Cell Reports
Resource
ACCESS
CellPress

Ultradeep Human Phosphoproteome Reveals a Distinct Regulatory Nature of Tyr and Ser/Thr-Based Signaling

Kirti Sharma,[1] Rochelle C.J. D'Souza,[1] Stefka Tyanova,[1] Christoph Schaab,[1] Jacek R. Wiśniewski,[1] Jürgen Cox,[1,*] and Matthias Mann[1,*]
[1]Department of Proteomics and Signal Transduction, Max-Planck Institute of Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany
*Correspondence: cox@biochem.mpg.de (J.C.), mmann@biochem.mpg.de (M.M.)
http://dx.doi.org/10.1016/j.celrep.2014.07.036

SUMMARY

Regulatory protein phosphorylation controls normal and pathophysiological signaling in eukaryotic cells. Despite great advances in mass-spectrometry-based proteomics, the extent, localization, and site-specific stoichiometry of this posttranslational modification (PTM) are unknown. Here, we develop a stringent experimental and computational workflow, capable of mapping more than 50,000 distinct phosphorylated peptides in a single human cancer cell line. We detected more than three-quarters of cellular proteins as phosphoproteins and determined very high stoichiometries in mitosis or growth factor signaling by label-free quantitation. The proportion of phospho-Tyr drastically decreases as coverage of the phosphoproteome increases, whereas Ser/Thr sites saturate only for technical reasons. Tyrosine phosphorylation is maintained at especially low stoichiometric levels in the absence of specific signaling events. Unexpectedly, it is enriched on higher-abundance proteins, and this correlates with the substrate K_M values of tyrosine kinases. Our data suggest that P-Tyr should be considered a functionally separate PTM of eukaryotic proteomes.

INTRODUCTION

Protein phosphorylation is of central importance in signaling systems and employed by cells to transiently alter protein properties such as the activity of enzymes, their interactions with other proteins, their localization and conformations, or to target them for destruction. When deregulated, it is also critically involved in disease processes, notably in cancer, where protein kinases are now the major class of drug targets (Rix and Superti-Furga, 2009; Zhang et al., 2009). An essential step in understanding the complex molecular circuitry of cellular signal transmission is to develop methods for measuring the extent and nature of phosphorylation events that occur in a cell. For this purpose, modern quantitative mass spectrometry (MS) has proved to be an ideal platform because it is a highly precise yet generic method for the global identification and quantification of proteins and their modifications (Choudhary and Mann, 2010; Hein et al., 2013; Junger and Aebersold, 2013; Lemeer and Heck, 2009). In eukaryotes, phosphorylation occurs almost exclusively on Ser, Thr, and Tyr residues, which represent approximately 17% of the total amino acids in an average human protein (Echols et al., 2002). On this basis, there are nearly 700,000 different potential phosphorylation sites (Ubersax and Ferrell, 2007). Recent MS-based studies have reported the identification of tens of thousands of phosphorylation sites in tissues and cultured cells (Humphrey et al., 2013; Lundby et al., 2012; Zhou et al., 2013), and it has been speculated that the total extent of the phosphoproteome is more than a million (Boersema et al., 2010). This vast and increasing number of identified phosphosites raises fundamental question about their properties and biological relevance and the scale of the complete phosphoproteome.

Based on autoradiography measurements S/T/Y phosphorylation ratios had been estimated many years ago as 90:10:0.05 (Hunter and Sefton, 1980). The vast majority of cellular protein phosphorylation events reported in MS-based studies occurs on Ser and Thr residues, whereas phosphotyrosines (P-Tyr) generally account for less than 1% of the identified sites. This is despite a large number of Tyr kinases encoded in the genome. There are several reasons for the relatively low number of Tyr phosphorylation events (Hunter, 2009). First, most Tyr kinases are only activated in specific circumstances and otherwise remain stringently negatively regulated. Second, unless protected by binding to SH2 or PTP domains (Sadowski et al., 1986), P-Tyr residues have a short half-life owing to high activity of phosphotyrosine phosphatases (PTPs). Finally, P-Tyr is primarily regulatory and rarely plays a structural role in proteins. P-Tyr based signaling pathways are a comparatively recent molecular innovation in evolutionary history and appear to be a hallmark of more complex organisms (Lim and Pawson, 2010). As a result of the relatively low proportion of Tyr phosphorylation sites, studies focused on this PTM have generally applied P-Tyr-specific antibodies (Boersema et al., 2010; Kettenbach and Gerber, 2011; Rush et al., 2005). To date, the precise relationship of Tyr phosphorylation to Ser and Thr phosphorylation in the context of proteome organization is not clear.

Despite the great successes of MS-based workflows in the study of PTMs, many challenges remain (Olsen and Mann,

2013). For instance, the identification of phosphorylation sites, especially on larger peptides, has remained challenging. Here, we address this problem by developing a new computational solution, implemented in the widely used MaxQuant framework and its Andromeda search engine (Cox and Mann, 2008; Cox et al., 2011). Furthermore, knowing the stoichiometry of phosphorylation sites, in addition to their confident identification and localization, would be very helpful in biological interpretation. This has only recently become possible, on the basis of stable-isotope labeling techniques (Olsen et al., 2010; Wu et al., 2011). We developed a label-free approach to quantify the modified peptides and determine their fractional occupancy. With these and other technological advances, we present an analysis of the nature of the phosphoproteome to very great depth by combining protein abundance measurements with phosphorylation changes across mitosis and epidermal growth factor (EGF) stimulation. Our parallel and in-depth investigation of proteome and phosphoproteome resolved to Ser, Thr, and Tyr residues revealed many general aspects of protein phosphorylation as a cellular control process and its potential interaction with other modifications.

RESULTS

Experimental Workflow for Deep Coverage of the Cellular Phosphoproteome

We chose to study phosphorylation in HeLa cells, because they are the most frequently used model systems in cell biology and in previous phosphoproteomics studies. To survey proteome and phosphoproteome at maximum possible depth, HeLa S3 cells were left untreated, mitotically arrested and released, or stimulated with epidermal growth factor (EGF) for 5 or 15 min (Figure 1A). We used a double thymidine block in combination with nocodazole arrest to obtain a mitotic phase population and confirmed synchronization by fluorescence-activated cell sorting (FACS) (Figure 1A, left panel). Successful growth factor stimulation by EGF was apparent from increased phosphorylation of the activation loop Thr/Tyr residues of MAPK1 and MAPK3. For robust statistics, we analyzed biological quadruplicates for each cellular condition and six replicates for the untreated and asynchronous control cells (see Figure 1B and Experimental Procedures for details).

We additionally performed immunoaffinity enrichment of Tyr phosphorylated peptides (Kettenbach and Gerber, 2011) from untreated, mitotic, EGF-stimulated, and pervanadate-treated (for PTP inhibition) cells.

The instrument parameters on the quadrupole Orbitrap instrument used (Q Exactive [Michalski et al., 2011]) were optimized for phosphopeptide identification (Figure S1). The total proteome and phosphoproteome data set comprised 273 liquid chromatography-tandem mass spectrometry (LC-MS/MS) experiments of 4 or 2 hr gradient duration (about 40 days of measuring time, 20 million MS/MS scans).

Computational Pipeline for Reliable Identification, Site Localization, and Absolute Stoichiometry

We included quality thresholds on the spectral matches by applying an Andromeda search engine score filter of 40, and we increased the uniqueness of the peptide spectra match by requiring an Andromeda delta score of 8 to the second best match with different sequence (Supplemental Experimental Procedures; see below and Figure S2A).

The distribution of search engine scores of identified phosphopeptides with a minimum score cutoff of 40 is shown in Figure S2B. A typical spectrum of a phosphopeptide with a score just above the filtering cutoff provides visual evidence that even in such a case there are sufficient peaks for phosphopeptide identification (Figure S2C). This is further supported by additional annotations on low- and high-scoring peptides from a recently developed "expert system" for computer-assisted annotation of MS/MS spectra (Neuhauser et al., 2012) (Figure S2C). The localization score for a site is then defined as the normalized sum of the probabilities for the cases where the site is carrying a modification (Figure 2C; Table S2).

We experimentally evaluated our computational pipeline for PTM identification and localization on results from a peptide library containing more than 100,000 unmodified peptides and their phosphorylated counterparts (Marx et al., 2013) (Figure S2A).

Previously, we had determined site occupancies from stable isotope labeling by amino acids in cell culture (SILAC) quantitation on the level of protein, unmodified peptide, and modified peptide (Olsen et al., 2010). We reasoned that label-free data on multiple conditions might also enable accurate site occupancy calculations. For each site in each condition, we select the most suitable "reference sample" as the one that produces the lowest error in the calculations involved in the occupancy formula ("Proportion" in Figure 2D). We further improved accuracy by using a weighted mean of occupancy calculations for all available reference samples. This occupancy error captures the spread produced by the structure of the formula independent of the statistical errors of the measured quantities. A value less than one indicates that errors are not magnified and therefore the occupancy value is highly reliable.

