Science Signaling
AAAS

Comparative Analysis Reveals Conserved Protein Phosphorylation Networks Implicated in Multiple Diseases

Chris Soon Heng Tan, Bernd Bodenmiller, Adrian Pasculescu, Marko Jovanovic, Michael O. Hengartner, Claus Jørgensen, Gary D. Bader, Ruedi Aebersold, Tony Pawson and Rune Linding (28 July 2009)
Science Signaling **2** (81), ra39. [DOI: 10.1126/scisignal.2000316]

Science Signaling (ISSN 1937-9145) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005.

COMPUTATIONAL BIOLOGY

Comparative Analysis Reveals Conserved Protein Phosphorylation Networks Implicated in Multiple Diseases

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

(Published 28 July 2009; Volume 2 Issue 81 ra39)

Protein kinases enable cellular information processing. Although numerous human phosphorylation sites and their dynamics have been characterized, the evolutionary history and physiological importance of many signaling events remain unknown. Using target phosphoproteomes determined with a similar experimental and computational pipeline, we investigated the conservation of human phosphorylation events in distantly related model organisms (fly, worm, and yeast). With a sequence-alignment approach, we identified 479 phosphorylation events in 344 human proteins that appear to be positionally conserved over ~600 million years of evolution and hence are likely to be involved in fundamental cellular processes. This sequence-alignment analysis suggested that many phosphorylation sites evolve rapidly and therefore do not display strong evolutionary conservation in terms of sequence position in distantly related organisms. Thus, we devised a network-alignment approach to reconstruct conserved kinase-substrate networks, which identified 778 phosphorylation events in 698 human proteins. Both methods identified proteins tightly regulated by phosphorylation as well as signal integration hubs, and both types of phosphoproteins were enriched in proteins encoded by disease-associated genes. We analyzed the cellular functions and structural relationships for these conserved signaling events, noting the incomplete nature of current phosphoproteomes. Assessing phosphorylation conservation at both site and network levels proved useful for exploring both fast-evolving and ancient signaling events. We reveal that multiple complex diseases seem to converge within the conserved networks, suggesting that disease development might rely on common molecular networks.

INTRODUCTION

Protein kinases recognize and phosphorylate linear motifs (*1*, *2*) in proteins. These molecular events can directly control the activities of other proteins and the dynamic assembly of directional protein-protein interaction networks. In combination with phosphatases, kinases regulate the phosphorylation-dependent binding of linear motifs to modular protein domains, such as the Src homology 2 (SH2) domain that recognizes phosphorylated tyrosine motifs and the BRCA1 C-terminal (BRCT) domain that recognizes phosphorylated serine and threonine motifs, and thereby create logic gates (*3*, *4*) that enable the cell to swiftly and precisely respond to both internal and external perturbations (*5*, *6*). Although interaction maps (*7–10*) provide useful information, it is the network dynamics and utilization that mediate cellular processing of environmental cues (*11*, *12*). Quantitative mass spectrometry (MS) measurements of phosphorylation networks and their dynamics are now rapidly unraveling thousands of cellular phosphorylation sites (*13–18*). With the functional and phenotypic characterization of previously unknown sites lagging behind their detection, a systematic way to highlight and prioritize important phosphorylation events is needed to guide functional experimental studies.

In addition, the conservation and evolutionary trace of most sites remain largely unknown. Unlike protein domains, which are conserved over long evolutionary distances, phosphorylation motifs are short and often reside in disordered fast-evolving regions (*19–23*). These properties render phosphorylation sites difficult to align and trace evolutionarily (*24–27*). Here, we assembled human phosphorylation sites previously identified in both large-scale MS [high throughput (HTP)] and low-throughput (LTP) targeted experiments (*28*, *29*) and explored their conservation with the phosphorylated proteins (phosphoproteomes) of three target model organisms (fly, worm, and yeast) that were measured with a similar experimental and computational pipeline. Through a combination of sequence-alignment and reconstructive, network-alignment approaches, we investigated the conservation of protein phosphorylation events at two distinct levels: sites that are conserved at similar positions (termed "positionally conserved") in orthologous proteins between human and at least one target species (termed "core sites") and those that are involved in conserved kinase-substrate regulatory networks but that are not necessarily constrained to the same location within phosphoproteins from humans and the model organisms.

RESULTS

To identify human sites that are conserved in distantly related model organisms and thereby likely to be important for fundamental cellular activities, we first identified positionally conserved sites with a full-length (global)

[1]Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada M5G 1X5. [2]Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada M5S 1A8. [3]Institute of Molecular Systems Biology, Eidgenössische Technische Hochschule (ETH), 8093 Zurich, Switzerland. [4]Institute of Molecular Biology, University of Zurich, 8057 Zurich, Switzerland. [5]Institute for Systems Biology, Seattle, WA 98103, USA. [6]Competence Center for Systems Physiology and Metabolic Diseases, ETH Zurich, 8093 Zurich, Switzerland. [7]Faculty of Science, University of Zurich, 8057 Zurich, Switzerland. [8]Cellular & Molecular Logic Team, Section of Cell and Molecular Biology, The Institute of Cancer Research (ICR), London SW3 6JB, UK.
*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: linding@icr.ac.uk

sequence-alignment algorithm (*30*) to map the experimentally identified phosphorylation sites from the target species (*Drosophila melanogaster*, *Caenorhabditis elegans*, and *Saccharomyces cerevisiae*) to orthologous human phosphoproteins (Fig. 1). This approach led to a conservative assessment of conserved sites because it requires the position of a site to be fixed within a multiple-sequence alignment. However, kinases can regulate cellular activities in ways that do not require their sites to occur at precise positions in protein sequences (*21*–*23*, *31*), as is the case in the threshold-dependent regulation of the Sic1 protein (*32*), for which phosphorylation at each of several sites promotes binding to Cdc4. Similarly, the ultrasensitive inactivation of Wee1 kinase is mediated by cyclin-dependent kinase 1 (Cdk1) decoy sites in both Wee1 and other proteins that "distract" CDK1 away from the causal sites in Wee1 (*33*). Therefore, we aimed to identify conserved human phosphorylation events that are not necessarily conserved at the same sites between orthologous kinases and substrates in the target species by deploying the NetworKIN (*34*) algorithm in combination with NetPhorest (*2*) to infer the relevant protein kinases for substrates identified in the phosphoproteomes of human and each target species. The computationally reconstructed human kinase-substrate network was subsequently overlaid with that of the target species to identify conserved kinase-substrate relationships. By taking two distinct approaches to assess phosphorylation conservation, we provide insight into the evolution of phosphorylation-based regulation with potential impact for our understanding of normal biological processes and complex diseases.

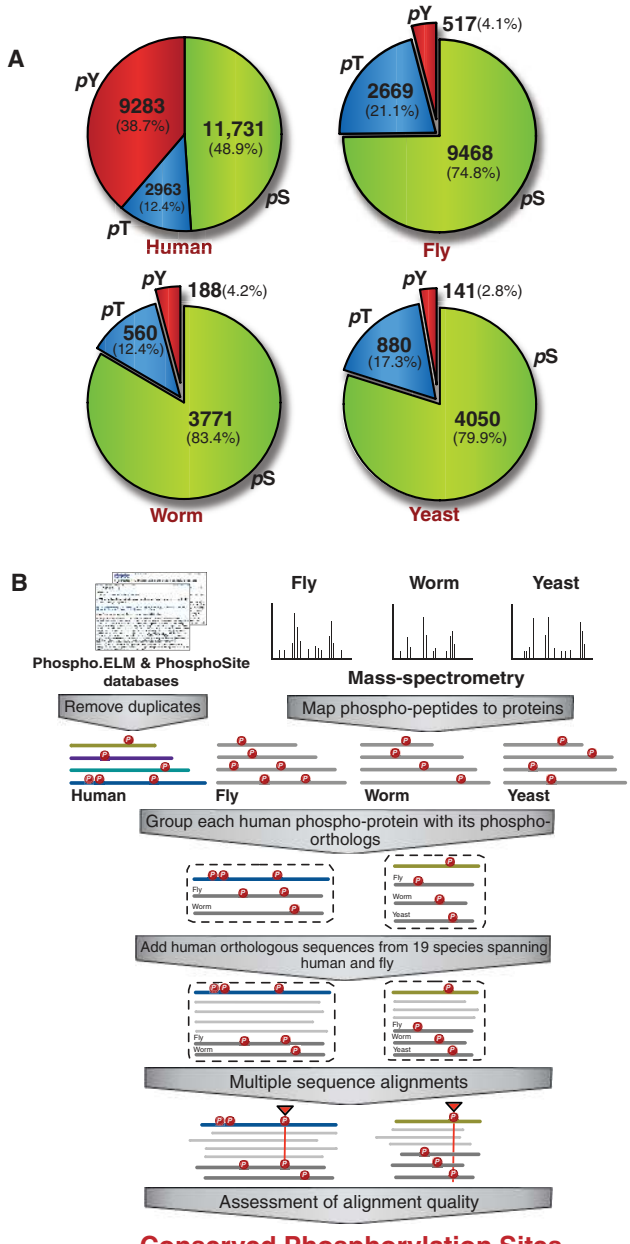

Fig. 1. Human and target species phosphoproteomes. (**A**) A human phosphorylation data set was assembled by combining data from the Phospho.ELM and PhosphoSite databases (*28*, *29*), resulting in 6456 human phosphoproteins (encoded by 6293 genes) with 23,977 sites. The fractions of phosphoserine, phosphothreonine, and phosphotyrosine are indicated. The human data are biased by HTP phosphotyrosine antibody–based studies (*17*, *36*); thus, the phosphotyrosine portion is artificially high. For comparison, we generated phosphoproteomes in three target species with a similar MS and computational pipeline (*15*) (see Supplementary Methods). (**B**) Schematic overview of core site detection. See Methods for further details.

Comparative phosphoproteomics reveal a conserved human phosphorylation core that is implicated in fundamental cellular processes

A total of 23,977 human phosphorylation sites found across 6456 phosphoproteins encoded by 6293 genes were assembled from the two primary online databases PhosphoSite (release 2.0) (*29*) and Phospho.ELM (release 7.0) (*28*). For *D. melanogaster*, *C. elegans*, and *S. cerevisiae*, we used phosphorylation site data that were generated with a similar experimental and computational pipeline (see Methods and Supplementary Materials) and are available via the PhosphoPep database (www.phosphopep.org) (*15*, *35*). Our study included 12,654, 4519, and 5071 phosphorylation sites for *D. melanogaster*, *C. elegans*, and *S. cerevisiae*, respectively.

We observed an exceptionally high fraction of phosphotyrosine sites in the assembled human phosphorylation data that can largely be attributed to HTP phosphotyrosine antibody–based studies (*17*, *36*). The portion of phosphoserine, phosphothreonine, and phosphotyrosine is shown in Fig. 1A.

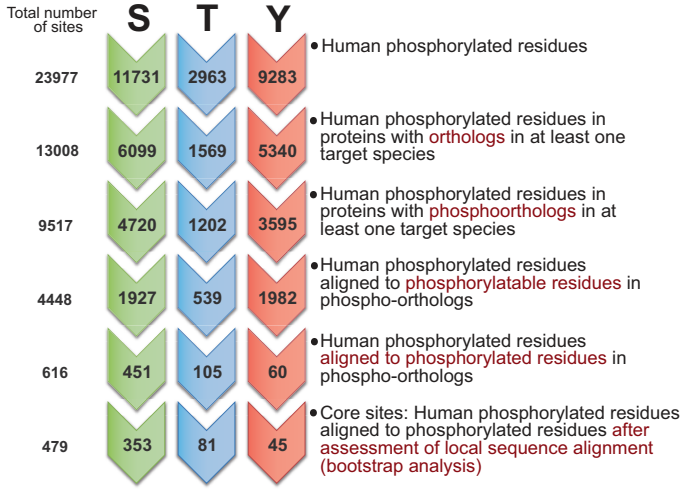

Fig. 2. Number of core sites. The number of human phosphorylation sites left at different stages of the core site detection protocol.

