the substrates contain the consensus phosphorylation sites, many yeast proteins with consensus sequences are not kinase targets. For example, 209 and 3,288 proteins on the array have Ark1 and Hrr25 consensus sites, yet only 8 and 14 proteins were recognized, respectively. Therefore, either additional sequences on the substrate help direct substrate recognition or the consensus phosphorylation site is not accessible to the kinase due to spatial and temporal restrictions. Nevertheless, these studies show the importance of directly assaying for protein phosphorylation with experimental tests.

The 4,200 different protein-kinase–substrate phosphorylations have been assembled into an *in vitro* phosphorylation network (Fig. 3). In many cases the identification of substrates helps to define the role of the kinase in yeast signalling networks more accurately. For example, the phosphorylation of Sla1 and She3 by Ark1 and Prk1, respectively, further explains the role of these kinases in actin regulation. Forty-nine kinase–substrate interactions are also present in the protein–protein interaction network. Presumably the low overlap between these data sets is because kinase–substrate interactions are expected to be transient with low binding affinities and not detectable by most protein–protein interaction assays.

Many of these interactions may represent *in vitro* targets that do not occur *in vivo*. Filtering the phosphorylation network to contain only the interactions in which kinase and substrate are present in the same cellular compartment or in the same functional categories results in 1,384 (33%; $P < 10^{-99}$) and 768 (18.4%; $P < 10^{-99}$) interactions, respectively (Fig. 3, Supplementary Table 2). Furthermore, of the 29 kinases with functional assignments and 9 or more substrates, 14 showed enrichment of substrates in the same categories as those of the protein kinase (Supplementary Table 3). Filtering is likely to enrich the data set for interactions that occur *in vivo*. However, this approach may also increase the number of false

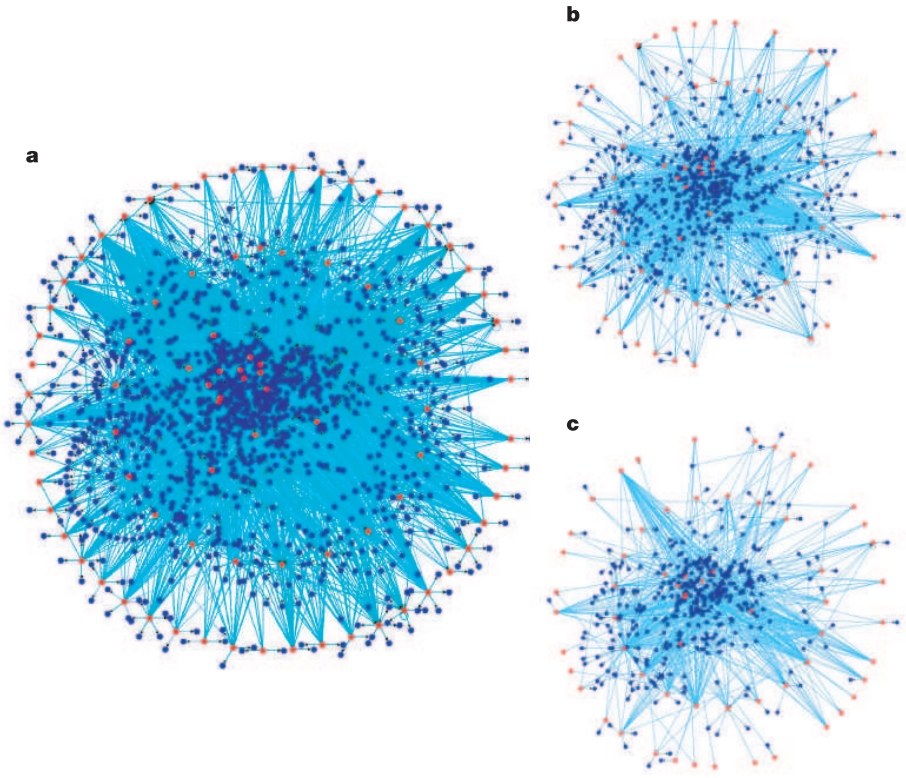

Figure 3 | An *in vitro* phosphorylation map of yeast. a, A map showing the connections between kinases and substrates. In all, 87 different kinases (red dots) and 1,325 substrates (blue dots) are represented in the map. **b**, Global localization data can be used to identify only those phosphorylation events occurring between proteins of the same cellular compartment. **c**, Functional data can be used to identify substrates with similar functions to those of the kinases phosphorylating them.