Phosphopeptides Can Be Identified for Nearly Every Cellular Protein

MaxQuant unambiguously identified 43% of all acquired MS/MS scans that corresponded to 145,340 nonredundant (phospho) peptide sequences originating from 10,801 protein groups (protein entries distinguishable on the basis of identified peptides), when filtering at a false discovery rate of 1% at peptide and protein levels (Table S1). After applying Andromeda score and delta score filters for modified peptides more than 50,000 unique phosphopeptides corresponding to 38,229 phosphorylation events on 7,832 proteins remained, which could be localized with high confidence to single amino acid sequence locations (class I sites, Experimental Procedures, and Figure 3A; Table S2). This number constitutes a lower limit on the size of a human cancer cell line phosphoproteome. Therefore, in contrast to commonly cited estimates that were based on in situ radiolabeling and 2D electrophoresis (Cohen, 2000; Pinna and Ruzzene, 1996), at least three-quarters of the expressed proteome can be phosphorylated. For a given condition such as mitosis, four replicate experiments were sufficient to saturate the number of sites identifiable with the technology used here (Figure S3A). Next, we plotted the number of proteins with at least one

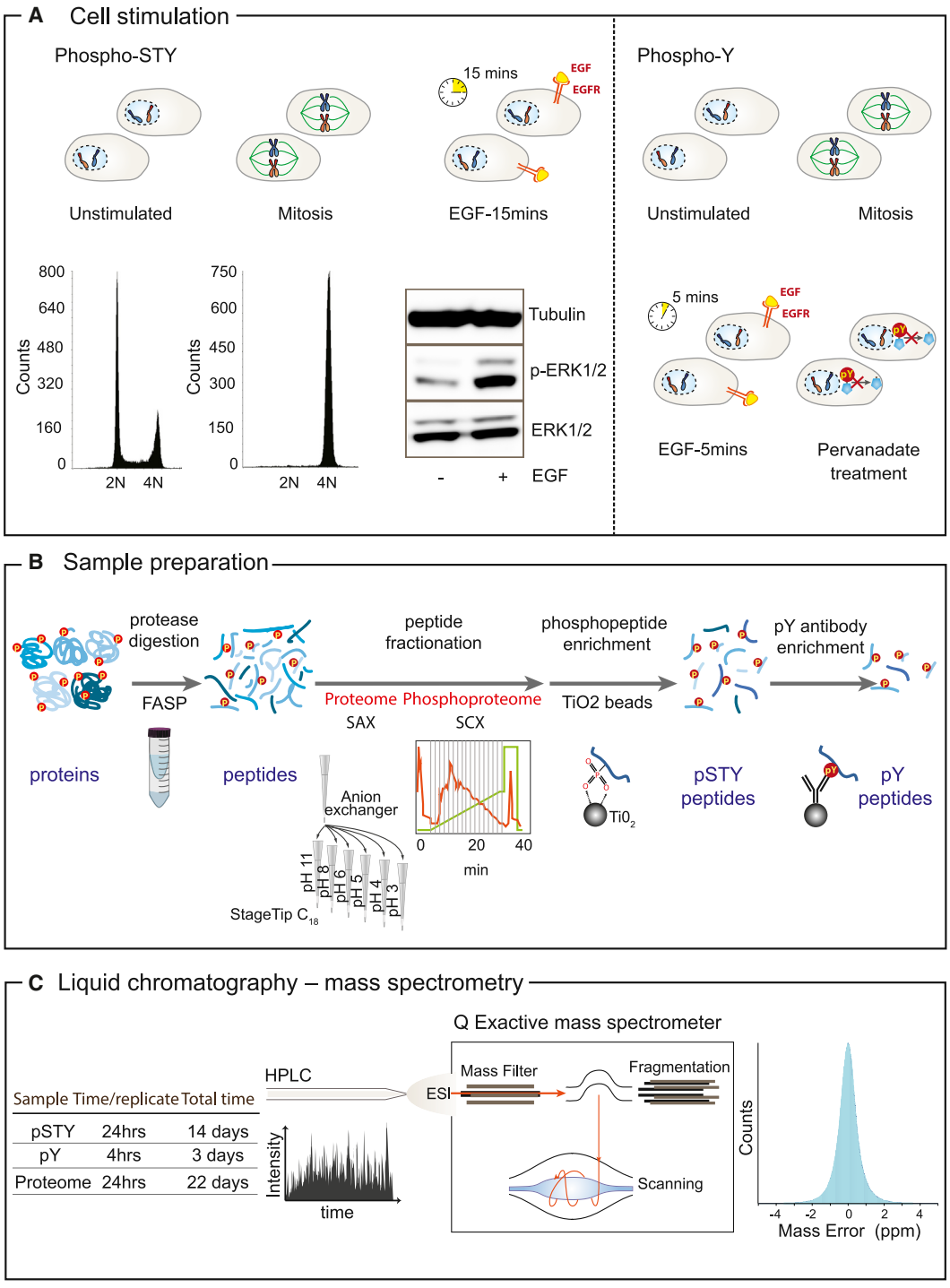

Figure 1. Workflow for Large-Scale Phosphoproteome

(A) Cell stimulation. HeLa S3 cells were synchronized using a double thymidine block and nocodazole arrest and released for 0.5 hr when they were in mitosis. FACS profiles of the synchronized HeLa S3 populations are shown. Two other sets of HeLa S3 cells were treated with EGF for 5 or 15 min. EGF-induced ERK1/2 activation was monitored by immunoblotting. A separate population of cells was treated with sodium pervanadate for in vivo inhibition of Tyr phosphatases.

(B) Sample preparation. Cell lysates from the above treatments were lysed in SDS buffer. For proteome measurements peptides were separated into six fractions, and, for total phosphoproteome analysis, phosphopeptides were enriched by strong cation exchange (SCX chromatography) and TiO_2 microbeads. P-Tyr peptides were immunoprecipitated from phosphopeptides using anti-P-Tyr antibodies.

(C) LC-MS analysis: all fractions were separated on a reverse-phase column and electrosprayed into a quadrupole-Orbitrap mass spectrometer, which was operated in a data-dependent mode, and produced ppm range mass accuracies for precursors and fragments.

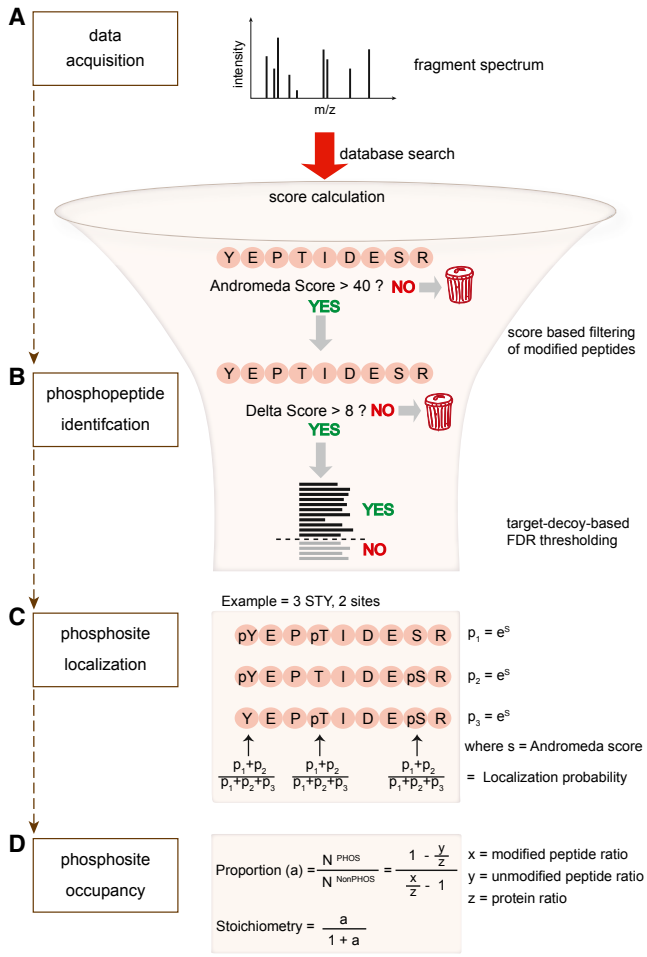

Figure 2. Computational Pipeline for High-Stringency Identification and Quantification of Phosphopeptides in MaxQuant

(A) Data acquisition with high resolution for precursors and fragments.

(B) Phosphopeptide identification. Only spectra matched to sequences with an Andromeda score >40 and a delta score of >8 were retained. Posterior error probabilities (PEP) were calculated based on the target decoy strategy to control the false discovery rate.

(C) Phosphosite localization. Localization of the modification site was achieved by looping through possible combinations for the phosphorylation on individual amino acid residues on the peptide for which the Andromeda score is calculated and exponentiated to obtain the localization probability.

(D) Phosphosite occupancy. The proportion of phosphorylated peptides was calculated based on the extracted signal differences of modified peptide, unmodified peptides, and corresponding protein ratios between biological states.

detected phosphorylation event as a function of the size of the set of intensity ranked phosphopeptides (Figure 3A, left panel). The curve quickly covers half of the detected proteome and then rises more slowly as more and more low-intensity phosphopeptides are added. The number of phosphorylation events found in our earlier studies at a depth of 6,600 (Olsen et al., 2006) and 20,443 phosphosites (Olsen et al., 2010) coincide with this curve, suggesting a general relationship. Clearly, the curve has not reached saturation, suggesting that adding further perturbations would add substantially to the phosphoproteome. In particular, we predict that at a depth of 100,000 sites more than 90% of the expressed proteome would have detectable phosphorylation sites (Figure 3A, left panel).