Of all the human sites assembled, 39.7% were in found in proteins orthologous to phosphoproteins detected in at least one target species (Fig. 2). Deploying a sequence-alignment protocol (Fig. 1B, see Methods) with the MAFFT program (*30*) on the three target phosphoproteomes and the human phosphorylation set (see Methods), we identified 479 sites (termed "core sites") that were conserved between human and at least one target species in 344 proteins encoded by 337 human genes (termed "core site genes," Fig. 3A). Of these core sites, 73.7% are phosphoserines, 16.9% are phos-

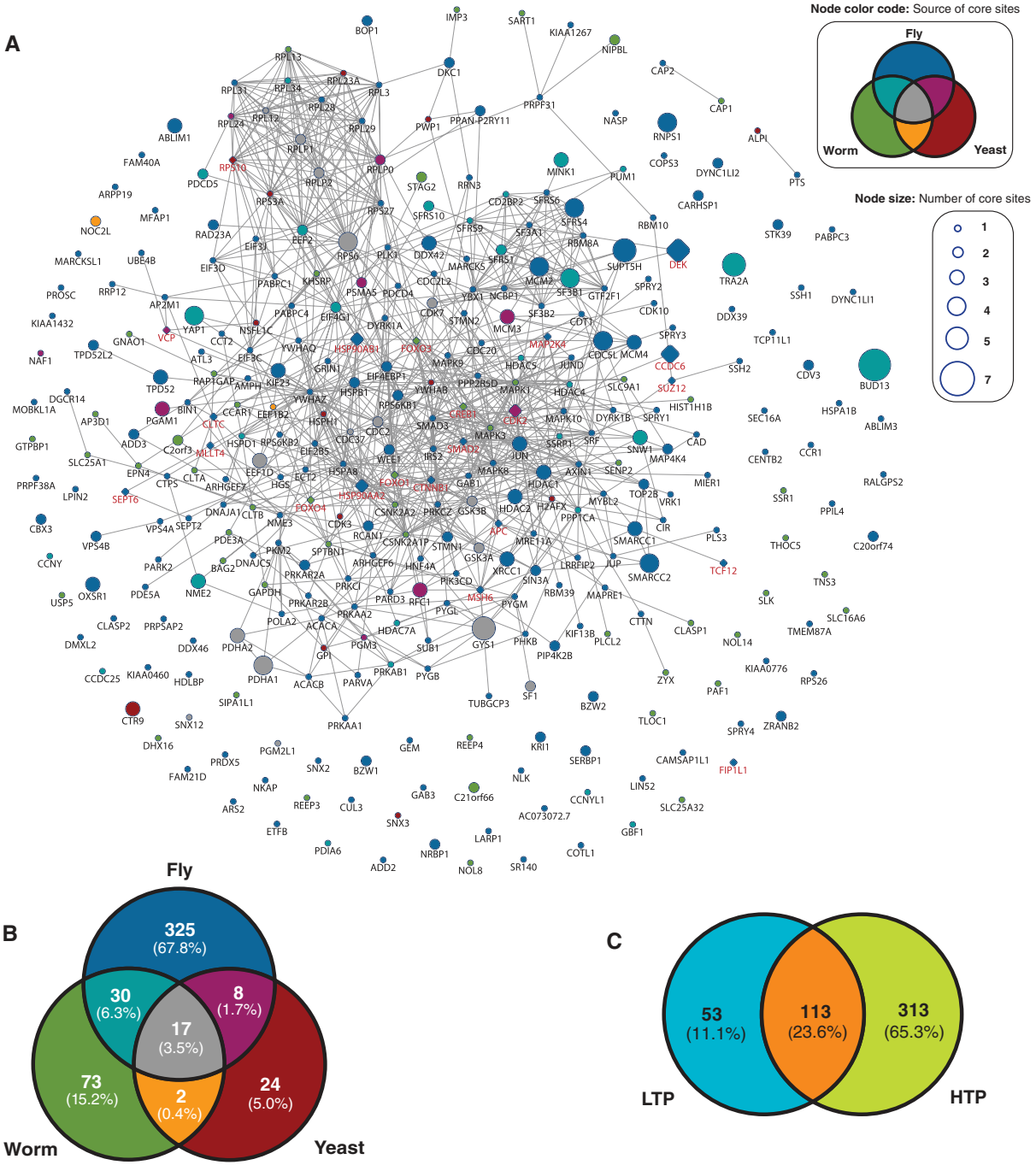

Fig. 3. Protein association map of the human phosphorylation core. (**A**) The comparative approach identified 479 phosphorylation sites in 344 proteins (mapping to 337 human genes). The STRING resource (*66*) was used to construct a functional association network of these proteins (using only high-confidence probabilistic associations). Nodes represent genes and are colored according to which target species contains the conserved phosphorylation site(s) and node size indicates the number of core sites that the encoded proteins have. Known cancer-associated genes are highlighted by red text and diamond nodes (see table S6). The underlying data are provided in the Supplementary Data. (**B**) Target species source of the 479 core sites. Overlap analysis of the target species showed that these data are likely incomplete. (**C**) Eighty-nine percent of the core sites were identified by recent HTP studies.

phothreonines, and 9.4% phosphotyrosines (Fig. 2). These sites make up 10.8% of the 4448 human phosphoresidues that were aligned to phosphorylatable residues in at least one target species, and in most cases, these numbers are significantly higher than expected by random chance from observed alignments (tables S1 to S3).

Among the 479 sites, 139 (~29%) were found within 75 protein domain families (compared to the global average of ~20% for all 29,977 human phosphorylation sites), 57 were conserved in at least two target species, and 17 were conserved in all three target species (Fig. 3B). We observed that core sites shared between humans and more than one target species have an increased tendency to be located within protein domains: 9 of the 17 omnipresent core sites occurred in domains from 6 families (dehydrogenase E1, phosphoglucomutase-phosphomannomutase, glycogen synthase, PhoX homologous, Cdc37 N-terminal kinase binding, 60S acidic ribosomal, and serine-threonine protein kinase catalytic domain), suggesting that the phosphorylation of these protein domains is of ancient origin. It should be noted that not all core sites are phosphorylated by kinases; for example, phosphorylation of the core site Ser^{175} in the phosphoglucomutase domain of human glucose-1,6-bisphosphate synthase likely happens by self-catalysis.

To analyze the functional context of core site genes, we constructed a functional association network among these genes with the STRING resource (Fig. 3A). This network revealed a tight cluster of functionally associated core site genes that encode components of various protein complexes and signaling networks, as well as singleton genes that were not confidently associated to any other core site gene. The β-catenin destruction complex and clathrin coat proteins of coated pits appear to be heavily regulated by protein phosphorylation of ancient origin because they contain core sites in four out of four and four out of five of their conserved protein components, respectively (tables S4 and S5). Function enrichment analysis with Gene Ontology (*37*) annotation revealed that core site genes are involved in fundamental cellular processes. For example, amino acid phosphorylation, RNA splicing, cell division, and translation were statistically enriched over the super set of human phosphoproteins that have orthologs in target species ($P < 0.05$, hypergeometric test, Benjamini and Hochberg false discovery rate correction; see the Supplementary Data). Thus, the observed enrichment suggests that even processes not previously appreciated as regulated by phosphorylation, such as the phosphorylation-mediated regulation of many RNA splicing proteins observed in human cells, arose early during evolution before the last common ancestor of fly and human.

Most core sites were only recently discovered by large-scale phosphoproteomics

Tracing the experimental sources of the core sites, we found that 65.3% of the core sites were detected in HTP experiments reported in the past 5 years (Fig. 3C) (*13*–*18*, *28*, *29*). Moreover, some of these newly discovered and highly conserved sites appear in extensively studied proteins. For example, Thr^{187} in human Wee1 (a major cell cycle regulator kinase) and Ser^{502} in human EEF2 (an essential factor for protein synthesis) with highly conserved flanking regions (defined as the −5 to +5 positions of a phosphorylated residue) of 80% and 100% identity, respectively, were conserved from human to fly. These observations suggest that our systematic and comparative approach reveals important clues for deciphering the functional phosphoregulatory events that occur in fundamental cellular processes.

Matching kinases to the human core sites provides insight into their regulation

The NetPhorest atlas, which currently consists of 179 probabilistic classifiers trained from known relationships between kinases and phosphorylation sites and in vitro proteomics experiments (*2*), matches experimentally validated phosphorylation sites to probabilistic sequence models of kinase consensus (specificity) motifs. To gain further insight into the regulation of core sites, we deployed the NetPhorest algorithm to delineate the kinases or kinase families that are likely to target human core sites. Although many phosphorylation sites can be targeted by multiple kinases or kinase families (*2*, *28*), here, we restricted our analysis to the top three predictions from NetPhorest that exceeded previously calibrated thresholds (*2*) (see Supplementary Data).

We found that CDK2 and CDK3 kinase family and casein kinase 2 (CK2) were the most frequently predicted kinases, each matching ~29% of the human core sites. In comparison, only ~8% and ~6% of all human phosphorylation sites were matched to CDK2 and CDK3 kinase family and CK2, respectively. The high proportion of core sites predicted to be targeted by CDK2 and CDK3 kinase family and CK2 is not unexpected, because these kinases are functionally pleiotropic (*34*) and are involved in several fundamental cell processes such as cell survival, proliferation, and differentiation. In addition, we found that kinases involved in the cellular response to stress, such as p38 and c-Jun N-terminal kinase (JNK) family members were predicted to phosphorylate ~24% and 19% of the core sites (compared to ~7% and ~5% for all human phosphorylation sites), respectively. Al-

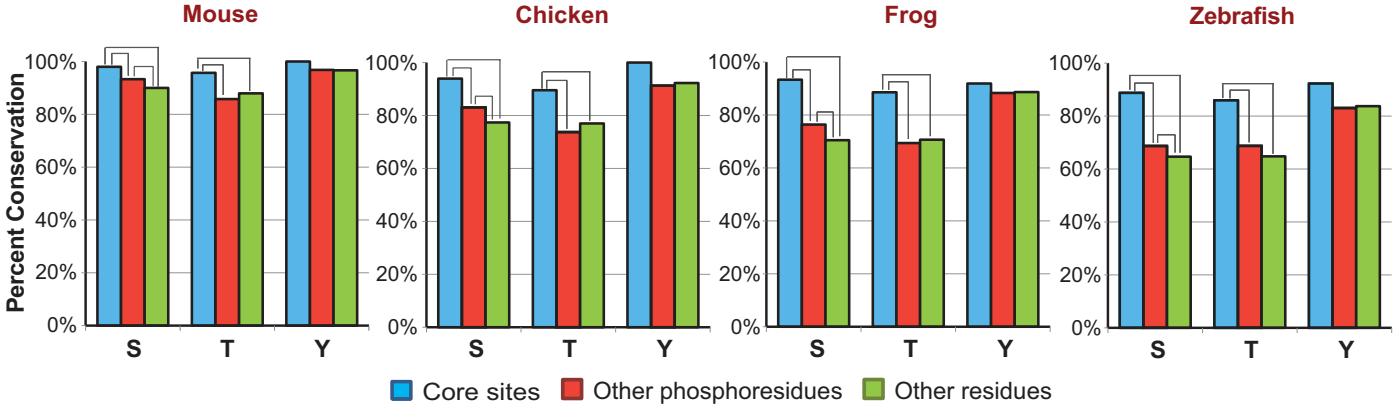

Fig. 4. Conservation of core sites. Proportion of conserved residues for different subsets of serine (S), threonine (T), and tyrosine (Y) in human core site proteins across orthologous proteins in selected species (*M. musculus*, *G. gallus*, *X. tropicalis*, and *D. rerio*) computed from MSAs (see Methods). Only human phosphoresidues with at least 20% identity from sequence position −5 to +5 of the residue (excluding position 0) to the orthologous sequence in the target species are included in the statistics. "Other residues" refers to those instances of the specified amino acid that are not known to be phosphorylated. Connectors linking two bars denote that the difference observed is statistically significant. $P < 0.05$, Fisher's exact test, one-tailed.

though one might expect ancient kinase families to target the core sites, we did not find strong evidence supporting this. Highly conserved core sites (sites with at least 80% sequence similarity within the flanking region) were predicted to be targeted by kinases of different evolutionary origin, such as the insulin receptor (InsR), Eph family members EphA3 through 6, and the nonreceptor tyrosine kinase Src (all of metazoan origin), and phosphoinositide kinase 1 (PDK1), serum- and glucocorticoid-inducible kinase (SGK), and NEK3 [NIMA (never in mitosis gene a)–related kinase 3] through 5 and 11 (all of primordial origin) (*39*).