negatives because the protein localization data may be incomplete and/or inaccurate. Furthermore, the functions of many protein kinases and substrates are either limited or unknown, and thus many substrates would not be expected to have functions assigned to the particular kinase. For example, two kinases involved in a G2/M cell-cycle control pathway, Hsl1 (ref. 16) and Swe1 (ref. 17), phosphorylate proteins involved in glucose metabolism, Gpm1 and Pfk26, respectively; these interactions were validated *in vivo*. Hsl1 and Swe1 localize to the septin ring and nucleus/spindle pole/bud neck, respectively, whereas Gpm1 and Pfk26 are cytosolic proteins involved in glucose metabolism. Presumably, these proteins interact in the cytoplasm and link cytoskeletal function and checkpoint control with glycolysis. The assignment of new functions to protein kinases probably reveals how different cellular pathways can be coordinated by a single regulator.

To understand how phosphorylation may be integrated into global regulatory networks, we also combined the phosphorylation data with transcription factor binding and protein interaction data and generated the first integrated regulatory network for yeast. We then searched this network for common regulatory modules[6–12,18]. Eight modules were observed, and six (modules 1–6) were of high statistical significance (Fig. 4). At least four modules (1, 2, 3 and 7) have been validated in our studies or from the literature. All of the modules involve kinase–substrate pairs, which we refer to as 'kinates' (kinase–substrate pairs). The modules are: (1) interacting kinates, (2) scaffolds, (3) kinase cascades, (4) transcription-factor-regulated kinates, (5) kinate regulon, (6, 7) feedback loops and (8) heterosubstrate regulation. Examples are shown in Fig. 4 and a comprehensive list is available at http://networks.gersteinlab.org/phosphorylome/. These results show the utility of integrating different data types; many potential novel regulatory networks, not evident from single data sets, have been identified.

This study is, to our knowledge, the first global investigation of protein phosphorylation by protein kinases using an unbiased approach. A total of 181 substrates of Cdc28–Clb2 were identified previously[19] by computationally searching for substrates with multiple cyclin-dependent kinase sites and assaying for phosphorylation of the glutathione *S*-transferase (GST)-tagged proteins in cell extracts with a Cdc28-*as* allele. Of 150 tested 24% were preferentially phosphorylated by a Clb28–Clb5 kinase[20]. We found overlap (10 and 7; $P < 10^{-6}$) between their Cdc28–Clb2 and Cdc28–Clb5 target lists and the 43 substrates we identified with Cdc28–Clb5. Differences between the lists might be due to one of the following: first, the sensitivity of the respective assays; second, an '*as* allele' as opposed to a wild-type kinase; third, the use of different cyclins in the initial screen; fourth, the fact that their assays were performed in cellular extracts and ours were not; and/or fifth, their use of a biased approach in comparison with our survey of most yeast proteins.

We also compared our substrate list with two other studies that mapped *in vivo* phosphorylation sites. The first[2] mapped phosphorylation sites on 98 proteins that were on our protein chips; our study provides candidate kinases for 50 of these proteins (51%). The second[21] identified 89 phosphorylation sites that are induced when yeast cells are treated with mating pheromone. Five proteins with these sites are substrates of Pho85–Pcl1, two are substrates for Kss1 and a single substrate was identified for Ste20 and Fus3. The remaining proteins whose phosphorylation is induced by treatment with pheromone may be targets either of additional kinases induced by the pheromone response (for example, Ste7 or Ste11) or of other kinases. Nonetheless, combining the data of our study with *in vivo* data from others provides strong candidates for the kinases that phosphorylate each substrate.