Because of the recent identification of Fam20 as a novel protein kinase that phosphorylates SXpS/E motifs in secreted proteins, there is much interest in secreted phosphoproteins. To illustrate the value of our data set, we selected proteins with phosphosites with SXpS/E motifs that disfavor proline at +1 position. To capture putative Fam20 substrates, we further filtered for secreted, extracellular matrix or Golgi apparatus proteins resulting in a set of 101 phosphosites on 64 proteins (Table S2). Likewise, our data set also constitutes a catalog of regulatory protein expression and phosphorylation events associated with mitotic and growth factor signaling. Tables with all identified proteins, phosphopeptides and phosphosites are provided (Tables S1 and S2 and data on proteome exchange). In addition, all the identified phosphosites, phosphopeptides, and proteins, including their high-resolution fragment spectra, are easily accessible through the MaxQB database (Schaab et al., 2012) via a user-friendly web interface at http://maxqb.biochem.mpg.de/mxdb/project/show/P007 (see Supplemental Experimental Procedures) (Figure S3B).

We found that more than 22,000 of the 38,000 phosphosites (55%) were identified in all the biological conditions and 20% were found exclusively in mitosis (Figure 3A, right panel). The preponderance of exclusive sites at this phase of the cell cycle may reflect the complexity of associated regulatory events (Table S2).

In signaling studies selected, functionally important sites are usually followed using site-specific antibodies. For comparison to our data set, we chose the sites for which antibodies exist in the PhosphositePlus database (Hornbeck et al., 2012) (Figure 3B, left panel). Our data set covered more than 50% of them, implying that it has reached considerable depth, especially considering that our analysis was performed with only three perturbations and in a single cell system, whereas antibodies have been generated for very diverse signaling contexts. Interestingly, the sites usually followed by antibodies are enriched on more abundant proteins with median intensity about half an order of magnitude higher than those of all identified sites (Figure 3B, middle panel). It follows that either functional phosphorylation events occur on more abundant proteins or that biological functions were historically first ascribed to phosphorylation of more abundant proteins. Intensity-ranked signal intensities of these 282 phosphosites span the entire dynamic range of MS analysis, but 90% of them are present within four orders of magnitude. By extension, the dynamic range achieved in our phosphoproteome analysis appears sufficient to sample most of the currently investigated cellular phosphoproteome.

Label-free Quantification Enables Precise Identification of Key Pathways across Biological States

We investigated if the label-free algorithms of MaxQuant could accurately capture signaling dynamics, similarly to SILAC (Mondetti et al., 2011; Olsen et al., 2006). Label-free method are particularly challenging in the analysis of PTMs because each peptide needs to be quantified by itself, unlike in proteome quantification, where several peptides generally contribute to protein quantification.

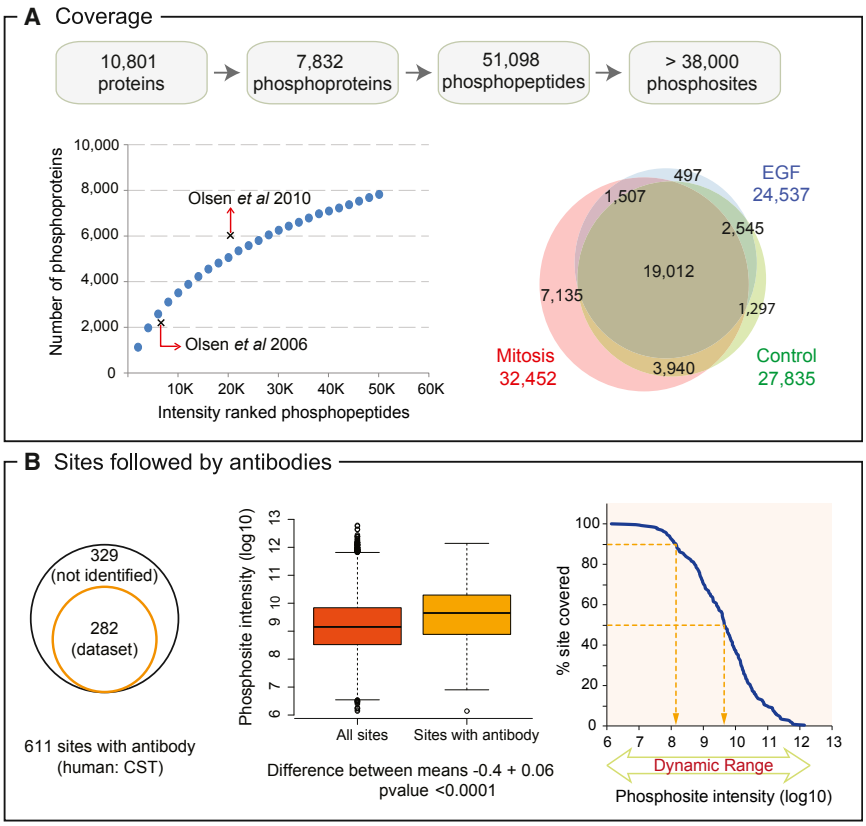

Figure 3. Overview of the Identified Phosphoproteome

(A) Detection of phosphoproteins as a function of the depth of phosphopeptides. Large-scale MS-based analysis resulted in the identification of 10,801 proteins of which 7,832 were phosphoproteins. A plot of the number of proteins that were detected to be phosphorylated as a function of phosphoproteome depth from this study with phosphorylated proportion identified in two previous studies are marked in the left panel (Olsen et al., 2006, 2010). A total of 51,098 phosphopeptides were identified of which >38,000 phosphorylation events were localized to specific S/T/Y residues. The Venn diagram depicts overlap among high-confidence phosphosites identified across the conditions studied (right panel).

(B) Sites with corresponding antibodies. Overview of sites for which sequence-specific antibodies are available identified in our data set (left panel). Distribution of intensities of all phosphosites identified in comparison to those with antibodies (middle panel). Coverage of intensity ranked phosphosites (followed by antibodies) plotted as a function of their abundance.

We found that the key phosphosites that are known to be activated by epidermal growth factor (EGF) ere quantified accurately, as illustrated by the activating phosphorylations on Tyr 1197 and Tyr 1172 on epidermal growth factor receptor (EGFR) that are upstream of RAS/RAF/MAPK signaling (Figure 4A, upper panels and Table S2). Phosphorylation of these sites increased significantly in cells treated with either EGF or pervanadate (50-fold or 5-fold, respectively; $p < 0.005$). Similarly, we observed a decrease in inhibitory phosphorylation on Thr14/Tyr15 in cyclin dependent kinase 1/2/3 and increase in activating phosphorylation of polo like kinase-1 (PLK1) only in the mitotic samples (Figure 4A, upper panels). The biological replicate measurements showed high quantitative reproducibility with a Pearson's correlation coefficient across the entire data set of more than 38,000 sites of at least 0.82 (Figure S4A).

We obtained cluster-specific footprints of kinase activation using a Fisher's exact test for kinase-substrate motifs (Table S2; heatmap in Figure 4A). In the mitotic cluster, the motifs for proline-directed kinase CDK1 and polo box domain showed the strongest enrichment (Figure 4B). The GO and KEGG categories characteristic of mitosis were strongly enriched (Figure 4A). Remarkably, comparative enrichment analysis (Cox and Mann, 2012) of entire phosphoproteome of mitotic cells versus asynchronous cells showed that all of the most highly enriched and most significant gene ontology terms ($p < 10^{-20}$) directly related to the cell cycle (cell-cycle phase, cell-cycle process G2 phase/mitotic cell cycle and establishment of chromosome localization; Figure 4A, lower panel, inset).

In summary, the accurate quantification across thousands of sites and enrichment of relevant pathways demonstrate that label-free workflows and algorithms can enable quantitative analysis of phosphoproteomes to great depth, yielding functional portraits of cellular signaling states. For each protein and phosphosite of interest, levels estimated by label-free quantification as well as their statistical significance can readily be visualized across the different biological states in the MaxQB database (Figure S3B).

Specific Signaling States Are Characterized by High Fractional Site Occupancies in Disordered Regions

Next, we extracted absolute, site-specific stoichiometries from label-free abundances of proteins, phosphorylated peptides, and their unmodified counterparts for more than 16,000 sites. We assigned fractional occupancy to 7,620 phosphorylation sites with high confidence (Table S2). In accordance with our SILAC-based analysis of the cell cycle (Olsen et al., 2010), we found that the mitotic phase of the cell cycle has a very large number of Ser/Thr phosphorylation events with high site occupancy (Figure 4B, upper panel). For half of all sites, we determined a fractional occupancy of at least 75%. In contrast, the vast majority of Ser/Thr phosphosites in control cells or EGF-stimulated cells have less than 25% fractional occupancy (Figures 4B and S4B). This raised the intriguing possibility that the high-occupancy mitotic sites might be detectable even without any phosphoenrichment. Indeed, we were able to identify more than 2,000 phosphosites in the total proteome measurements of mitotic cells in stark contrast to only a few tens of phosphosites in control or EGF-stimulated proteome measurements.