Core sites are under evolutionary constraint

Tracing the conservation of the 479 core sites across 19 eukaryotic species spanning between human and the target species in evolution confirmed that the core sites are highly conserved, implying that many core sites are under negative selection and are likely important for fundamental cellular processes. For example, we found that 92.3% of the human core site phosphoresidues were preserved in the distantly related *Xenopus tropicalis* compared to 73.6% of other phosphorylatable residues between the same species. When human and mouse (*Mus musculus*) were compared, these numbers were 97.8% and 90.4%, respectively (Fig. 4 shows the conservation for the respective residue in selected species). Human tyrosine residues in general are highly conserved probably because of their roles in maintaining protein structure; thus, core site phosphotyrosines do not appear much more conserved than other tyrosines (Fig. 4).

Although changes in the flanking regions could reveal diverged sequence specificities of kinases that have evolved from yeast to human, such an anal-

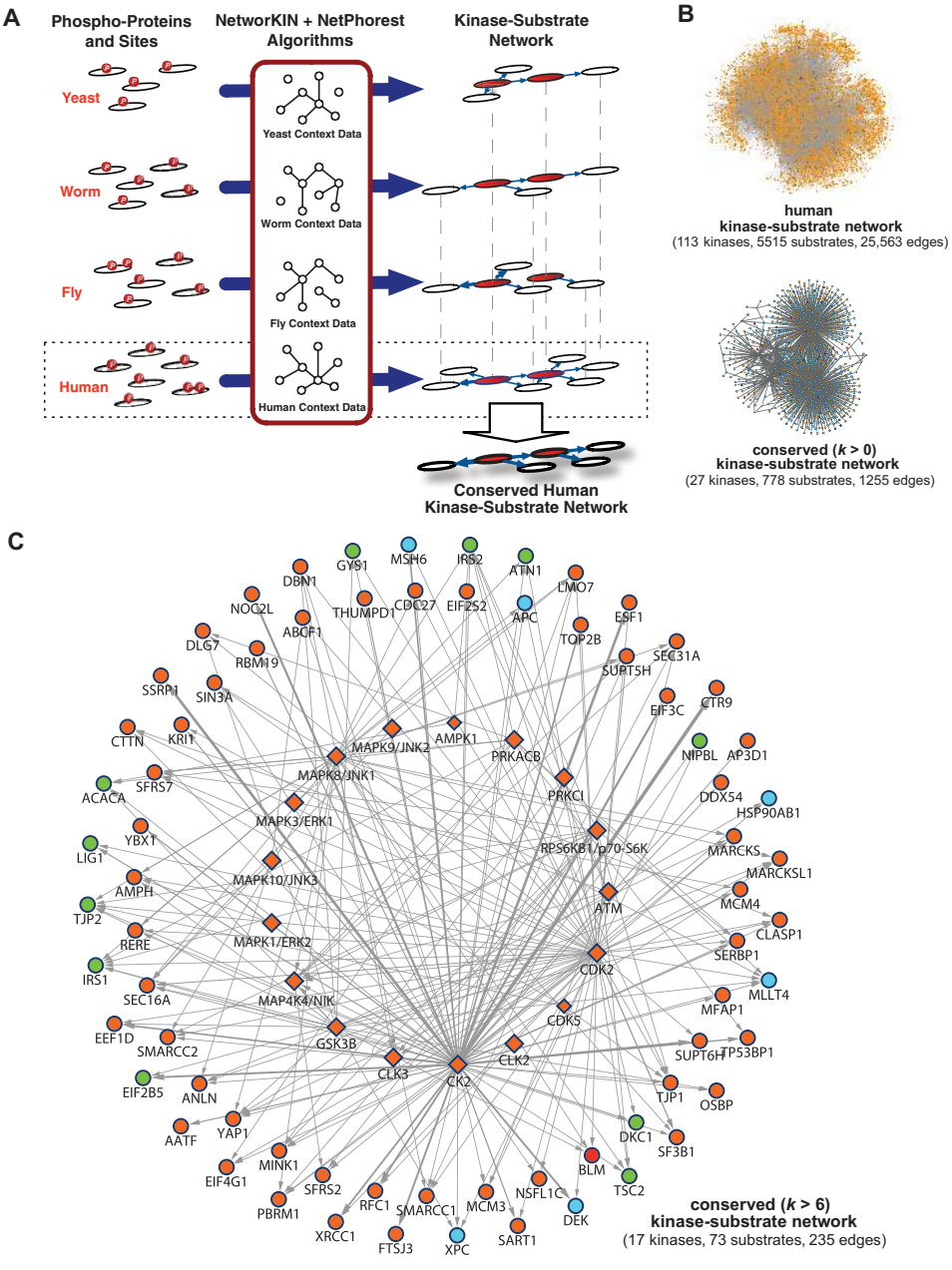

Fig. 5. Phosphorylation core net. Phosphorylation generates dynamic protein interaction networks; therefore, we analyzed conservation of phosphorylation at the network level rather than the positional (site) level. (**A**) First, we applied the NetworKIN and NetPhorest algorithms to the target species phosphoproteomes to reconstruct kinase-substrate networks. Second, interactions within these networks were superimposed (or aligned) with each other. Finally, for each substrate, we defined a phosphorylation conservation propensity *k* of the number of phosphorylation events supported by orthologous kinase-substrate phosphorylation in the target species. (**B**) The initial and increasingly conserved human phosphorylation networks. (**C**) Increasingly conserved human phosphorylation networks could be isolated on the basis of increasing *k*. Here, we show a conserved human phosphorylation network of $k > 6$. The thickness of the edges corresponds to the number of conserved interactions between the kinase and substrate across the target species. Diamond nodes represent kinases predicted to target the phosphoproteins. Proteins known to be implicated in cancer and other diseases are colored blue and green, respectively.

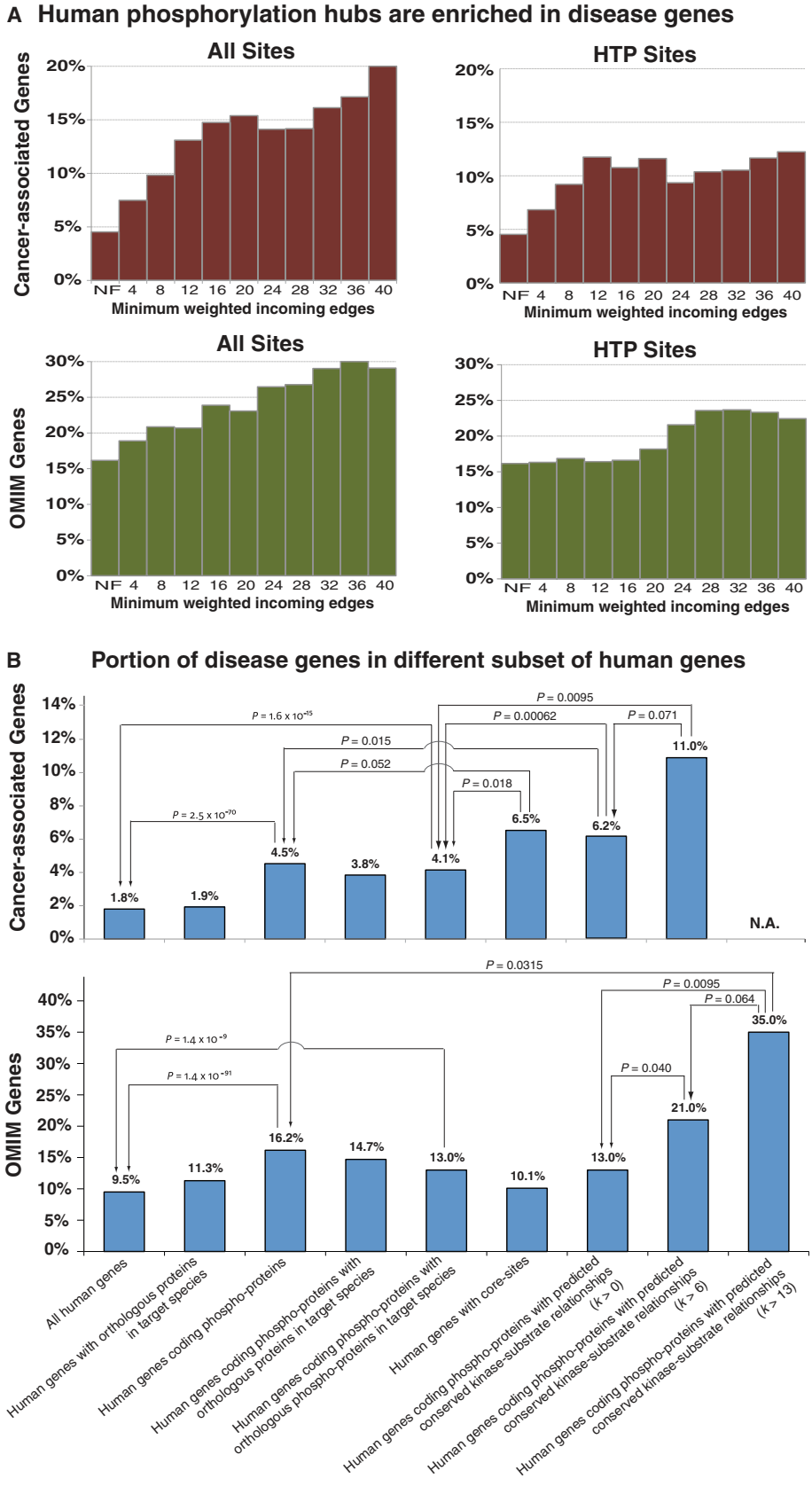

Fig. 6. Disease gene analysis of human phosphorylation networks. (**A**) Genes encoding human signaling hub proteins are enriched in disease genes. A directed kinase-substrate regulatory network is first inferred from assembled human phosphorylation data by NetworKIN. A phosphorylation propensity score *n* is computed for each gene, which is the sum of weighted incoming edges of kinases phosphorylating the gene's products. The weight of an incoming edge from each kinase to a gene is defined as the number of sites in the gene's products inferred to be targeted by a kinase. Human genes are then filtered by this score *n* to assess association with cancer-associated and disease genes from OMIM. *n* is computed from the entire set of human phosphorylation, as well as its subset from HTP studies. NF, not filtered. (**B**) Although human phosphoproteins are enriched in cancer-related genes, both core site genes and core net genes ($k > 0$) are statistically more enriched in cancer-associated genes than background phosphoproteins (top: hypergeometric test, with the protein group of the arrow target used as background). In addition, we observed that core net genes with a higher *k* are more enriched in cancer-associated genes. A similar trend is observed for other disease-associated genes (bottom). These results suggest that aberrant signaling through conserved phosphorylation networks contributes to disease.

ysis is confounded by the possibility that phosphorylation sites can evolve independently from their effector kinases. For example, the presence of multiple sites on a single substrate targeted by a single kinase could create functional redundancy that allows mutations to accumulate in the phosphorylation sites (*31*). Thus, using NetPhorest, we instead analyzed the conservation of kinase (or kinase family) consensus motifs matching the core sites between human and the target species. We estimated the proportion of aligned core site pairs with sufficient conservation within the flanking region for the kinase (or kinase family) predicted for each site of an aligned pair to be identical. This revealed that 67.4% of the aligned site pairs had identical kinases (or families) assigned, and 70% of these sites were predicted to be targeted by the CDK2, CDK3, or CK2 kinases (see Supplementary Data).