Proteome chips offer many advantages for studying protein phosphorylation. Thousands of proteins can be rapidly screened for enzyme–substrate relationships in an unbiased fashion with very small amounts of reagents and under a variety of test conditions. In addition, closely related kinases with known redundant functions can be readily differentiated at the molecular level on the basis of their substrate profiles.

Although we were able to identify many known substrates of protein kinases, two-thirds of reported phosphorylations were not observed. This may be due to the absence of 30% of yeast proteins on the array. In addition, the substrates may not be present in sufficient quantity for substrate phophorylation to be observed. Alternatively, although purified from yeast, the kinases and/or substrates may lack functional adaptors, scaffold proteins or modifications. In principle,

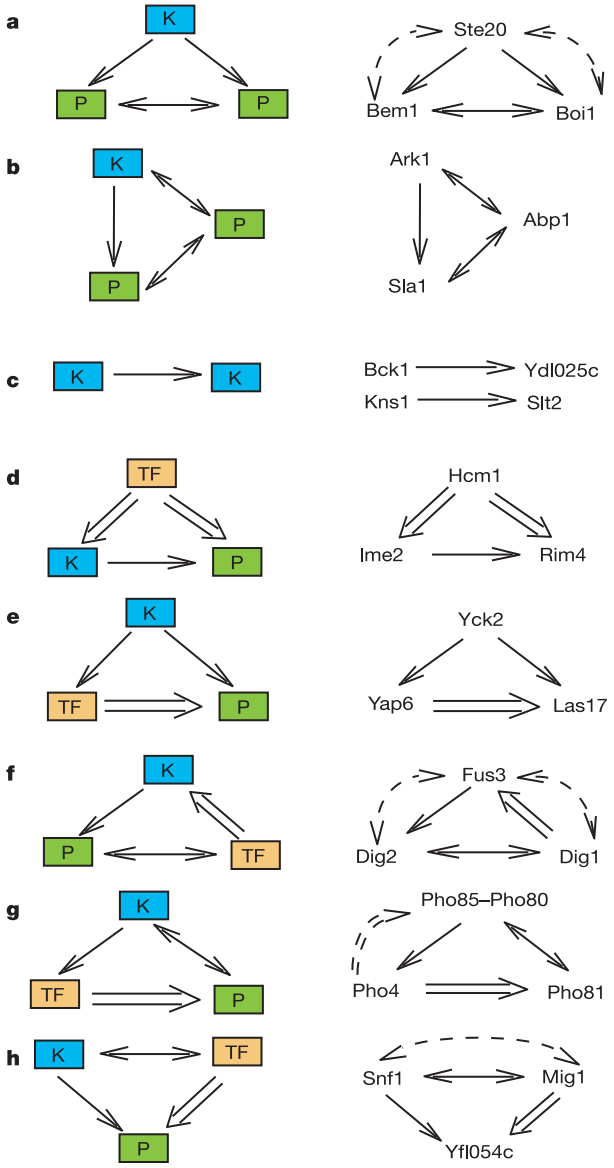

Figure 4 | Integration of other biological data reveals regulatory modules. When protein interaction, transcription factor binding and expression data are considered, many modules within the phosphorylation data are identified. Shown are protein–protein interactions (↔), kinase phosphorylations (⟶) and transcription factor (TF) regulation (⇒). K, kinase; P, protein. In the following list the modules are numbered from 1 to 8 with their common name in parentheses; also listed are the numbers of occurrences (n) and the statistical significances of such events. **a**, Module 1 (interacting kinates module), $n = 1{,}563$, $P < 10^{-99}$; **b**, module 2 (scaffold module), $n = 2{,}448$, $P < 10^{-99}$; **c**, module 3 (kinase cascade module), $n = 147$, $P < 10^{-99}$; **d**, module 4 (TF-regulated kinate module), $n = 145$, $P < 10^{-99}$; **e**, module 5 (kinate regulon module), $n = 92$, $P < 10^{-99}$; **f**, module 6 (kinate feedback loop I module), $n = 25$, $P = 10^{-3}$; **g**, module 7 (kinate feedback loop II module), $n = 11$, not enriched; **h**, module 8 (heterosubstrate regulation module), $n = 14$, not enriched.

the kinase or substrates can be purified from cells grown under different conditions to provide proper modifications on the kinase or substrates. In addition, adaptor proteins (such as cyclins), when identified, can be added to the reactions.