We next analyzed structural properties of identified phosphorylation sites in relation to their occupancies across different

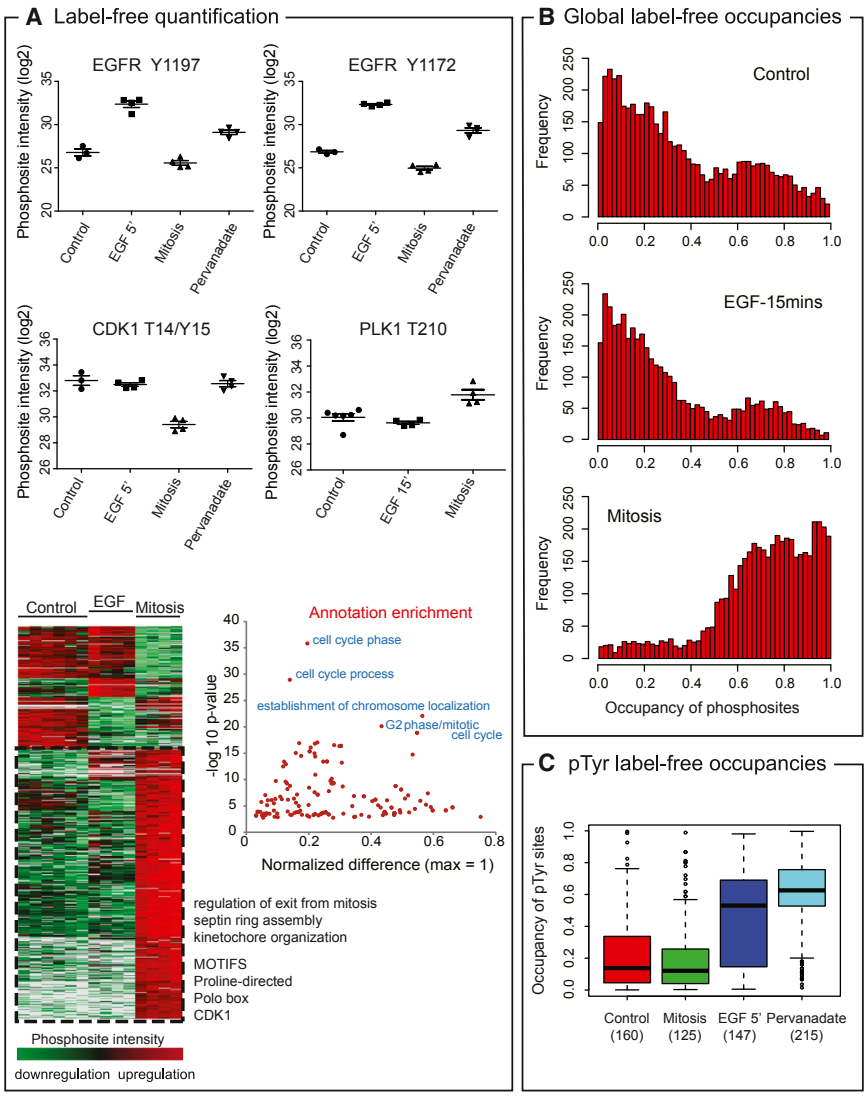

Figure 4. Relative and Absolute Label-free Quantification

(A) Label-free quantification of regulated phosphopeptides. Label-free quantification of individual replicates (mean ± SD of replicates) for the conditions studied is shown for Y1197 and Y1172 on EGFR, Y15/Y16 on CDK1/2/3, and T210 on PLK1 (upper panel). Heatmap of phosphosites regulated across different conditions (color scale from green to red indicating decreased and increased phosphorylation) with kinase motifs and categories enriched (lower panel). Inset in the lower panel shows 1D annotation analysis of annotation terms of proteins significantly upregulated during mitosis. The data points corresponding to annotation terms whose members are regulated with very high significance are labeled ($p < 10^{-20}$).

(B) Label-free site occupancies: frequency distribution of phosphosite occupancies for control, EGF-treated, and mitotic samples.

(C) Label-free site occupancies of P-Tyr sites: distribution of P-Tyr occupancies across different biological conditions.

signaling states. Disorder state was predicted using the DisoPred software (Ward et al., 2004) for all proteins, and the corresponding state was mapped to each phosphorylation site in our data set (Table S2). The majority of the sites (86%) were found in disordered regions, as expected (Iakoucheva et al., 2004). Without stimulation there was very little difference between ordered and disordered sites as judged by 2D enrichment analysis (Cox and Mann, 2012), suggesting that protein phosphorylation tends to occur independent of structure in unstimulated cells. In mitosis or EGF treatment, in contrast, the occupancies for ordered and disordered regions were significantly different ($p < 10^{-6}$). High fractional occupancy sites preferentially occurred on disordered regions in these states (Table S2), presumably due to activation of upstream of kinase activity favoring disordered regions (Tyanova et al., 2013).

In contrast to Ser/Thr site, absolute occupancy of Tyr sites has not yet been reported from global measurements. We were able to reliably estimate occupancies for 312 Tyr-specific phosphorylation events in at least one cellular condition (Table S2). Most mitotic and control phosphosites have less than 10% fractional occupancy, whereas half of those from EGF-treated samples show fractional occupancy of 50% or higher (Figure 4C). For pervanadate-treated cells, three-quarters of Tyr phosphorylation events have a fractional occupancy greater than 50%.

Our results show generally low fractional occupancy in untreated cells and that Tyr phosphorylation is maintained at especially low stoichiometric levels in the absence of specific signaling events (Figure 4C). Together, the above results show that site occupancies correlate with cellular signaling state.

An 80-20 Rule for Protein Phosphorylation

The distribution of the MS-signals of the identified phosphopeptides are a function of protein abundance and site stoichiometry and therefore can be used as a proxy for the amount of ATP transferred as phosphate moieties onto proteins in the form of Ser, Thr, and Tyr modifications. The MS signals of the phosphopeptides span six orders of magnitude, demonstrating the sensitivity with which MS-based proteomics can now capture them. However, 70% of the quantified phosphopeptides are contained within only one order of magnitude above or below the median phosphopeptide abundance (Figure 5A, left panel). The 150 most abundant phosphopeptides already composed 20% of total phospho, whereas the most abundant 4,963 phosphopeptides constituted 80% of the cumulative phosphopeptide signal (Figure 5A, right panel). The remaining 89% of the

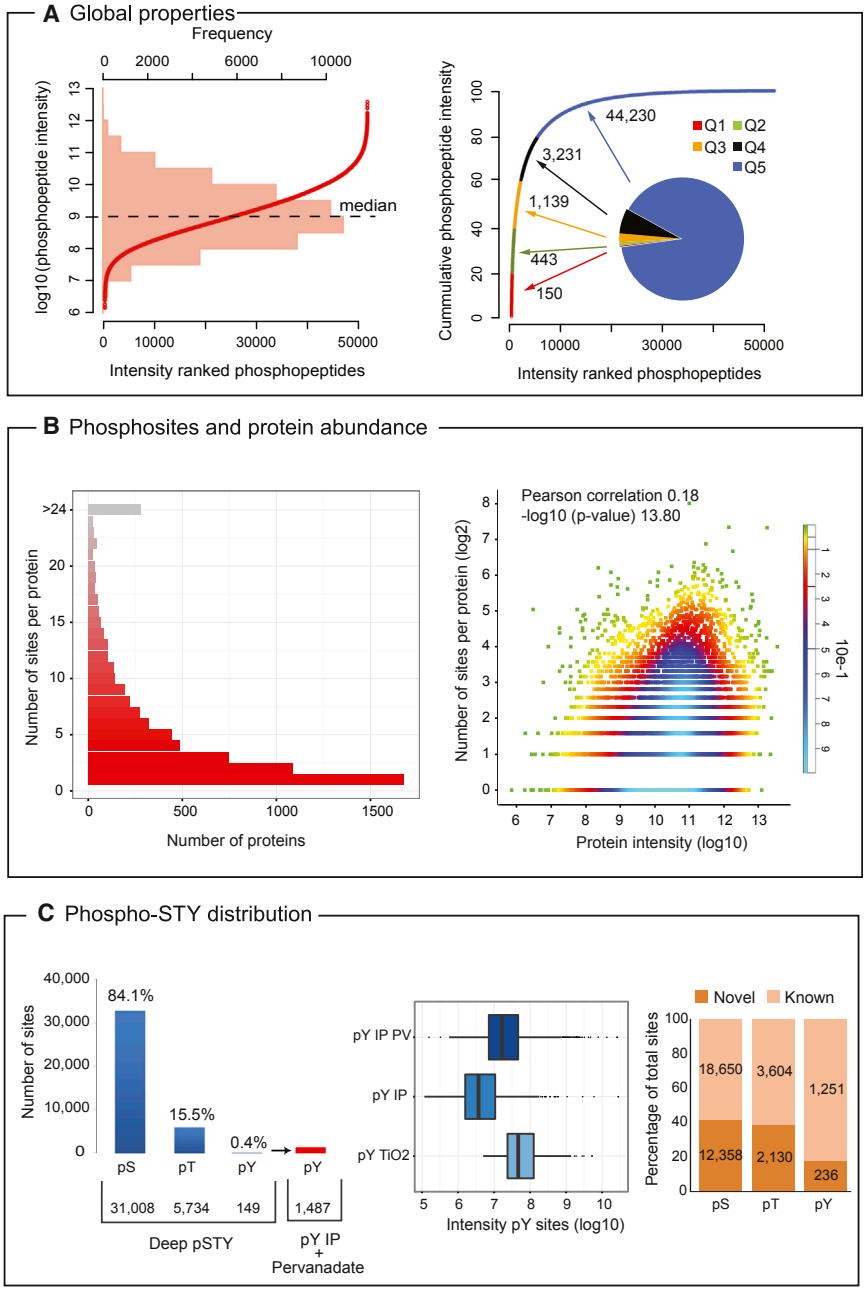

Figure 5. Properties of the Phosphoproteome

(A) Dynamic range of the phosphoproteome. Histogram of phosphopeptide intensities showing median intensity on which ranked phosphopeptide abundances from decreasing to increasing abundance are overlaid (left panel). Cumulative phosphopeptide abundance from the highest to the lowest abundance with pie chart separating the abundances into five intensity quantiles (right panel).