Relaxing the analysis to include the top two or three best predictions showed that 81.6% and 86.8% of the core site pairs shared the same kinases or kinase families, respectively (see Supplementary Data). The kinases that regulate the remaining (~13 to 18%) core sites may have changed during evolution. This potential rewiring of the core phosphorylation networks could enable cells to utilize the same core sites to relay signals from different kinases in response to new environmental cues or stimuli. However, we cannot conclusively argue this point because we do not have consensus motifs for all kinases and thus may miss pairs of aligned sites that match conserved but hitherto unknown phosphorylation (kinase consensus) motifs. To explore this further, we performed an orthogonal analysis by clustering core sites on the basis of sequence similarity within their flanking regions between human and target species to identify potential previously unknown phosphorylation (kinase consensus) motifs. First, we grouped aligned sites that shared similar conserved flanking residues (see Methods). Next, we visualized the grouped core sites as sequence motif logos (*2*) and manually organized them into proline-based, arginine-based, and acidic-based phosphoserine or phosphothreonine motifs (fig. S1). Most of the revealed motifs resembled known human kinase consensus motifs (*2*), such as that of PDK1, suggesting the possibility of exploiting comparative phosphoproteomics to discover kinase consensus motifs.

Non–core sites are implicated in a conserved regulatory network

Linear motifs, such as phosphorylation sites, often reside in disordered regions that can change rapidly or undergo convergent evolution (*19*–*23*). We observed that ~50% of human phosphorylation sites in proteins with orthologous phosphoproteins were not aligned to phosphorylatable residues in any of the target species (Fig. 2) and that ~64% of the sites in these proteins were located in intrinsically disordered regions, in agreement with previous reports (*40*, *41*). These observations suggest that many phosphorylation sites are fast evolving and, therefore, do not exhibit strong evolutionary conservation at the sequence position level in distantly related organisms. Even within well-aligned regions (as assessed by the overall sequence identity of residues flanking the phosphorylatable residues), we noticed that phosphorylatable residues in disordered regions are less conserved than phosphorylatable residues in ordered regions (fig. S2). These properties render phosphorylation sites (linear motifs) located in disordered regions difficult to align and trace during evolution (*24*–*27*), which is further supported by the observation that core sites in disordered regions were underrepresented (fig. S3).

A key role of kinases is to modulate cellular signaling networks (for example, by creating binding sites for SH2 domains). Because these events may not require phosphorylation events to occur at precise positions in protein sequences (*21*–*23*, *31*), we investigated the evolutionary conservation of phosphorylation at the level of protein networks rather than strictly focusing on the positionally conserved sites in individual proteins. Specifically, we sought to identify phosphorylation events on orthologous proteins that are mediated by orthologous kinases between human and the target species.

The NetworKIN algorithm can computationally reconstruct phosphorylation networks (*34*) by modeling kinase specificity from contextual information for phosphoproteins and kinases in tandem with sequence models of kinase consensus motifs. The kinome coverage of NetworKIN was extended by the NetPhorest atlas (*2*). A potential concern in using these tools on nonhuman data relates to whether the orthologous kinases in yeast, worm, and fly have similar consensus motifs. NetworKIN made reliable predictions in yeast (*34*) and for several yeast kinases with known human orthologs, the motifs appear identical (B. Turk, personal communication), which is in agreement with the observations reported above for core sites. Furthermore, this conservation of kinase consensus motifs is expected from evolutionary principles: Consensus motifs of pleiotropic kinases (*34*) must be under strong selective pressure because a motif change could potentially affect the complete target and function space for that kinase. Finally, NetworKIN filters predictions on the basis of context; thus, even if a motif falsely matches a site in a target species, it is not likely that the context data would allow inclusion of this prediction.

By deploying the NetworKIN (*34*) algorithm in combination with NetPhorest (*2*), we predicted protein kinases for all phosphoproteins identified in human and the target species. We used the default parameters for NetworKIN (see Methods), which allows a single site to be phosphorylated by multiple kinases and then overlaid the human phosphorylation network with those of the target species (Fig. 5A) to obtain a human phosphorylation network limited to those phosphoproteins and kinases that were conserved in at least one target species (core net). We further quantified phosphorylation conservation by defining a propensity (denoted as k) for each human substrate, which represented the number of a human substrate's phosphorylation events that were supported by orthologous kinase-substrate relationships in target species (fig. S4). Thus, k captures the phosphorylation events on a human protein that are supported by orthologous (conserved) kinase-substrate relationships predicted in the target species. Due to gene duplication that occurred along the lineages of human and target species, multiple kinase-substrate relationships in human may be supported by single kinase-substrate relationship in target species. Conversely, a single kinase-substrate relationship in human may be supported by multiple kinase-substrate relationships in target species.

The initial ($k \geq 0$) human phosphorylation network contained 25,563 interactions between 113 kinases and 5515 substrates (Fig. 5B, top panel), whereas the human phosphorylation network resulting from overlaying the networks from the target species, for $k > 0$, had 1255 interactions between 27 human kinases and 778 substrates (encoded by 759 genes, termed "core net genes") (Fig. 5B, bottom panel), of which 1105 interactions (88%) and 698 substrates (encoded by 682 genes) were not attributed to core sites. Randomized network analysis (see Methods) revealed that this overlap was unlikely to occur by random chance (empirical $P < 0.001$; fig. S5).

Core site genes and core net genes are enriched in genes associated with cancer

The two methods yielded different but somewhat overlapping sets of genes. The alignment-based approached identified the 337 core site genes and the kinase-substrate, network-based approach identified the larger set of the 759 core net genes, which included 525 genes that were not part of the core site gene set. We analyzed each of these gene sets to determine if they were enriched in genes associated with cancer (Fig. 6).

First, human genes encoding phosphoproteins were statistically enriched in cancer-associated genes (see Methods) over the entire protein set (4.5% versus 1.8% background, $P < 0.05$, hypergeometric test; Fig. 6B, bottom). However, the core site gene set was more enriched in cancer-associated (see

Methods) genes over the entire set of genes encoding phosphoproteins ($P = 0.05$, hypergeometric test; Fig. 6B). The 22 cancer-associated genes include *FOXO1*, 3, and 4, *CREB1*, *SMAD2*, and *HSP90* (table S6). This enrichment occurred despite the fact that the subset of human genes encoding phosphoproteins with orthologous proteins in target species was not more enriched in cancer-associated genes than the entire set of human phosphoproteins, regardless of whether their orthologous proteins are phosphorylated (4.1% versus 4.5%) or not (3.8% versus 4.5%). We speculate that some core sites in the products of these genes may be aberrantly regulated in transformed cells. For example, phosphorylation of the core site Ser^{315} in FOXO3A by SGK1 prevents FOXO3A from inducing cell cycle arrest and apoptosis (*38*), thereby promoting cell proliferation. Hence, it is plausible that deregulated phosphorylation of Ser^{315} in FOXO3A could contribute to neoplastic growth. That 15 core sites in these cancer-associated genes were only recently detected in large-scale MS experiments (table S6) suggests that investigation of these sites may provide clues to further understand the functional role of these proteins in normal and malignant cells.

Similarly, core net genes were statistically enriched for cancer-associated genes ($P = 1.5 \times 10^{-2}$ when compared to all human genes encoding phosphoproteins and $P = 6.2 \times 10^{-4}$ when compared to human genes encoding phosphoproteins with orthologous phosphoproteins in the target species, hypergeometric test; Fig. 6B, top graph), identifying approximately one fold more cancer-associated genes than did the alignment-based method (47 versus 22) with a slight drop in specificity (6.5% versus 6.2%; Fig. 6B, top graph). This suggests that the network comparison approach can identify potentially important phosphorylation events occurring in less conserved protein regions. Note that the predicted conserved effector kinases of phosphoproteins from the 759 genes were not included in the enrichment analysis unless they were among the 759 genes. In total, 52 unique cancer-associated genes were identified in the combined set of core site and core net genes (see Supplementary Data).

Highly connected regulatory hubs in the core net are associated with multiple complex diseases

Analysis of the topological features of predicted human phosphorylation networks revealed that the number of kinase-substrate relationships of human phosphoproteins correlated positively with enrichment in cancer-associated genes in the entire human phosphoproteome, as well as to a lesser degree in its HTP subset (Fig. 6A, top graphs). A weaker positive correlation was observed for other diseases in general, as defined in Online Mendelian Inheritance in Man (OMIM) (Fig. 6A, bottom graphs). Hence, a highly phosphorylated regulatory hub protein is more likely to be encoded by a gene implicated in disease. Moreover, there seems to be a strong linear correlation between this likelihood and the signal integration properties of the proteins. A concern was that this observation could stem from ascertainment bias because disease genes are extensively studied. We therefore interrogated the human phosphoproteome with the conservation phosphorylation propensity (k) associated with each human protein, given that the measure is computed with target phosphoproteomes from unbiased systematic studies. We found that k correlated positively with cancer-associated genes and OMIM genes (Fig. 6B). Thus, it appears that genes encoding proteins that receive and integrate many signaling events have an increased tendency to be implicated in disease, which agrees with similar suggestions (*42*, *43*), and that their signal integration properties are conserved in the target species. Another possibility is that these genes encode products that need to be tightly regulated by protein phosphorylation in human and target species and that are vulnerable to deregulation likely caused by mutations or changes in protein abundance.

Accordingly, we identified in the core net proteins that are involved in several complex diseases, which may be suitable for experimental and therapeutic studies. A complete list is available in the Supplementary Data; several examples of proteins that are predicted substrates for kinases involved in various diseases are mentioned here. We identified proteins related to Alzheimer's disease, SEPT1 ($k = 4$) and DBN1 ($k = 7$), which are supported by evidence that misregulation of phosphorylation is important in neurological disorders (*44*). We identified proteins related to viral infection, the human immunodeficiency virus 1 (HIV-1) infection–related proteins SFRS2, SFRS5, and SFRS7 ($k = 13$, 6, and 13, respectively). We identified proteins associated with the cell polarity, TJP1 and TJP2 and MINK1 ($k = 10$, 17, and 7, respectively). We identified proteins implicated in controlling cell and organ size, the Hippo-associated protein YAP1 ($k = 12$), and metabolism, the insulin receptor substrate proteins IRS1 and IRS2 ($k = 16$ and 14, respectively). All these proteins are predicted substrates for the following kinases that are involved in the same set of diseases: CDK2 (cancer and HIV infection), MAP4K4 (cancer and insulin resistance), ATM (cancer), PRKACA and GSK3 (diabetes, cancer, Alzheimer's disease, and HIV), MAPK8 (HIV infection and Alzheimer's disease), and RPS6KB1 (RNA splicing and HIV infection).