Our assays measure substrate specificity directly at the level of kinase–substrate interaction, which is highly selective because a discrete set of substrates are recognized by each kinase. Nonetheless, phosphorylations that do not normally occur *in vivo* may be identified from this assay. These false positives may be due to either *in vitro* phosphorylation of proteins by kinases that normally reside in other cellular compartments and/or are expressed at different times, or through the absence of adaptor proteins that limit the kinase–substrate interactions. It is unlikely that many false positives are observed by co-purification of a substrate with the intended protein on the array; 80% of all substrates identified can be validated by solution-based assays for substrate phosphorylation and the remainder do not reveal phosphorylation of a co-purifiying protein (Supplementary Fig. 2). Combining our data with other information provides a useful method of detecting interactions likely to occur *in vivo*. Because many kinase signalling pathways are highly conserved from fungi to humans[1], our comprehensive identification of the phosphorylation regulatory network in yeast will not only serve as valuable resource for yeast research but will also provide much insight into this important regulatory network in all eukaryotes.

METHODS

Protein purification and proteome arrays. Yeast proteome chips were prepared that were similar to that described previously with the use of yeast strains that overexpressed yeast proteins as GST fusions[4]. About 4,400 yeast strains that consistently express protein of the correct size were used to prepare protein chips (Supplementary Information). Proteins were spotted in duplicate on surface-modified microscope slides by using a 48-pin contact printer (Genomic Solutions). The protein arrays were manufactured at Invitrogen (Branford, Connecticut). Protein kinases containing inactivating mutations in their catalytic domains were prepared by site-directed mutagenesis to mutate an absolutely conserved catalytic residue, Asp, to Ala (ref. 22).

The Pho85 kinases were purified from insect cells[23]; the remaining 81 kinases were purified from yeast. Yeast protein kinases were expressed as GST fusions and purified as described previously[5]. Cells were grown in 50–500-ml cultures, harvested and lysed with glass beads in lysis buffer (100 mM Tris-HCl pH 7.4, 100 mM NaCl, 1 mM EGTA, 0.1% 2-mercaptoethanol, 0.1% Triton X-100, protease cocktail (Roche), 1 mM EDTA, 50 mM NaF, 10 mM sodium β-glycerophosphate, 1 mM Na_3VO_4). Kinases were bound to glutathione beads and eluted into kinase buffer (100 mM Tris-HCl pH 8.0, 100 mM NaCl, 10 mM $MgCl_2$, 20 mM glutathione, 20% glycerol). Although attempts were made to purify more than 110 protein kinases, only those that were highly active in *in vitro* assays were tested on the protein chips.

Kinase assay and data acquisition. Each kinase was assayed to determine the concentration needed to achieve optimal signal:noise ratio using test protein chips (see Supplementary Information). Proteome arrays were blocked in Superblock (Pierce) with 0.1% Triton X-100 for 1 h at 4 °C and probed in duplicate for every kinase. Optimized conditions (see Supplementary Information) consisted of diluting the kinase into kinase buffer plus 0.5 mg ml^{-1} bovine serum albumin, 0.1% Triton X-100 and 2 μl of [γ-^{33}P]ATP (33.3 nM final concentration). Each kinase in buffer was overlaid on two arrays, covered with a coverslip and placed in a humidified chamber at 30 °C for 1 h. The slides were washed twice with 10 mM Tris-HCl pH 7.4, 0.5% SDS and once with doubly distilled water before being spun dry and exposed to X-ray film (Kodak). For each experiment, two additional arrays were incubated with kinase buffer in the absence of kinase, which served as autophosphorylation reference slides. Three sets of exposures were taken for each kinase assayed: 1, 3 and 7 days. The X-ray film was scanned at 1,800 dots per inch and each kinase was analysed with Genepix 3.0. The optimal exposure was selected for each kinase and compared with the corresponding autophosphorylation slides.