(B) Overview of phosphorylation sites per protein. Distribution of phosphoproteins based on number of phosphorylation sites per protein (left panel) and density scatterplot of protein abundance versus number of sites per protein. The color code indicates the percentage of points that are included in a region of a specific color (right panel).

(C) Phosphosite distribution across S/T/Y residues. Distribution of the S/T/Y phosphorylation events by global phosphoproteomics and P-Tyr immune precipitation (IP) (left panel). Box plots of P-Tyr peptide intensities from global (TiO_2-based), P-Tyr IP, and IP^+ pervanadate-treated samples (middle panel). Distribution of the known and novel P-Tyr/P-Ser/P-Thr sites after matching with PhosphositePlus database (right panel).

quantified phosphopeptides, including both the low- and high-occupancy phosphorylation events, correspond to only 20% of total cellular ATP transferred as phosphate groups. This suggests that a power law and specifically a Pareto Principle, or 80-20 rule, holds true for cellular phosphorylation, and it reveals that the vast majority of phosphorylation events together consume less than 20% of cellular ATP consumed in protein phosphorylation.

Our parallel measurement of more than 50,000 phosphopeptides and about 11,000 proteins allowed us to investigate a possible correlation between a protein's abundance and its propensity to be phosphorylated. We observed a weak but highly significant tendency for a greater number of identified phosphosites with increasing protein abundance (Figure 5B, Pearson correlation 0.18, $p < 10^{-13}$). About 15% of identified proteins were phosphorylated on just one residue, whereas the remaining 85% were phosphorylated at multiple sites (Figure 5B). Remarkably, half of the phosphorylated proteins had six or more detected phosphorylation sites. These multiple phosphorylation events may either be functional—for instance, reflecting cross-talk downstream of multiple signaling pathways—or they may reflect background phosphorylation due to low-level kinase activity or both.

Because we observed that many proteins could potentially be multiply phosphorylated, we next asked how frequently two sites on the same protein molecule are phosphorylated simultaneously. This can be inferred indirectly from our data set. First, if we detect two or more sites on a multiply phosphorylated peptide, they must be simultaneously phosphorylated (indicated by multiplicity greater than 1 in Table S2). Second, we used occupancies data to identify a minimum set of coexisting phosphosites on a protein. Assuming that a protein with two modification sites, both of which are more than 50% occupied, the two sites cannot be completely exclusive. This assumption can be applied to any binary combination of modified sites on a protein. Applying this logic, we collated a list of sites that are more than 50% occupied during mitosis and should coexist (Table S2).

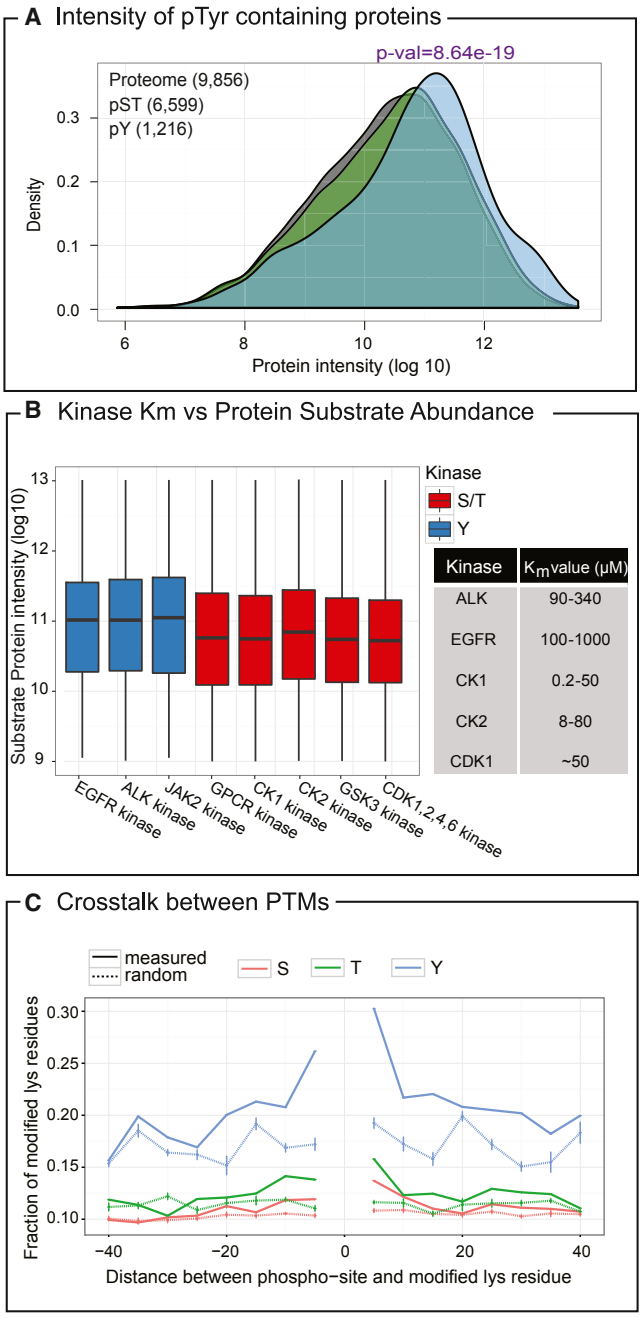

Figure 6. Comparison of Phosphorylation on S/T versus Y Residues

(A) Density distribution of intensities of all proteins identified (black) compared to proteins with Ser/Thr phosphorylation (green) and Tyr phosphorylation (blue).

(B) Violin plot distributions of protein intensities of substrates phosphorylated by the indicated kinases. Tyr kinases substrates are depicted in blue and those for S/T kinases in red. A table of kinases and their K_M values (Donella-Deana et al., 2005; Fan et al., 2005; Sarno et al., 1996; Ubersax et al., 2003).

(C) Distribution of modified lysine residues over the protein sequence (intervals [−10;−5], [−5;0], [0;5], [5;10]) compared to random occurrences (dashed lines) plotted separately for S, T, and Y.

In Vivo Inhibition of Tyr Phosphatases and Immunoenrichment Enables Comprehensive Coverage of P-Tyr

The residue-specific phosphorylation pattern found here (Ser 84.1%, Thr 15.5%, and Tyr 0.4%; Figure 5C) closely corresponds to classical radioisotope based estimates (Hunter and Sefton, 1980). However, successive quintiles of phosphopeptide abundance contained proportionally fewer detected P-Tyr sites, so that the overall the P-Tyr proportion was much lower than that in a previous phosphoproteome investigation in which we identified about 6,600 sites (Olsen et al., 2006).

Next, to ensure maximal representation of Tyr phosphorylation events, we immunoenriched Tyr phosphorylated peptides (Kettenbach and Gerber, 2011). Apart from the three cellular states (EGF stimulation, mitotis, and untreated control cells), we also enriched pTyr peptides from pervanadate-treated cells. Inhibition of Tyr phosphatases by pervanadate increased the number of identified P-Tyr residues and also increased their overall MS signal by an order of magnitude (Figure 5C, middle panel). Together, this strategy allowed us to cover 10-fold more P-Tyr residues through identification of more than 2,000 Tyr phosphorylated peptides on about 1,300 proteins. Comparison between TiO_2 enrichment and anti-pTyr antibody enrichment across three conditions (control, mitosis, and EGF stimulation) revealed that two-thirds of the sites identified using TiO2 enrichment (149 sites) were identified by the pTyr antibodies. In contrast, only 11% of the sites identified by anti-pTyr antibodies (839 sites) were identified using the TiO_2 enrichment method. The P-Tyr sites were found on key proteins across multiple signaling pathways, suggesting a global representation of these events in our data set (Table S3). Although 60% of the identified Ser and Thr phosphorylation events were novel, as compared to the PhosphositePlus database (Hornbeck et al., 2012), this was true for only 18% of the P-Tyr residues (Figure 5C, right panel). We conclude that, as depth of coverage increases, detectable Tyr sites are covered relatively rapidly, whereas the increase in Ser/Thr sites only appears to saturate for technical reasons.

Tyrosine Phosphorylation Occurs on Higher Abundance Proteins

Our data set revealed that proteins carrying P-Tyr residues were significantly more abundant ($p < 10^{-16}$), whereas proteins with phosphorylated Thr or Ser residues were similar in abundance to all identified proteins (Figure 6A). When we analyzed the data by kinase motifs, again the Tyr kinase substrates were more abundant than those of Ser/Thr kinases (Figure 6B).

The higher abundance of tyrosine phosphorylated proteins is unlikely to be a consequence of poor detection of low abundant P-Tyr peptides because the P-Tyr peptides found after in vivo inhibition of Tyr phosphatases (Figure 5C) were also present on more abundant proteins. Further, if our data were technically biased toward detecting only highly abundant P-Tyr phosphopeptides, we would tend to identify Tyr phosphosites with high fractional occupancies compared to Ser/Thr phosphosites. However, we did not find evidence for such a bias. Instead, the occupancy of Tyr-specific phosphosites was rather determined by the signaling state of the cell (Figures 6A and S4B).