Phosphorylation conservation patterns differ by cellular function

Given the disease associations observed in both core site and core net gene sets, we investigated the prevalence of phosphorylation conservation at both the site and the network levels across different cellular functions. That is, we aimed to identify cellular processes in which phosphorylation events are preferentially positionally conserved across orthologs or preferentially mediated by orthologous kinases (conserved kinase-substrate relationship). Specifically, we compared functions of core site genes and core net genes against the complete set of human phosphoproteins that are orthologous to known phosphoproteins in the target species. There are a total of 337 core site genes and 758 core net genes with 233 common genes between the two gene sets. We find that core site genes are statistically enriched (hypergeometric test, Benjamini and Hochberg false discovery rate correction; see Supplementary Data) in genes encoding proteins involved in amino acid phosphorylation ($P = 8.0 \times 10^{-5}$) and RNA splicing ($P = 1.9 \times 10^{-3}$) and encoding cytosolic ribosomal proteins ($P = 2.0 \times 10^{-2}$). Manual inspection revealed that of the core sites present in protein kinases, 26 are located within activation loop regions, which are important for the regulation of kinase activity (fig. S6). Hence, some core sites are structurally constrained for allosteric regulation, suggesting why this particular subset is positionally conserved.

In contrast, core net genes were enriched (hypergeometric test, Benjamini and Hochberg false discovery rate correction; see Supplementary Data) in genes associated with the cell cycle ($P = 1.4 \times 10^{-4}$), chromosome organization and biogenesis ($P = 5.4 \times 10^{-4}$), DNA-dependent regulation of transcription ($P = 3.3 \times 10^{-3}$), macromolecular complex assembly ($P = 2.4 \times 10^{-3}$), and protein targeting ($P = 1.6 \times 10^{-2}$). In 403 out of 688 core net genes with localization annotation, core net genes were strongly enriched in genes encoding proteins that localize to the nucleus ($P = 1.7 \times 10^{-15}$), which correlates with the finding that core net genes are strongly associated with chromosome organization and biogenesis and DNA-dependent regulation of transcription. Correspondingly, our results support the notion that functional conservation of phosphorylation does not necessitate positional conservation: For example, protein phosphorylation in cell cycle–associated proteins in yeast can be conserved, yet dynamic as a result of site relocation (*23*). Correspondingly, our analysis of the core net sites has identified more cellular activities that may be subject to a similar mode of evolution in phosphorylation regulation.

DISCUSSION

To assess the evolutionary history of phosphorylation sites, it is essential to appreciate that the lack of evidence for a phosphorylation event does not infer a nonphosphorylated site. Rather, the site could be phosphorylated but below the limits currently detectable; alternatively, phosphorylation may depend on a missing environmental cue or the site may become dephosphorylated under the experimental conditions used. In addition, some sites are only phosphorylated in specific cell types, rendering their detection even more difficult. Thus, phosphorylation events are highly context dependent (*34*, *45*) and dynamic (*46*). Indeed, Gygi and co-workers derived a phosphoproteome from fly embryos and compared it to the one used here (derived from the Kc-167 cell line) (*47*) and found about 25% overlap despite the fact that the same species was analyzed. Although this difference can partly be explained by different experimental and computational pipelines, it undoubtedly also reflects differences between the biological systems studied (for example, complete embryos contain many specialized cells in contrast to the defined Kc-167 cell line). This highlights that a large number of additional phosphorylation sites are likely to be discovered by continued and improved phosphoproteomic analysis. In particular, studies of the utilization, dynamics, and functional roles of the sites will be important (*48*, *49*), because these reflect cellular information processing much more directly than the number of sites itself.

Although we began with over 40,000 combined phosphorylation sites in both human and the target species (yeast, worm, and fly), we only identified 479 positionally conserved phosphorylation events in human. About 45% (10,969) of the non–core sites were found in human phosphoproteins with no detectable orthologous proteins in any of the target species. Of the remaining 13,008 human phosphoresidues, only 4448 aligned to phosphorylatable residues in at least one target species, of which 479 are aligned to phosphorylated residues (Fig. 2). This limited overlap is presumably due to the large evolutionary distance and actual physiological differences between human and the target species (more than ~600 million years) and the incomplete coverage of the phosphorylation mapping data sets due to mass spectrometer limitations (for example, sensitivity) and the limited number of experimental conditions or biological contexts analyzed (for example, developmental stages).

The incompleteness of the data is illustrated by the composition of the phosphorylated residues of the human phosphoproteome analyzed here (Fig. 1A), which is biased toward pY [39% observed versus an average of about 4% observed in the target species phosphorylation data and in other large-scale phosphoproteomic studies (*18*, *47*)]. This overrepresentation of pY can be attributed to the use of pY antibodies in several HTP studies [for example, (*17*)] and to the notion that phosphotyrosine peptides are more easily detected with MS than are phosphoserine- or phosphothreonine-containing peptides (*14*). Therefore, this pY overrepresentation can be used to estimate a lower bound on the coverage of the human phosphorylation data, by computing how many additional (pS and pT) sites are needed to "dilute" the fraction of pY down to the large-scale average of 4% phosphotyrosine (which is likely an overestimate). Thus, we estimate that there are at least 200,000 more pS and pT sites yet to be discovered in the human phosphoproteome. This estimate will rise with additional discovery of pY sites. A caveat here is that many phosphorylation events are detected in transformed cells, or cells exposed to growth factors or other stimuli (*14*, *18*), which is likely to change the relative amounts of pY, pT, and pS compared to those of nonstimulated cells, as observed by Hunter and co-workers (*50*). Nonetheless, this estimate illustrates the incompleteness of current phosphoproteomic data. In addition, 160 core sites were only found in one target species, although phosphorylatable (S, T, or Y) residues are present at orthologous positions in at least one other target species.

Our analysis supports the notion that many phosphorylation sites evolve quickly (*41*, *51*, *52*) and, therefore, lack strong conservation at the sequence and position levels—65% of human phosphorylated residues in proteins with orthologous proteins in target species were not aligned to phosphorylatable residues. Some of the missed phosphorylation sites could be phosphorylation events that need not be positionally conserved (*21–23*, *32*, *33*). To this end, we investigated the conservation of sites involved in regulatory networks by overlaying predicted kinase-substrate relationships. The two complementary approaches (network versus site alignment) highlight phosphorylation events that are conserved across species spanning long evolutionary distances and, hence, are likely functionally important for fundamental cellular activities. The utility of these approaches is underlined by the identification of multiple low-abundance signaling proteins and disease-related genes. Consequently, we identify genes encoding products that need to be tightly regulated by protein phosphorylation in human and target species and that are vulnerable to deregulation likely caused by mutations or changes in protein abundance. Surprisingly, we did not see any enrichment in disease association in the subset of human phosphoproteins with orthologs that are phosphorylated in our phosphoproteomes over the background set of all human phosphoproteins. In contrast, both the network and the site-alignment approaches identified a subset of genes encoding phosphoproteins that were significantly more enriched in disease-associated genes over the entire set of human phosphoproteins. We also noticed that core site genes were not enriched in OMIM genes compared to the global set of genes encoding phosphoproteins with orthologous proteins in the target species, regardless of whether the orthologous proteins are phosphorylated or not. On the contrary, cross-species signaling hubs among the core net genes (those with high k) had an increased tendency to be implicated in both cancer and other diseases. This suggests that core site genes are implicated in a narrower set of diseases than are core net genes.

Whereas earlier work (*27*) has studied the conservation of phosphorylation sites across diverse species where the data have come from diverse experimental approaches, our work focused on querying human phosphorylation sites with phosphoproteomes from three model organisms generated on a similar experimental and computational platform. This resulted in a higher coverage of conserved phosphorylation events, as exemplified by the identification of substantially more positionally conserved phosphorylation sites identified between human and yeast than the previous study (51 versus 1) (*27*). Our comparison is still relatively rough as we, for example, compare a full multicellular organism (worm), a single-cell organism (yeast), and various cell lines (fly and human). Therefore, we expect the numbers of conserved phosphorylation sites to increase as comparative phosphoproteomics develops in the future. Reports on the investigation of the sequence conservation of phosphorylation sites have reached conflicting conclusions: Gnad *et al*. (*53*) and Malik *et al*. (*25*) reported that experimentally validated phosphorylated residues were more conserved than other phosphorylatable residues; whereas Jimenez *et al*. (*24*) suggested the opposite. Furthermore, it has been suggested that sites identified with large-scale MS are less likely to be functionally important unless they display conservation at the sequence level (*41*, *53*). However, we argue that such strategies will filter away many biologically important phosphorylation sites that need not be positionally conserved. Our network-alignment approach enables studies of phosphorylation events that are not necessarily positionally conserved and underlines the importance of assessing phosphorylation conservation at both site and network level.

In summary, we have systematically investigated the conservation of phosphorylation sites in human regulatory networks by comparison to distantly related model organisms. We identified cross-species phosphorylation events that occur on proteins that have an increased tendency to be implicated in diseases caused by mutations. This result suggests that a simi-

lar approach could be taken to identify networks misregulated in cancer, diabetes, or mental illnesses. We note that multiple diseases seem to converge on the conserved regulatory network (core net). Therefore, we argue that it is important to consider conserved kinase-substrate relationships rather than just conservation of phosphoproteins when searching for disease-related genes. Furthermore, these results suggest that multiple diseases might be targeted using common therapeutic agents (*54*). This idea is supported by a recent study in mice indicating that type 1 diabetes can be suppressed by imatinib (*55*), a small-molecule tyrosine kinase inhibitor developed as a cancer drug. Similar supportive evidence is emerging related to the role of the kinase AMPK (adenosine monophosphate–activated protein kinase) in cancer and diabetes. Therefore, we envisage human regulatory network analysis similar to those used here may be useful for future network medicine endeavors (*56*).

METHODS

Generation of phosphopeptide data

A detailed description of the generation of the phosphopeptide data is provided in the Supplementary Material.

Assembly of nonredundant human phosphorylation site data set

Human phosphorylation sites were collected from the two major online databases PhosphoSite [release 2.0 (*29*)] and Phospho.ELM [release 7.0 (*28*)]. As the two databases use protein sequences from different releases of SwissProt to track the positions of phosphorylation sites, all data were mapped into a reference sequence set from Ensembl [release 46, 2007 (*57*)]. This helped to resolve cases where identical sites had different positions due to revisions of the SwissProt sequence referenced and to identify and remove redundant sites. The mapping between SwissProt primary accessions and its corresponding Ensembl human protein identifiers (release 46) was obtained from Ensembl through its BioMart interface. Finally, the positions of the phosphorylation sites in the Ensembl protein sequences were identified by exact string matching (using the peptide from −7 to +7 surrounding the phosphorylated central residue as defined in the Phospho.ELM or PhosphoSite database). This procedure resulted in 23,977 nonredundant (at 100% identity level) human phosphorylation sites for the comparative analysis.