Data analysis. A computer algorithm was written to identify substrates (see Supplementary Information). Substrates were identified that were two standard deviations above background for at least three of the four protein spots from the two slides assayed for a given kinase. Analysis from the autophosphorylation slides was used to remove proteins showing autophosphorylation. Positive signals were also inspected visually to ensure that each spot was not caused by the presence of an artefact.

A network was generated by combining the substrates for all kinases assayed by using Osprey version 1.2.0 (ref. 24). For each kinase, the substrates were filtered on the basis of functional data from the Munich Information Center for Protein Sequences (MIPS)[18] and localization data in ref. 25. Functional enrichment was performed for the substrates of all the kinases. Substrates enriched with $P < 0.05$ were considered enriched and were then compared with the functional annotation for the kinase. Alignment of the substrates of the Tpks and Pho85 isoforms was performed to obtain a substrate profile for each of the kinases. Modules were identified between transcription factors, kinases and substrates. The requirement for the three-element module was that it should contain at least one phosphorylation interaction and have a total of at least three interactions with all proteins having two interactions. For module 3, kinase–kinase interactions were searched. Common phosphorylation motifs from sets of substrates were identified with a Pratt algorithm[15].

***In vivo* substrate validation.** The phosphorylating kinase was deleted for potential substrates of Ark1, Swe1, Hsl1, Bck1, Prk1 and Ste20 by using the available chromosomally tandem affinity purification (TAP-tagged) strains (see Supplementary Information)[26,27]. Verified strains were grown and analysed in parallel with the corresponding wild-type TAP-tag strain. TAP-tagged proteins were purified with lysis buffer as above and isolated from the lysates with the use of IgG beads. The bound IgG beads were washed with lysis buffer containing 250 mM NaCl. The beads were heated to 70 °C in the presence of NuPAGE loading buffer and eluates were analysed on 10% NuPAGE gels (Invitrogen). Immunoblots were prepared and probed with anti-phosphoserine, anti-phosphothreonine (Qiagen) and anti-phosphoTyr (Upstate) antibodies to detect loss of phosphorylation. Immunoblots were also probed with anti-calmodulin-binding-peptide antibody (Upstate), which recognizes a portion of the TAP tag, to identify gel mobility shifts. All validated substrates were tested at least twice, and for Glo3, She3 and Gpm1 independent transformants were tested.

Received 21 June; accepted 1 September 2005.