We reasoned that possible explanation may be found in the K_M values, an important factor in determining the activity of a kinase toward its substrates. To ensure efficient phosphorylation, substrates of a kinase should be present at concentrations above their K_M values. Indeed, we found that abundance of the P-Tyr and P-Ser/Thr proteins correlates well with the relatively high K_M values of the Tyr kinases compared to those of Ser/Thr kinases (Figure 6B, table within figure). Differences in median site occupancy for Tyr and Ser/Thr phosphorylation events in untreated cells can likewise be explained by the K_M values for Ser/Thr compared to those of Tyr kinase (Figure S4B).

Phosphorylated Residues Are Preferentially Located near Modifiable Lysines

Crosstalk between different types of posttranslational modification is an emerging theme in signaling biology (Hunter, 2007). To investigate our data set for evidence of relations between phosphorylated residues and lysine modifications, we aligned our sites with those of lysines that have been documented as modified by acetylation, ubiquitination or sumoylation (see Experimental Procedures). If there was crosstalk between phosphorylation and these lysine modifications, we would expect a nonrandom distance distribution between them, and this was indeed the case (Figure 6C). At short distances from the phosphorylation sites, the measured fractions (solid lines) lie above the randomized data (dashed lines) demonstrating a tendency of modified lysine residues to occur with a higher preference in the surroundings of phosphorylated residues. Furthermore, the fraction of modified to nonmodified lysine residues decreased with the distance to the phosphorylation site. Interestingly, this effect was by far strongest for P-Tyr residues, where the fold difference between the measured and the randomized data was 1.57, than for either P-Ser or P-Thr residues (1.26 and 1.36, respectively). This striking finding is in agreement with the notion that our measured P-Tyr are primarily regulatory and stringently controlled whereas the proportion of functional P-Ser and P-Thr sites may be much lower. We found distances between Ser/Thr/Tyr and lysine residues to be significantly shorter ($p < 10^{-15}$) on average in phosphopeptides than in unphosphorylated ones. Thus, we confirmed that our observation is not biased even if it is observed on tryptic phosphopeptides, which tend to have modification sites in the vicinity of their C-terminal amino acid. In any case, the highly nonrandom distribution of measured phosphosites in relation to reported lysine modification sites provides strong evidence for the notion of multiple modifications acting in concert to regulate cellular processes.

DISCUSSION

Comprehensive study of PTMs is more challenging than proteome analysis because of the required enrichment steps, low abundance of modified peptides, and more complex LC-MS/MS and computational analysis.

Here, we have developed an experimental and computation pipeline to probe the phosphoproteome of a human cancer cell line in unprecedented depth. We demonstrated accurate identification of phosphopeptides and validated our strategy on a large synthetic phosphopeptide library. We also achieved efficient relative quantification and determination of absolute site stoichiometry from label-free data. These latter advances are incorporated into MaxQuant and are thus freely available to the community for the analysis of phosphorylation or other PTMs. Likewise, our data set of more than 50,000 distinct phosphopeptides, including quantitative and contextual information is readily accessible through the MaxQB database (http://maxqb.biochem.mpg.de/mxdb/project).

We find that at least three-fourths of the detected proteome (7,832 out of 10,801 proteins) can be phosphorylated. Combining protein abundance measurements with phosphorylation changes across mitosis and epidermal growth factor (EGF) stimulation allowed determination of the occupancy of thousands of phosphorylation sites using a label-free quantification approach. This revealed that low-occupancy phosphorylation sites prevail in unstimulated cells. In contrast, specific signaling states often entail high site occupancies, implying that nearly all of the relevant signaling molecules in the cell carry the modification. More than 80% of cellular ATP involved in protein phosphorylation is concentrated on less than 20% of phosphorylation site, thus phosphorylation appears to follow a Pareto Rule. Conversely, the large majority of phosphorylation events together consume only a minor fraction of cellular ATP. Thus, our data provide direct experimental evidence that these events are not energy expensive to the cell and support the notion that controlling these events may be cost inefficient (Levy et al., 2012). Another relevant factor in this regard are the rates by which the phosphates may turn over on these sites (see, for example, Kleiman et al., 2011). However, such a high turnover rate is associated with a limited subset of regulated phosphorylation events (Kleiman et al., 2011). Therefore, our observation based on more than 50,000 phosphopeptides should be a reliable estimate of cellular ATP utilization in protein phosphorylation.

In vivo inhibition of Tyr-phosphatases followed by antibody-based enrichment of phosphopeptides allowed recording of tyrosine phosphorylation sites in a broad range of pathways rather than mainly adding nonfunctional sites. Comparison of our data to already known phosphorylation events (Hornbeck et al., 2012) suggests that coverage of Tyr phosphorylation is very comprehensive already. In contrast, modeling of the number of phosphorylated proteins as a function of Ser/Thr phosphorylated peptides demonstrates that the Ser/Thr phosphoproteome is far from complete, and saturation in a given condition is rather due to technical factors. Tyr phosphorylated residues were preferentially found closer to modifiable lysines than Ser/Thr phosphorylated residues, implying a preferential role in PTM crosstalk. Beltrao et al. previously hypothesize that PTM crosstalk can be used as a feature to assign functional relevance to modifications (Beltrao et al., 2012, 2013; Swaney et al., 2013), and Swaney et al. (2013) implicated co-occurrence of phosphorylated residues with ubiquitinylated in the regulation of protein degradation.

Despite the large number of Tyr kinases in the genome, P-Tyr residues account for <1% of total phosphorylation events. This and many other lines of evidence suggest that specificity in Tyr phosphorylation is strictly maintained in cellular systems (Hunter, 2009). Tyrosine phosphatases are thought to have come into

existence before the corresponding kinases (Lim and Pawson, 2010). Here, we directly determined fractional occupancy of P-Tyr sites in cells and showed that the observed low fractional occupancy of P-Tyr sites in unstimulated cells indeed correlates with cellular control mechanisms determining specificity. The global inhibition of Tyr phosphatases by pervanadate should lead to a loss in specificity of Tyr phosphorylation events and does not reflect a particular biological state. Nevertheless, the observed phosphorylated pool is a collection of Tyr residues that remain phosphorylated in the absence of phosphatase action and represents in vivo substrates of Tyr kinases. Although it was expected that we will identify a large number of such tyrosine phosphorylation events (as reported in Rush et al., 2005 and Kettenbach and Gerber, 2011), their occupancies was a matter of speculation. In this study, we unexpectedly found that such events tend to have high site occupancy.

The extent of site-specific phosphorylations in vivo depends on various factors such as the local concentrations of the kinase, the substrate, antagonizing phosphatase(s), protein-protein interactions as well as other potential substrates "competing" for the same kinase (Ubersax and Ferrell, 2007). An example of multiple mechanisms contributing to kinase selectivity can be found in case of cell-cycle control by cyclins, which include a difference in kinase K_M toward substrates and interaction with specific motif on substrates (Loog and Morgan, 2005). Interestingly, we found that proteins phosphorylated on Tyr residues are on average more abundant compared to the entire proteome. The observed difference in phosphoproteins abundance correlates with the substrate K_M values of Tyr kinases. K_M differences have important consequences in the cell, where the protein kinases are exposed to varying substrate concentrations. We speculate that high abundance of substrate proteins coupled with relatively lower efficiency of Tyr kinases could buffer against harmful effects of the occasional stray phosphorylation of functionally important sites. In contrast to Ser/Thr kinases, the Tyr kinases might not be significantly inhibited by competition from general substrates because of their high K_M values for general substrates, allowing them to efficiently phosphorylate a subset of low K_M substrates. Consequently, the low activity of Tyr kinases toward general targets, combined with high abundance of their specific targets, contributes to specificity in Tyr-based signaling events. The information content in many tyrosine kinase recognition motifs can be low, and therefore other control mechanisms for specificity are required. A recent study revealed that an ensemble of fine-tuned weak interactions control cellular decisions as exemplified by cell-fate control by RTKs (Findlay et al., 2013). Therefore, in order for multiple, coincident ligand interactions to appropriately specify stem cell differentiation, individual ligand affinities must be selected to remain below a certain threshold. Here, we observed fine-tuned substrate protein abundances correlating with weak tyrosine kinase activity. We propose that the tailored protein abundances of Tyr kinase substrates coupled with low substrate affinities provide yet another mechanism controlling specificity to these signaling systems. In summary, our findings highlight the nature of P-Tyr as a separate functional regulatory posttranslational modification of eukaryotic proteomes.

EXPERIMENTAL PROCEDURES

Cell Culture and Treatment

HeLa S3 cells were cultured in RPMI 1640 (Gibco) supplemented with 10% fetal bovine serum (Invitrogen) and 1% penicillin/streptomycin (Invitrogen). A double thymidine block in combination with nocodazole was used to obtain a homogenous mitotic phase population and the efficiency of mitotic arrest was monitored by FACS analysis. Cells were washed and suspended in PBS and treated with 100 ng/ml EGF for 5 or 15 min. For tyrosine inhibition of phosphatases, cells were treated with 1 mM pervanadate and 50 ng/ml calyculin A for 15 min at 37°C as described earlier (Rush et al., 2005).