Identification of phosphorylated orthologs for human phosphoproteins in the three target species

Ortholog information of human phosphoproteins inferred by Ensembl (release 46) ortholog detection pipeline was obtained from Ensembl's BioMart interface. Specifically, Ensembl identifiers of genes orthologous to human genes together with identifiers of their translated protein products were retrieved. The details of the ortholog detection pipeline are described at http://aug2007.archive.ensembl.org/info/data/compara/homology_method.html. Briefly, gene families are identified from all sequences in the database by WU-Blastp and Smith–Waterman searches, followed by construction of a phylogenetic tree for each gene family to identify orthology and paralogy relationships between gene pairs. Finally, we used this information to identify human phosphoproteins with orthologs that were phosphorylated in at least one of the target species (we termed these "phosphoorthologs"). Subsequently, the sequences of these human phosphoproteins were aligned with those of their target species phosphoorthologs to identify conserved core sites.

Identification of core sites

The phosphorylation core sites were detected from multiple sequence alignments (MSAs) of each human phosphoprotein with all its detected phosphoorthologs (as described above). To improve each MSA, we included the protein sequence of the longest splice variant (or an arbitrarily chosen longest if several exist with identical length) of one-to-one orthologous genes from 19 eukaryotic species spanning the evolution between *Homo sapiens* and *D. melanogaster* (*Aedes aegypti*, *Anopheles gambiae*, *Bos taurus*, *Canis familiaris*, *Ciona intestinalis*, *Ciona savignyi*, *Danio rerio*, *Gallus gallus*, *Gasterosteus aculeatus*, *Macaca mulatta*, *Monodelphis domestica*, *Mus musculus*, *Ornithorhynchus anatinus*, *Oryzias latipes*, *Pan troglodytes*, *Rattus norvegicus*, *Takifugu rubripes*, *Tetraodon nigroviridis*, and *X. tropicalis*). For the sake of completeness, we also included the orthologous protein sequences for each target species that had no detected phosphorylation. Finally, these sequences were aligned using the MAFFT (v6.240, E-INS-i option with default parameters) algorithm on an IBM x366 running CentOS (LINUX). The resulting MSAs were subsequently processed by a Perl script to identify the human phosphoresidues that are aligned in the same column with a phosphoresidue observed in any target species (we termed these phosphorylation sites "core sites"). We did not require the aligned phosphoresidues to be identical amino acids to allow detecting cases where one phosphoresidue is converted to another during evolution (for example, pT to pS or pY).

Assessing local alignment quality of core sites with shuffled phosphoortholog sequences

We repeated the MSA with shuffled sequences of phosphoorthologs to identify spurious core sites that could arise from poorly aligned regions in the sequence alignment by random chance alone. First, we identified pairs of aligned phosphoresidues lying in potential poorly aligned regions, which we defined as those having less than 50% identity between human and the target species in the sequence region −5 to +5 (excluding position 0) relative to the human phosphoresidue. For each of these pairs of aligned phosphoresidues, we then computed the BLOSUM62 alignment score between human and target species of sequence region −5 to +5 relative to the human phosphoresidue, and repeated the MSA, as outlined above, 500 times but with the sequence of the phosphoortholog shuffled randomly each time. We then computed the empirical *P* value for the BLOSUM62 computed alignment score of the aligned phosphoresidues as the fraction of trials in which the shuffled phosphoortholog sequence aligned to the same region in the human phosphoprotein to a phosphorylatable residue (S, T or Y) with equal or better BLOSUM62 score than the actual phosphoortholog sequence. Finally, we used these values to only consider core sites that have an empirical *P* value <0.05 resulting in 479 core sites.

Assessing the statistical significance on the number of observed aligned phosphoresidues

We adopted a simple probabilistic model to estimate the statistical significance of the number of observed aligned phosphoserine, phosphothreonine, and phosphotyrosine residues between human and each target species. First, we computed the number of aligned phosphoresidues expected by random chance between human and each target species in the nonshuffled MSA (separate analyses were performed for phosphoserine, phosphothreonine, and phosphotyrosine). Here, we illustrate, as an example, how the number of aligned phosphotyrosines expected by random chance between human and fly was derived: Let *A* be the set of human tyrosine-phosphorylated proteins whose orthologs in fly are tyrosine phosphorylated. Correspondingly, let *B* be the set of fly tyrosine-phosphorylated proteins that correspond to *A*. Next, let *PA* and *PB* be the proportion of tyrosines in protein set *A* and *B*, respectively, that are phosphorylated, and let *NAB* be the total number of tyrosines in *A* that are aligned to tyrosines in *B* as observed in the MSA (described above). It then follows that the number of human phosphotyrosines aligning to phosphotyrosines in fly expected by random chance, assuming joint probability of two independent events, is computed as *PA* ×

PB × *NAB*. Finally, we assessed the statistical significance of the difference between expected random occurrence and observed number of aligned phosphotyrosines by a χ^2 test. Similar analyses were then performed between human and each target species for serine, threonine, and tyrosine separately. The computed statistics are shown in tables S1 to S3.

Clustering analysis of flanking region conservation of core site

For every pair of aligned phosphorylated residues, a consensus sequence of the local alignment from −5 to 5 of the aligned phosphorylated residues is first defined. For example, ..RK.SP..D. is the consensus pattern of GTRKG*p*SPLKDE aligned to NERKV*p*SPDEDM. Next, a consensus pattern *S* encoded as a vector set $V = (v_{-5}, v_{-4}, \ldots, v_4, v_5)$ is defined, where vector v_i is a vector of the 20 elements coding for number of specific amino acids appearing at position *i* among the consensus sequences. The similarity between vector set V_x and V_y is computed as the sum of cosine similarities of all corresponding vectors across the two sets, as follows. First, a vector set is encoded for every consensus sequence. Second, the similarity between pairs of vector sets are computed, and the most similar pair is then merged into a new vector set by summing up the corresponding vectors across the two old sets. The previous step is iteratively performed, and if the two most similar vector sets at each iteration encode 10 or more core sites, they are output and removed from further computation. Lastly, core sites in human and target species represented by output vector sets are then visualized separately with sequence logo for manual inspection and classification (fig. S1).

Prediction of intrinsic protein disorder

We used the DISOPRED2 predictor [http://bioinf.cs.ucl.ac.uk/disopred/ (*58*)] to predict disordered regions in human protein sequences by inputting these to the predictor. The nonredundant (NR) protein sequence database required for the predictor to run was obtained from the National Center for Biotechnology Information in November 2007. The NR database was filtered for transmembrane protein regions with the *pfilt* program provided with DISOPRED2. Subsequently, we analyzed the output with custom Perl scripts and SQL queries.

Gene Ontology analysis

Gene Ontology (GO) term enrichment analyses were performed with the BiNGO [v2.00 (*59*)] plugin for Cytoscape [v2.5.2 (*60*)]. The GO annotations of human genes were retrieved from Ensembl (release 48) and the statistical significance of overrepresented GO terms was determined with hypergeometric distribution tests (corrected for multiple hypothesis testing with false discovery rate). The statistical significance of GO terms associated with core site genes was estimated by comparing the GO terms of two sets of human genes encoding phosphoproteins: those that have orthologs in at least one target species and its subset of genes that have phosphoorthologs in the target species. The statistical significance of GO terms associated with human core net genes (substrates with inferred conserved kinase-substrate relationships in target species) was estimated by comparing it to the entire set of human genes encoding phosphoproteins that have phosphoorthologs in at least one target species, and the phosphoorthologs that have kinase-substrate relationship predicted by NetworKIN.

Assembly of disease-related gene data set

We obtained a list of cancer-associated genes annotated in four peer-reviewed publications (*61*–*64*) from CancerGene [http://cbio.mskcc.org/cancergenes (*65*)]. The first two publications reviewed genes important in cancer development, maintenance, and metastasis, and the last two reported genes with mutations causally implicated in oncogenesis as observed in primary neoplasms. As the cancer-associated genes reported in (*64*) form the basis of cancer-associated genes in Cancer Gene Census (www.sanger.ac.uk/genetics/ CGP/Census/), we obtained the latest list from the database. Subsequently, the gene symbols and aliases obtained were mapped to Ensembl gene entries with the alias mapping file provided by the STRING database [http://string.embl.de (*66*)], resulting in a final set of 413 cancer-related genes. In addition we assembled a data set of genes involved in genetic diseases from the OMIM database [www.ncbi.nlm.nih.gov/omim/ (*67*)]. These genes were obtained from OMIM and mapped to gene identifiers in Ensembl database (release 46). This resulted in a total set of 2174 human genes associated with disease.

Computational reconstruction of conserved human kinase-substrate networks

We used the NetworKIN algorithm [v2.0b (*34*, *68*)] to predict the kinases that may phosphorylate the phosphorylation sites in the four species (*H. sapiens*, *D. melanogaster*, *S. cerevisiae*, and *C. elegans*), resulting in four directed and weighted kinase-substrate networks. We used default parameters for NetworKIN, setting the ranking score cutoff to 0.7 for human and 0.5 for target species. This setting was an empirical decision made on the basis of the relatively weak association data in worm and fly compared to yeast and human. In addition, we expected the conservation quantitation to counteract spurious protein-protein associations. Many predictions from NetworKIN are based on indirect probabilistic associations of proteins; thus, a direct physical interaction is not an absolute prerequisite for the algorithm to associate a substrate with a kinase. Because STRING utilizes evidence transfer between the target species, our approach will be somewhat biased toward these associations. However, the systematic analysis of the phosphoproteomes of the target species and the use of linear motif from NetPhorest serve as unbiased starting material for the NetworKIN prediction algorithm, minimizing this issue.

Each edge in the networks represents a predicted kinase-substrate relationship. The weight of the edge is proportional to the total number of sites among spliced variants of the substrate gene product predicted to be phosphorylated by the kinase. The human kinase-substrate network is compared across the three target species to infer a network of evolutionary conserved kinase-substrate relationships in human. Each inferred evolutionary conserved kinase-substrate relationship in human is further scored (see fig. S4 for details). For each predicted human kinase-substrate relationship (*a*,*b*), kinase *a* and substrate *b* are orthologous to kinase-encoding gene set A_x and substrate-encoding gene set B_x in target species *x* (fly, worm, or yeast), where A_x and B_x can be an empty set. Let *n* be the edge weight of (*a*,*b*) and m_x be the maximum edge weight among kinase-substrate pairs from A_x and B_x in *x*'s weighted kinase-substrate network. The human kinase-substrate relationship (*a*,*b*) is considered conserved in target species *x* if $m_x > 0$ (kinase-substrate relationship between members of A_x and B_x is predicted by NetworKIN based on phosphorylation data in target species *x*). The conservation score C_x of kinase-substrate relationship (*a*,*b*) across target species *x* is then selected as the smaller number of *n* and m_x. The final conservation score C_{total} of kinase-substrate relationship (*a*,*b*) in human across the three target species is the sum of C_{fly}, C_{worm}, and C_{yeast}. Finally, the conserved phosphorylation propensity *k* of a substrate *b* is calculated as the sum of C_{total} of each conserved kinase-substrate relationship that *b* is implicated in (see fig. S4 for a schematic illustration). Finally, we chose not to compress multiple orthologous kinases into a single node, such as JNK1 and JNK2 into a JNK group (*2*), because it is possible for functional divergence to occur after duplication such that the initial set of substrates targeted by an ancient kinase become uniquely targeted among the duplicated kinases.