1. Manning, G., Plowman, G. D., Hunter, T. & Sudarsanam, S. Evolution of protein kinase signalling from yeast to man. *Trends Biochem. Sci.* **27,** 514–520 (2002).
2. Ficarro, S. B. *et al.* Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. *Nature Biotechnol.* **20,** 301–305 (2002).
3. Cohen, P. The regulation of protein function by multisite phosphorylation—a 25 year update. *Trends Biochem. Sci.* **25,** 596–601 (2000).
4. Zhu, H. *et al.* Global analysis of protein activities using proteome chips. *Science* **293,** 2101–2105 (2001).
5. Zhu, H. *et al.* Analysis of yeast protein kinases using protein chips. *Nature Genet.* **26,** 283–289 (2000).
6. Gavin, A. C. *et al.* Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415,** 141–147 (2002).
7. Ito, T. *et al.* A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl Acad. Sci. USA* **98,** 4569–4574 (2001).
8. Uetz, P. *et al.* A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. *Nature* **403,** 623–627 (2000).
9. Xenarios, I. *et al.* DIP: the database of interacting proteins. *Nucleic Acids Res.* **28,** 289–291 (2000).
10. Bader, G. D. & Hogue, C. W. BIND—a data specification for storing and describing biomolecular interactions, molecular complexes and pathways. *Bioinformatics* **16,** 465–477 (2000).
11. Horak, C. E. *et al.* Complex transcriptional circuitry at the G1/S transition in *Saccharomyces cerevisiae*. *Genes Dev.* **16,** 3017–3033 (2002).
12. Lee, T. I. *et al.* Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* **298,** 799–804 (2002).
13. Toda, T., Cameron, S., Sass, P., Zoller, M. & Wigler, M. Three different genes in *S. cerevisiae* encode the catalytic subunits of the cAMP-dependent protein kinase. *Cell* **50,** 277–287 (1987).
14. Pan, X. & Heitman, J. Cyclic AMP-dependent protein kinase regulates pseudohyphal differentiation in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **19,** 4874–4887 (1999).
15. Jonassen, I., Collins, J. F. & Higgins, D. G. Finding flexible patterns in unaligned protein sequences. *Protein Sci.* **4,** 1587–1595 (1995).
16. Hanrahan, J. & Snyder, M. Cytoskeletal activation of a checkpoint kinase. *Mol. Cell* **12,** 663–673 (2003).
17. Barral, Y., Parra, M., Bidlingmaier, S. & Snyder, M. Nim1-related kinases coordinate cell cycle progression with the organization of the peripheral cytoskeleton in yeast. *Genes Dev.* **13,** 176–187 (1999).
18. Mewes, H. W., Albermann, K., Heumann, K., Liebl, S. & Pfeiffer, F. MIPS: a database for protein sequences, homology data and yeast genome information. *Nucleic Acids Res.* **25,** 28–30 (1997).
19. Ubersax, J. A. *et al.* Targets of the cyclin-dependent kinase Cdk1. *Nature* **425,** 859–864 (2003).
20. Loog, M. & Morgan, D. O. Cyclin specificity in the phosphorylation of cyclin-dependent kinase substrates. *Nature* **434,** 104–108 (2005).

21. Gruhler, A. *et al.* Quantitative phosphoproteomics applied to the yeast pheromone signalling pathway. *Mol. Cell Proteomics* **4,** 310–327 (2005).
22. Huse, M. & Kuriyan, J. The conformational plasticity of protein kinases. *Cell* **109,** 275–282 (2002).
23. Moffat, J. & Andrews, B. Late-G1 cyclin-CDK activity is essential for control of cell morphogenesis in budding yeast. *Nature Cell Biol.* **6,** 59–66 (2004).
24. Breitkreutz, B. J., Stark, C. & Tyers, M. Osprey: a network visualization system. *Genome Biol.* **4,** R22.1–4 (2003).
25. Huh, W. K. *et al.* Global analysis of protein localization in budding yeast. *Nature* **425,** 686–691 (2003).
26. Ghaemmaghami, S. *et al.* Global analysis of protein expression in yeast. *Nature* **425,** 737–741 (2003).
27. Winzeler, E. A. *et al.* Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285,** 901–906 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Tang for help with the initial Cdc28 preparations, and D. Gelperin, J. Mok and K. Wise for comments on the manuscript. M.S. was funded by grants from the NIH; J.P., G.D. and J.F. were funded by NIH predoctoral fellowships, and B.A. and M.T. were funded by grants from the Canadian Institutes of Health Research. M.J.R.S. was funded by a project grant from the Wellcome Trust, UK.

Author Contributions Assay development was performed by H.Z., J.P., G.D., G.M. and M.S. Proteome chips were prepared by G.M., B.S. and P.F.P. at Invitrogen. G.J. contributed transcription factors for the arrays. Kinase assays were performed by J.P., G.D., H.Z. and M.S. Most kinases were prepared by J.P. and G.D. Additional kinases were provided by A.B., R.S., R.R.M., M.C.S., N.R., S.J.L., A.S.M., M.J.R.S., D.F.S, C.D.V., M.T. and B.A. Data analysis was performed by X.Z., J.P. and G.D. Consensus mapping was by H.G. *In vitro* solution validations were performed by G.M. and L.M. *In vivo* substrate validations were performed by J.P., G.D. and J.F. Most assays and analyses were performed in the laboratory of M.S. with contributions from M.G.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare competing financial interests: details accompany the paper at www.nature.com. Correspondence and requests for materials should be addressed to M.S. (michael.snyder@yale.edu).