Total Proteome and Phosphoproteome Sample Preparation and MS Analyses

A total of 12.5 mg of protein lysate per experiment was digested by the FASP method (Wiśniewski et al., 2009b). For proteome analysis, 30 μg of the peptides was separated on a pipette-tip based SAX column (Wiśniewski et al., 2009a). For phosphopeptide enrichment, 6 mg peptides obtained from FASP was fractionated by strong cation exchange (SCX) chromatography (Olsen et al., 2006) and subjected to phosphopeptide enrichment using TiO_2 beads (Zhou et al., 2011). Briefly, peptides were dissolved in 80% acetonitrile (ACN) and 6% triflouroacetic acid (TFA) and incubated with TiO_2 beads (1:5 peptides to bead ratio) for 20 min. The beads were washed with 80% ACN and 0.1% TFA and phosphopeptide elution was carried out under basic pH using ammonia. For P-Tyr enrichment, 8 mg of digested peptides were subjected to TiO_2 enrichment before immunoprecipitation with a mix of anti-P-Tyr antibodies as described (Kettenbach and Gerber, 2011). The (phospho)peptides were desalted on C_{18} StageTips (Rappsilber et al., 2003).

Reverse-Phase Chromatography and Mass Spectrometry

Peptides were separated on a 50 cm reversed phase column (75 μm inner diameter, packed in-house with ReproSil-Pur C18-AQ 1.9 μm resin [Dr. Maisch GmbH]) over a 120 or 240 min gradient of 5%–60% buffer B (0.1% [v/v] formic acid, 80% [v/v] acetonitrile) using the Proxeon Ultra EASY-nLC system. The LC system was directly coupled online with a Q Exactive instrument (Thermo Fisher Scientific) via a nano-electrospray source. The mass spectrometer was programmed to acquire in a data-dependent mode using a fixed ion injection time strategy (Kelstrup et al., 2012). Full scans were acquired in the Orbitrap mass analyzer with resolution 70,000 at 200 m/z. For the full scans, 3E6 ions were accumulated within a maximum injection time of 20 ms and detected in the Orbitrap analyzer. The ten most intense ions with charge states ≥ 2 were sequentially isolated to a target value of 1e6 with a maximum injection time of 60 or 80 ms and fragmented by HCD (Nagaraj et al., 2010) in the collision cell (normalized collision energy of 25%) and detected in the Orbitrap analyzer at 17,500 resolution.

Data Processing and Analysis

Raw mass spectrometric data were analyzed in the MaxQuant environment (Cox and Mann, 2008) v.1.5.0.0 and employed Andromeda for database search (Cox et al., 2011). The MS/MS spectra were matched against the human Uniprot FASTA database v.2/25/2012 (81,213 entries). Enzyme specificity was set to trypsin, and the search included cysteine carbamidomethylation as a fixed modification and N-acetylation of protein, oxidation of methionine, and/or phosphorylation of Ser, Thr, Tyr residue (STY) as variable modifications. Up to two missed cleavages were allowed for protease digestion, and peptides had to be fully tryptic. For detailed explanation of PTM analysis in MaxQuant, see Supplemental Experimental Procedures.

Downstream Bioinformatics Analyses

Bioinformatic analysis was done in the Perseus software environment, which is part of MaxQuant. Hierarchical clustering of proteins or phosphosites was performed on logarithmized intensities. For ANOVA analysis, replicates were grouped, and the statistical test was performed with a permutation-based false discovery rate (FDR) cutoff of 0.01.

Categorical annotation was supplied in the form of GO biological process, molecular function, and cellular component, KEGG pathways for pathway annotation and human protein reference database (HPRD) for kinase substrate

motifs. Enrichment for these categories was evaluated by Fisher's exact test to obtain p values. The annotation matrix algorithm was used to compute the difference of any significant protein annotation term from the overall intensity distribution as described (Cox and Mann, 2012). We used a 2D version of the nonparametric Mann-Whitney test. Multiple hypothesis testing was controlled with a Benjamini-Hochberg FDR threshold of 0.05.

Determination of Crosstalk

Ubiquitination, acetylation, and sumoylation data sets were obtained from the public repository PhosphoSitePlus (Hornbeck et al., 2012). The modified lysine residues were mapped to the phosphorylation data set. The fraction of modified to nonmodified lysine residues at the flanking regions of each phosphosite was computed. Flanking regions of different lengths (from 5 to 40 amino acids in intervals such as [−10; −5], [−5;0][0;5], [5;10]) were analyzed. Next, the positions of the modified lysine residues were randomized over all lysine residues in the corresponding proteins, and the resulting fractions of modified to nonmodified lysines were computed. The randomization was repeated 1,000 times, creating a background distribution of random distances. The measured and randomized fractions were plotted for each P- Ser, P- Thr, and P-Tyr residue.

ACCESSION NUMBERS

The MS-based proteomics data were deposited at the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository with the data set identifier PXD000612. The data set can also be accessed through the MaxQB database (http://maxqb.biochem.mpg.de/mxdb/project/show/P007).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, four figures, and three tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2014.07.036.

AUTHOR CONTRIBUTIONS

K.S. and M.M. conceived the study. K.S. and R.C.J.D. did the experiments and developed the phosphoproteomics workflow. J.C. developed the MaxQuant software modules for label-free PTM analysis and occupancy calculation. K.S., R.C.J.D., and S.T. implemented the bioinformatics analysis and interpreted the data. J.R.W. intellectually contributed toward data interpretation. C.S. extended MaxQB for PTM data visualization. K.S., R.C.J.D., S.T., J.C., and M.M. wrote the manuscript.

ACKNOWLEDGMENTS

We thank Paul Boersema, Sean Humphrey, Nadin Neuhauser, and Chunaram Choudhary for helpful discussions and Martin Dodel, Igor Paron, and Korbinian Mayr for technical assistance. This work was partially supported by European Union 7th Framework project PROSPECTS (Proteomics Specification in Time and Space, grant HEALTH-F4-2008-201645).

Received: January 8, 2014
Revised: May 30, 2014
Accepted: July 22, 2014
Published: August 21, 2014

REFERENCES

Beltrao, P., Albanèse, V., Kenner, L.R., Swaney, D.L., Burlingame, A., Villén, J., Lim, W.A., Fraser, J.S., Frydman, J., and Krogan, N.J. (2012). Systematic functional prioritization of protein posttranslational modifications. Cell *150*, 413–425.

Beltrao, P., Bork, P., Krogan, N.J., and van Noort, V. (2013). Evolution and functional cross-talk of protein post-translational modifications. Mol. Syst. Biol. *9*, 714.

Boersema, P.J., Foong, L.Y., Ding, V.M.Y., Lemeer, S., van Breukelen, B., Philp, R., Boekhorst, J., Snel, B., den Hertog, J., Choo, A.B.H., and Heck, A.J. (2010). In-depth qualitative and quantitative profiling of tyrosine phosphorylation using a combination of phosphopeptide immunoaffinity purification and stable isotope dimethyl labeling. Mol. Cell. Proteomics *9*, 84–99.

Choudhary, C., and Mann, M. (2010). Decoding signalling networks by mass spectrometry-based proteomics. Nat. Rev. Mol. Cell Biol. *11*, 427–439.

Cohen, P. (2000). The regulation of protein function by multisite phosphorylation—a 25 year update. Trends Biochem. Sci. *25*, 596–601.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. *26*, 1367–1372.

Cox, J., and Mann, M. (2012). 1D and 2D annotation enrichment: a statistical method integrating quantitative proteomics with complementary high-throughput data. BMC Bioinformatics *13* (*Suppl 16*), S12.

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R.A., Olsen, J.V., and Mann, M. (2011). Andromeda: a peptide search engine integrated into the MaxQuant environment. J. Proteome Res. *10*, 1794–1805.

Donella-Deana, A., Marin, O., Cesaro, L., Gunby, R.H., Ferrarese, A., Coluccia, A.M.L., Tartari, C.J., Mologni, L., Scapozza, L., Gambacorti-Passerini, C., and Pinna, L.A. (2005). Unique substrate specificity of anaplastic lymphoma kinase (ALK): development of phosphoacceptor peptides for the assay of ALK activity. Biochemistry *44*, 8533–8542.

Echols, N., Harrison, P., Balasubramanian, S., Luscombe, N.M., Bertone, P., Zhang, Z., and Gerstein, M. (2002). Comprehensive analysis of amino acid and nucleotide composition in eukaryotic genomes, comparing genes and pseudogenes. Nucleic Acids Res. *30*, 2515–2523.

Fan, Y.X., Wong, L., and Johnson, G.R. (2005). EGFR kinase possesses a broad specificity for ErbB phosphorylation sites, and ligand increases catalytic-centre activity without affecting substrate binding affinity. Biochem. J. *392*, 417–423.

Findlay, G.M., Smith, M.J., Lanner, F., Hsiung, M.S., Gish, G.D., Petsalaki, E., Cockburn, K., Kaneko, T., Huang, H., Bagshaw, R.D., et al. (2013). Interaction domains of Sos1/Grb2 are finely tuned for cooperative control of embryonic stem cell fate. Cell *152*, 1008–1020.

Hein, M.Y., Sharma, K., Cox, J., and Mann, M. (2013). Proteomic Analysis of Cellular Systems. In Handbook of Systems Biology, *Chapter 1* (San Diego: Academic Press), pp. 3–25.