Assessing statistical significance of inferred conserved kinase-substrate relationships

To assess the statistical significance of the human kinase-substrate relationships inferred to be conserved in the target species, we repeated the procedure described above 2000 times, using randomized kinase-substrate networks of the three target species with the predicted human kinase-substrate network. Each time, randomized kinase-substrate networks in target species are created by switching all originally predicted substrates of each kinase with that of another randomly selected kinase within the same species. The empirical P value is then computed as the fraction of trials that have the same or more inferred conserved human kinase-substrate relationships than the original analysis. The distribution of the number of observed human kinase-substrate relationships from the trials is shown as a box plot in fig. S5.

SUPPLEMENTARY MATERIALS

www.sciencesignaling.org/cgi/content/full/2/81/ra39/DC1
Methods
Fig. S1. Clustering analysis of core phosphorylation sites.
Fig. S2. Phosphorylatable residues in disordered regions are fast evolving.
Fig. S3. Phosphorylation site disorder analysis.
Fig. S4. Schematic diagram of how the conserved phosphorylation propensity *k* of each human substrate is computed.
Fig. S5. Box plot of number of human substrate relations observed as conserved in target species from randomized trials.
Fig. S6. Core sites observed in activation loops of protein kinases.
Table S1. Statistical significance of core sites observed between human and yeast.
Table S2. Statistical significance of core sites observed between human and fly.
Table S3. Statistical significance of core sites observed between human and worm.
Table S4. Core sites identified in components of the β-catenin destruction complex.
Table S5. Core sites identified in components of the clathrin coat of coated pits.
Table S6. Core site identified in 22 cancer-associated genes.
Table S7. Correlation of cancer-associated genes with conserved phosphorylation propensity *k* computed for individual target species.
Table S8. Correlation of OMIM genes with conserved phosphorylation propensity *k* computed for individual target species.
References
Supplementary Data

REFERENCES AND NOTES

1. T. Hunt, Protein sequence motifs involved in recognition and targeting: A new series. *Trends Biochem. Sci.* **15**, 305 (1990).
2. M. L. Miller, L. J. Jensen, F. Diella, C. Jørgensen, M. Tinti, L. Li, M. Hsiung, S. A. Parker, J. Bordeaux, T. Sicheritz-Ponten, M. Olhovsky, A. Pasculescu, J. Alexander, S. Knapp, N. Blom, P. Bork, S. Li, G. Cesareni, T. Pawson, B. E. Turk, M. B. Yaffe, S. Brunak, R. Linding, Linear motif atlas for phosphorylation-dependent signaling. *Sci. Signal.* **1**, ra2 (2008).
3. J. E. Dueber, B. J. Yeh, R. P. Bhattacharyya, W. A. Lim, Rewiring cell signaling: The logic and plasticity of eukaryotic protein circuitry. *Curr. Opin. Struct. Biol.* **14**, 690–699 (2004).
4. R. P. Bhattacharyya, A. Reményi, B. J. Yeh, W. A. Lim, Domains, motifs, and scaffolds: The role of modular interactions in the evolution and wiring of cell signaling circuits. *Annu. Rev. Biochem.* **75**, 655–680 (2006).
5. B. T. Seet, I. Dikic, M.-M. Zhou, T. Pawson, Reading protein modifications with interaction domains. *Nat. Rev. Mol. Cell Biol.* **7**, 473–483 (2006).
6. K. Miller-Jensen, K. A. Janes, J. S. Brugge, D. A. Lauffenburger, Common effector processing mediates cell-specific responses to stimuli. *Nature* **448**, 604–608 (2007).
7. J.-F. Rual, K. Venkatesan, T. Hao, T. Hirozane-Kishikawa, A. Dricot, N. Li, G. F. Berriz, F. D. Gibbons, M. Dreze, N. Ayivi-Guedehoussou, N. Klitgord, C. Simon, M. Boxem, S. Milstein, J. Rosenberg, D. S. Goldberg, L. V. Zhang, S. L. Wong, G. Franklin, S. Li, J. S. Albala, J. Lim, C. Fraughton, E. Llamosas, S. Cevik, C. Bex, P. Lamesch, R. S. Sikorski, J. Vandenhaute, H. Y. Zoghbi, A. Smolyar, S. Bosak, R. Sequerra, L. Doucette-Stamm, M. E. Cusick, D. E. Hill, F. P. Roth, M. Vidal, Towards a proteome-scale map of the human protein-protein interaction network. *Nature* **437**, 1173–1178 (2005).
8. A.-C. Gavin, P. Aloy, P. Grandi, R. Krause, M. Boesche, M. Marzioch, C. Rau, L. J. Jensen, S. Bastuck, B. Dümpelfeld, A. Edelmann, M.-A. Heurtier, V. Hoffman, C. Hoefert, K. Klein, M. Hudak, A.-M. Michon, M. Schelder, M. Schirle, M. Remor, T. Rudi, S. Hooper, A. Bauer, T. Bouwmeester, G. Casari, G. Drewes, G. Neubauer, J. M. Rick, B. Kuster, P. Bork, R. B. Russell, G. Superti-Furga, Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440**, 631–636 (2006).
9. U. Stelzl, U. Worm, M. Lalowski, C. Haenig, F. H. Brembeck, H. Goehler, M. Stroedicke, M. Zenkner, A. Schoenherr, S. Koeppen, J. Timm, S. Mintzlaff, C. Abraham, N. Bock, S. Kietzmann, A. Goedde, E. Toksöz, A. Droege, S. Krobitsch, B. Korn, W. Birchmeier, H. Lehrach, E. E. Wanker, A human protein-protein interaction network: A resource for annotating the proteome. *Cell* **122**, 957–968 (2005).
10. C. von Mering, R. Krause, B. Snel, M. Cornell, S. G. Oliver, S. Fields, P. Bork, Comparative assessment of large-scale data sets of protein–protein interactions. *Nature* **417**, 399–403 (2002).
11. P. H. Huang, A. Mukasa, R. Bonavia, R. A. Flynn, Z. E. Brewer, W. K. Cavenee, F. B. Furnari, F. M. White, Quantitative analysis of EGFRvIII cellular signaling networks reveals a combinatorial therapeutic strategy for glioblastoma. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12867–12872 (2007).
12. K. A. Janes, S. Gaudet, J. G. Albeck, U. B. Nielsen, D. A. Lauffenburger, P. K. Sorger, The response of human epithelial cells to TNF involves an inducible autocrine cascade. *Cell* **124**, 1225–1239 (2006).
13. A. Wolf-Yadlin, S. Hautaniemi, D. A. Lauffenburger, F. M. White, Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 5860–5865 (2007).
14. S. A. Beausoleil, M. Jedrychowski, D. Schwartz, J. E. Elias, J. Villén, J. Li, M. A. Cohn, L. C. Cantley, S. P. Gygi, Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 12130–12135 (2004).
15. B. Bodenmiller, J. Malmstrom, B. Gerrits, D. Campbell, H. Lam, A. Schmidt, O. Rinner, L. N. Mueller, P. T. Shannon, P. G. Pedrioli, C. Panse, H.-K. Lee, R. Schlapbach, R. Aebersold, Phosphopep—A phosphoproteome resource for systems biology research in *Drosophila* Kc167 cells. *Mol. Syst. Biol.* **3**, 139 (2007).
16. S. Matsuoka, B. A. Ballif, A. Smogorzewska, E. R. McDonald III, K. E. Hurov, J. Luo, C. E. Bakalarski, Z. Zhao, N. Solimini, Y. Lerenthal, Y. Shiloh, S. P. Gygi, S. J. Elledge, ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. *Science* **316**, 1160–1166 (2007).
17. K. Rikova, A. Guo, Q. Zeng, A. Possemato, J. Yu, H. Haack, J. Nardone, K. Lee, C. Reeves, Y. Li, Y. Hu, Z. Tan, M. Stokes, L. Sullivan, J. Mitchell, R. Wetzel, J. Macneill, J. M. Ren, J. Yuan, C. E. Bakalarski, J. Villen, J. M. Kornhauser, B. Smith, D. Li, X. Zhou, S. P. Gygi, T.-L. Gu, R. D. Polakiewicz, J. Rush, M. J. Comb, Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. *Cell* **131**, 1190–1203 (2007).
18. J. V. Olsen, B. Blagoev, F. Gnad, B. Macek, C. Kumar, P. Mortensen, M. Mann, Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* **127**, 635–648 (2006).
19. R. Linding, R. B. Russell, V. Neduva, T. J. Gibson, Globplot: Exploring protein sequences for globularity and disorder. *Nucleic Acids Res.* **31**, 3701–3708 (2003).
20. P. Puntervoll, R. Linding, C. Gemünd, S. Chabanis-Davidson, M. Mattingsdal, S. Cameron, D. M. Martin, G. Ausiello, B. Brannetti, A. Costantini, F. Ferrè, V. Maselli, A. Via, G. Cesareni, F. Diella, G. Superti-Furga, L. Wyrwicz, C. Ramu, C. McGuigan, R. Gudavalli, I. Letunic, P. Bork, L. Rychlewski, B. Küster, M. Helmer-Citterich, W. N. Hunter, R. Aasland, T. J. Gibson, ELM server: A new resource for investigating short functional sites in modular eukaryotic proteins. *Nucleic Acids Res.* **31**, 3625–3630 (2003).
21. V. Neduva, R. B. Russell, Linear motifs: Evolutionary interaction switches. *FEBS Lett.* **579**, 3342–3345 (2005).
22. A. M. Moses, M. E. Liku, J. J. Li, R. Durbin, Regulatory evolution in proteins by turnover and lineage-specific changes of cyclin-dependent kinase consensus sites. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 17713–17718 (2007).
23. L. J. Jensen, T. S. Jensen, U. de Lichtenberg, S. Brunak, P. Bork, Co-evolution of transcriptional and post-translational cell-cycle regulation. *Nature* **443**, 594–597 (2006).
24. J. L. Jiménez, B. Hegemann, J. R. Hutchins, J.-M. Peters, R. Durbin, A systematic comparative and structural analysis of protein phosphorylation sites based on the mtcPTM database. *Genome Biol.* **8**, R90 (2007).
25. R. Malik, E. A. Nigg, R. Körner, Comparative conservation analysis of the human mitotic phosphoproteome. *Bioinformatics* **24**, 1426–1432 (2008).
26. B. Macek, F. Gnad, B. Soufi, C. Kumar, J. V. Olsen, I. Mijakovic, M. Mann, Phosphoproteome analysis of *E. coli* reveals evolutionary conservation of bacterial Ser/Thr/Tyr phosphorylation. *Mol. Cell. Proteomics* **7**, 299–307 (2008).
27. J. Boekhorst, B. van Breukelen, A. J. Heck, B. Snel, Comparative phosphoproteomics reveals evolutionary and functional conservation of phosphorylation across eukaryotes. *Genome Biol.* **9**, R144 (2008).
28. F. Diella, C. M. Gould, C. Chica, A. Via, T. J. Gibson, Phospho.ELM: A database of phosphorylation sites—Update 2008. *Nucleic Acids Res.* **36**, D240–D244 (2008).
29. P. V. Hornbeck, I. Chabra, J. M. Kornhauser, E. Skrzypek, B. Zhang, PhosphoSite: A bioinformatics resource dedicated to physiological protein phosphorylation. *Proteomics* **4**, 1551–1561 (2004).
30. K. Katoh, H. Toh, Recent developments in the MAFFT multiple sequence alignment program. *Brief. Bioinform.* **9**, 286–298 (2008).
31. L. J. Holt, J. E. Hutti, L. C. Cantley, D. O. Morgan, Evolution of Ime2 phosphorylation sites on Cdk1 substrates provides a mechanism to limit the effects of the phosphatase Cdc14 in meiosis. *Mol. Cell* **25**, 689–702 (2007).