Hornbeck, P.V., Kornhauser, J.M., Tkachev, S., Zhang, B., Skrzypek, E., Murray, B., Latham, V., and Sullivan, M. (2012). PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. Nucleic Acids Res. *40*, D261–D270.

Humphrey, S.J., Yang, G., Yang, P., Fazakerley, D.J., Stöckli, J., Yang, J.Y., and James, D.E. (2013). Dynamic adipocyte phosphoproteome reveals that Akt directly regulates mTORC2. Cell Metab. *17*, 1009–1020.

Hunter, T. (2007). The age of crosstalk: phosphorylation, ubiquitination, and beyond. Mol. Cell *28*, 730–738.

Hunter, T. (2009). Tyrosine phosphorylation: thirty years and counting. Curr. Opin. Cell Biol. *21*, 140–146.

Hunter, T., and Sefton, B.M. (1980). Transforming gene-product of rous-sarcoma virus phosphorylates tyrosine. Proc. Natl. Acad. Sci. USA *77*, 1311–1315.

Iakoucheva, L.M., Radivojac, P., Brown, C.J., O'Connor, T.R., Sikes, J.G., Obradovic, Z., and Dunker, A.K. (2004). The importance of intrinsic disorder for protein phosphorylation. Nucleic Acids Res. *32*, 1037–1049.

Junger, M.A., and Aebersold, R. (2013). Mass spectrometry-driven phosphoproteomics: patterning the systems biology mosaic. Wiley Interdiscip. Rev. Dev. Biol. *3*, 83–112.

Kelstrup, C.D., Young, C., Lavallee, R., Nielsen, M.L., and Olsen, J.V. (2012). Optimized fast and sensitive acquisition methods for shotgun proteomics on a quadrupole orbitrap mass spectrometer. J. Proteome Res. *11*, 3487–3497.

Kettenbach, A.N., and Gerber, S.A. (2011). Rapid and reproducible single-stage phosphopeptide enrichment of complex peptide mixtures: application to general and phosphotyrosine-specific phosphoproteomics experiments. Anal. Chem. *83*, 7635–7644.

Kleiman, L.B., Maiwald, T., Conzelmann, H., Lauffenburger, D.A., and Sorger, P.K. (2011). Rapid phospho-turnover by receptor tyrosine kinases impacts downstream signaling and drug binding. Mol. Cell *43*, 723–737.

Lemeer, S., and Heck, A.J.R. (2009). The phosphoproteomics data explosion. Curr. Opin. Chem. Biol. *13*, 414–420.

Levy, E.D., Michnick, S.W., and Landry, C.R. (2012). Protein abundance is key to distinguish promiscuous from functional phosphorylation based on evolutionary information. Philos. Trans. R. Soc. Lond. B Biol. Sci. *367*, 2594–2606.

Lim, W.A., and Pawson, T. (2010). Phosphotyrosine signaling: evolving a new cellular communication system. Cell *142*, 661–667.

Loog, M., and Morgan, D.O. (2005). Cyclin specificity in the phosphorylation of cyclin-dependent kinase substrates. Nature *434*, 104–108.

Lundby, A., Secher, A., Lage, K., Nordsborg, N.B., Dmytriyev, A., Lundby, C., and Olsen, J.V. (2012). Quantitative maps of protein phosphorylation sites across 14 different rat organs and tissues. Nat. Commun. *3*, 876.

Marx, H., Lemeer, S., Schliep, J.E., Matheron, L., Mohammed, S., Cox, J., Mann, M., Heck, A.J.R., and Kuster, B. (2013). A large synthetic peptide and phosphopeptide reference library for mass spectrometry-based proteomics. Nat. Biotechnol. *31*, 557–564.

Michalski, A., Damoc, E., Hauschild, J.P., Lange, O., Wieghaus, A., Makarov, A., Nagaraj, N., Cox, J., Mann, M., and Horning, S. (2011). Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. Mol. Cell. Proteomics *10*, 011015.

Monetti, M., Nagaraj, N., Sharma, K., and Mann, M. (2011). Large-scale phosphosite quantification in tissues by a spike-in SILAC method. Nat. Methods *8*, 655–658.

Nagaraj, N., D'Souza, R.C.J., Cox, J., Olsen, J.V., and Mann, M. (2010). Feasibility of large-scale phosphoproteomics with higher energy collisional dissociation fragmentation. J. Proteome Res. *9*, 6786–6794.

Neuhauser, N., Michalski, A., Cox, J., and Mann, M. (2012). Expert system for computer-assisted annotation of MS/MS spectra. Mol. Cell. Proteomics *11*, 1500–1509.

Olsen, J.V., and Mann, M. (2013). Status of large-scale analysis of post-translational modifications by mass spectrometry. Mol. Cell. Proteomics *12*, 3444–3452.

Olsen, J.V., Blagoev, B., Gnad, F., Macek, B., Kumar, C., Mortensen, P., and Mann, M. (2006). Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. Cell *127*, 635–648.

Olsen, J.V., Vermeulen, M., Santamaria, A., Kumar, C., Miller, M.L., Jensen, L.J., Gnad, F., Cox, J., Jensen, T.S., Nigg, E.A., et al. (2010). Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. Sci. Signal. *3*, ra3.

Pinna, L.A., and Ruzzene, M. (1996). How do protein kinases recognize their substrates? Biochim. Biophys. Acta *1314*, 191–225.

Rappsilber, J., Ishihama, Y., and Mann, M. (2003). Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. Anal. Chem. *75*, 663–670.

Rix, U., and Superti-Furga, G. (2009). Target profiling of small molecules by chemical proteomics. Nat. Chem. Biol. *5*, 616–624.

Rush, J., Moritz, A., Lee, K.A., Guo, A., Goss, V.L., Spek, E.J., Zhang, H., Zha, X.M., Polakiewicz, R.D., and Comb, M.J. (2005). Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. Nat. Biotechnol. *23*, 94–101.

Sadowski, I., Stone, J.C., and Pawson, T. (1986). A noncatalytic domain conserved among cytoplasmic protein-tyrosine kinases modifies the kinase function and transforming activity of Fujinami sarcoma virus P130gag-fps. Mol. Cell. Biol. *6*, 4396–4408.

Sarno, S., Vaglio, P., Meggio, F., Issinger, O.G., and Pinna, L.A. (1996). Protein kinase CK2 mutants defective in substrate recognition. Purification and kinetic analysis. J. Biol. Chem. *271*, 10595–10601.

Schaab, C., Geiger, T., Stoehr, G., Cox, J., and Mann, M. (2012). Analysis of high accuracy, quantitative proteomics data in the MaxQB database. Mol. Cell. Proteomics *11*, 014068.

Swaney, D.L., Beltrao, P., Starita, L., Guo, A., Rush, J., Fields, S., Krogan, N.J., and Villén, J. (2013). Global analysis of phosphorylation and ubiquitylation cross-talk in protein degradation. Nat. Methods *10*, 676–682.

Tyanova, S., Cox, J., Olsen, J., Mann, M., and Frishman, D. (2013). Phosphorylation variation during the cell cycle scales with structural propensities of proteins. PLoS Comput. Biol. *9*, e1002842.

Ubersax, J.A., and Ferrell, J.E. (2007). Mechanisms of specificity in protein phosphorylation (vol 8, pg 530, 2007). Nat. Rev. Mol. Cell Biol. *8*, 665–665.

Ubersax, J.A., Woodbury, E.L., Quang, P.N., Paraz, M., Blethrow, J.D., Shah, K., Shokat, K.M., and Morgan, D.O. (2003). Targets of the cyclin-dependent kinase Cdk1. Nature *425*, 859–864.

Ward, J.J., McGuffin, L.J., Bryson, K., Buxton, B.F., and Jones, D.T. (2004). The DISOPRED server for the prediction of protein disorder. Bioinformatics *20*, 2138–2139.

Wiśniewski, J.R., Zougman, A., and Mann, M. (2009a). Combination of FASP and StageTip-based fractionation allows in-depth analysis of the hippocampal membrane proteome. J. Proteome Res. *8*, 5674–5678.

Wiśniewski, J.R., Zougman, A., Nagaraj, N., and Mann, M. (2009b). Universal sample preparation method for proteome analysis. Nat. Methods *6*, 359–362.

Wu, R., Haas, W., Dephoure, N., Huttlin, E.L., Zhai, B., Sowa, M.E., and Gygi, S.P. (2011). A large-scale method to measure absolute protein phosphorylation stoichiometries. Nat. Methods *8*, 677–683.

Zhang, J., Yang, P.L., and Gray, N.S. (2009). Targeting cancer with small molecule kinase inhibitors. Nat. Rev. Cancer *9*, 28–39.

Zhou, H., Low, T.Y., Hennrich, M.L., van der Toorn, H., Schwend, T., Zou, H., Mohammed, S., and Heck, A.J.R. (2011). Enhancing the identification of phosphopeptides from putative basophilic kinase substrates using Ti (IV) based IMAC enrichment. Mol. Cell. Proteomics *10*, 006452.

Zhou, H., Di Palma, S., Preisinger, C., Peng, M., Polat, A.N., Heck, A.J.R., and Mohammed, S. (2013). Toward a comprehensive characterization of a human cancer cell phosphoproteome. J. Proteome Res. *12*, 260–271.