32. P. Nash, X. Tang, S. Orlicky, Q. Chen, F. B. Gertler, M. D. Mendenhall, F. Sicheri, T. Pawson, M. Tyers, Multisite phosphorylation of a CDK inhibitor sets a threshold for the onset of DNA replication. *Nature* **414**, 514–521 (2001).
33. S. Y. Kim, J. E. Ferrell Jr., Substrate competition as a source of ultrasensitivity in the inactivation of Wee1. *Cell* **128**, 1133–1145 (2007).
34. R. Linding, L. J. Jensen, G. J. Ostheimer, M. A. van Vugt, C. Jørgensen, I. M. Miron, F. Diella, K. Colwill, L. Taylor, K. Elder, P. Metalnikov, V. Nguyen, A. Pasculescu, J. Jin, J. G. Park, L. D. Samson, J. R. Woodgett, R. B. Russell, P. Bork, M. B. Yaffe, T. Pawson, Systematic discovery of in vivo phosphorylation networks. *Cell* **129**, 1415–1426 (2007).
35. B. Bodenmiller, D. Campbell, B. Gerrits, H. Lam, M. Jovanovic, P. Picotti, R. Schlapbach, R. Aebersold, PhosphoPep—A database of protein phosphorylation sites in model organisms. *Nat. Biotechnol.* **26**, 1339–1340 (2008).
36. J. Rush, A. Moritz, K. A. Lee, A. Guo, V. L. Goss, E. J. Spek, H. Zhang, X. M. Zha, R. D. Polakiewicz, M. J. Comb, Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. *Nat. Biotechnol.* **23**, 94–101 (2005).
37. M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, G. Sherlock, Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).
38. A. Brunet, J. Park, H. Tran, L. S. Hu, B. A. Hemmings, M. E. Greenberg, Protein kinase SGK mediates survival signals by phosphorylating the forkhead transcription factor FKHRL1 (FOXO3a). *Mol. Cell. Biol.* **21**, 952–965 (2001).
39. G. Manning, G. D. Plowman, T. Hunter, S. Sudarsanam, Evolution of protein kinase signaling from yeast to man. *Trends Biochem. Sci.* **27**, 514–520 (2002).
40. M. O. Collins, L. Yu, I. Campuzano, S. G. N. Grant, J. S. Choudhary, Phosphoproteomic analysis of the mouse brain cytosol reveals a predominance of protein phosphorylation in regions of intrinsic sequence disorder. *Mol. Cell. Proteomics* **7**, 1331–1348 (2008).
41. C. R. Landry, E. D. Levy, S. W. Michnick, Weak functional constraints on phosphoproteomes. *Trends Genet.* **25**, 193–197 (2009).
42. I. W. Taylor, R. Linding, D. Warde-Farley, Y. Liu, C. Pesquita, D. Faria, S. Bull, T. Pawson, Q. Morris, J. L. Wrana, Dynamic modularity in protein interaction networks predicts breast cancer outcome. *Nat. Biotechnol.* **27**, 199–204 (2009).
43. D. Rambaldi, F. M. Giorgi, F. Capuani, A. Ciliberto, F. D. Ciccarelli, Low duplicability and network fragility of cancer genes. *Trends Genet.* **24**, 427–430 (2008).
44. M. P. Mazanetz, P. M. Fischer, Untangling tau hyperphosphorylation in drug design for neurodegenerative diseases. *Nat. Rev. Drug Discov.* **6**, 464–479 (2007).
45. C. Jørgensen, R. Linding, Directional and quantitative phosphorylation networks. *Brief. Funct. Genomic Proteomic* **7**, 17–26 (2008).
46. A. Wolf-Yadlin, S. Hautaniemi, D. A. Lauffenburger, F. M. White, Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 5860–5865 (2007).
47. B. Zhai, J. Villén, S. A. Beausoleil, J. Mintseris, S. P. Gygi, Phosphoproteome analysis of *Drosophila melanogaster* embryos. *J. Proteome Res.* **7**, 1675–1682 (2008).
48. N. Kumar, A. Wolf-Yadlin, F. M. White, D. A. Lauffenburger, Modeling HER2 effects on cell behavior from mass spectrometry phosphotyrosine data. *PLoS Comput. Biol.* **3**, e4 (2007).
49. K. Schmelzle, S. Kane, S. Gridley, G. E. Lienhard, F. M. White, Temporal dynamics of tyrosine phosphorylation in insulin signaling. *Diabetes* **55**, 2171–2179 (2006).
50. T. Hunter, B. M. Sefton, Transforming gene product of Rous sarcoma virus phosphorylates tyrosine. *Proc. Natl. Acad. Sci. U.S.A.* **77**, 1311–1315 (1980).
51. P. Beltrao, J. C. Trinidad, D. Fiedler, A. Roguev, W. A. Lim, K. M. Shokat, A. L. Burlingame, N. J. Krogan, Evolution of phosphoregulation: Comparison of phosphorylation patterns across yeast species. *PLoS Biol.* **7**, e1000134 (2009).
52. C. S. H. Tan, A. Pasculescu, W. A. Lim, T. Pawson, G. D. Bader, R. Linding, Positive selection of tyrosine loss in metazoan evolution. *Science* 9 July 2009 (10.1126/science.1174301).
53. F. Gnad, S. Ren, J. Cox, J. V. Olsen, B. Macek, M. Oroshi, M. Mann, PHOSIDA (phosphorylation site database): Management, structural and evolutionary investigation, and prediction of phosphosites. *Genome Biol.* **8**, R250 (2007).
54. O. Fedorov, B. Marsden, V. Pogacic, P. Rellos, S. Müller, A. N. Bullock, J. Schwaller, M. Sundström, S. Knapp, A systematic interaction map of validated kinase inhibitors with Ser/Thr kinases. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 20523–20528 (2007).
55. C. Louvet, G. L. Szot, J. Lang, M. R. Lee, N. Martinier, G. Bollag, S. Zhu, A. Weiss, J. A. Bluestone, Tyrosine kinase inhibitors reverse type 1 diabetes in nonobese diabetic mice. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 18895–18900 (2008).
56. T. Pawson, R. Linding, Network medicine. *FEBS Lett.* **582**, 1266–1270 (2008).
57. P. Flicek, B. L. Aken, K. Beal, B. Ballester, M. Caccamo, Y. Chen, L. Clarke, G. Coates, F. Cunningham, T. Cutts, T. Down, S. C. Dyer, T. Eyre, S. Fitzgerald, J. Fernandez-Banet, S. Gräf, S. Haider, M. Hammond, R. Holland, K. L. Howe, K. Howe, N. Johnson, A. Jenkinson, A. Kähäri, D. Keefe, F. Kokocinski, E. Kulesha, D. Lawson, I. Longden, K. Megy, P. Meidl, B. Overduin, A. Parker, B. Pritchard, A. Prlic, S. Rice, D. Rios, M. Schuster, I. Sealy, G. Slater, D. Smedley, G. Spudich, S. Trevanion, A. J. Vilella, J. Vogel, S. White, M. Wood, E. Birney, T. Cox, V. Curwen, R. Durbin, X. M. Fernandez-Suarez, J. Herrero, T. J. P. Hubbard, A. Kasprzyk, G. Proctor, J. Smith, A. Ureta-Vidal, S. Searle, Ensembl 2008. *Nucleic Acids Res.* **36**, D707–D714 (2008).
58. J. J. Ward, J. S. Sodhi, L. J. McGuffin, B. F. Buxton, D. T. Jones, Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol.* **337**, 635–645 (2004).
59. S. Maere, K. Heymans, M. Kuiper, BiNGO: A Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* **21**, 3448–3449 (2005).
60. P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
61. W. C. Hahn, R. A. Weinberg, Rules for making human tumor cells. *N. Engl. J. Med.* **347**, 1593–1603 (2002).
62. B. Vogelstein, K. W. Kinzler, Cancer genes and the pathways they control. *Nat. Med.* **10**, 789–799 (2004).
63. F. Mitelman, Recurrent chromosome aberrations in cancer. *Mutat. Res.* **462**, 247–253 (2000).
64. P. A. Futreal, L. Coin, M. Marshall, T. Down, T. Hubbard, R. Wooster, N. Rahman, M. R. Stratton, A census of human cancer genes. *Nat. Rev. Cancer* **4**, 177–183 (2004).
65. M. E. Higgins, M. Claremont, J. E. Major, C. Sander, A. E. Lash, CancerGenes: A gene selection resource for cancer genome projects. *Nucleic Acids Res.* **35**, D721–D726 (2007).
66. C. von Mering, L. J. Jensen, M. Kuhn, S. Chaffron, T. Doerks, B. Krüger, B. Snel, P. Bork, STRING 7—Recent developments in the integration and prediction of protein interactions. *Nucleic Acids Res.* **35**, D358–D362 (2007).
67. P. Hermanns, A. A. Bertuch, T. K. Bertin, B. Dawson, M. E. Schmitt, C. Shaw, B. Zabel, B. Lee, Consequences of mutations in the non-coding RMRP RNA in cartilage-hair hypoplasia. *Hum. Mol. Genet.* **14**, 3723–3740 (2005).
68. R. Linding, L. J. Jensen, A. Pasculescu, M. Olhovsky, K. Colwill, P. Bork, M. B. Yaffe, T. Pawson, NetworKIN: A resource for exploring cellular phosphorylation networks. *Nucleic Acids Res.* **36**, D695–D699 (2008).

69. We thank L. J. Jensen (CPR), K. Colwill (SLRI), S. Quirk, and J. Koh for comments on the manuscript. We are further indebted to B. Turk (Yale), S. Michnick (University of Montreal), P. Beltrao and N. Krogan (University of California, San Francisco) for sharing unpublished results. This project was in part supported by Genome Canada through Ontario Genomics Institute and the Canadian Institutes of Health Research (MOP-84324). We also thank the Functional Genomics Center Zurich for generous support with mass spectrometry resources. This project has been funded in part by ETH Zurich, the Swiss National Science Foundation under grant 31000-10767, with Federal (U.S.) funds from the National Heart, Lung, and Blood Institute, NIH, under contract N01-HV-28179, and by the Center for Model Organism Proteomics of the University of Zurich, the Swiss Initiative for Systems Biology. M.O.H. is the Ernst Hadorn Chair of Molecular Biology. R.A. was supported in part by a grant from F. Hoffmann-LaRoche (Basel) provided to the Competence Center for Systems Physiology and Metabolic Disease. B.B. is the recipient of a fellowship from the Boehringer Ingelheim Fonds. M.J. was supported by a fellowship from the Research Foundation of the University of Zurich and by a fellowship from the Roche Research Foundation. Author contributions: R.L. conceived the project. R.L., C.S.H.T., T.P., R.A., B.B., G.B., and C.J. conceived and designed the experiments. R.L., C.S.H.T., and A.P. performed the computational experiments. B.B. and M.J. performed the proteomics experiments. R.A., C.S.H.T., B.B., C.J., G.B., T.P., and R.L. wrote the paper.

Submitted 10 March 2009
Accepted 6 July 2009
Final Publication 28 July 2009
10.1126/scisignal.2000316
Citation: C. S. H. Tan, B. Bodenmiller, A. Pasculescu, M. Jovanovic, M. O. Hengartner, C. Jørgensen, G. D. Bader, R. Aebersold, T. Pawson, R. Linding, Comparative analysis reveals conserved protein phosphorylation networks implicated in multiple diseases. *Sci. Signal.* **2**, ra39 (2009).