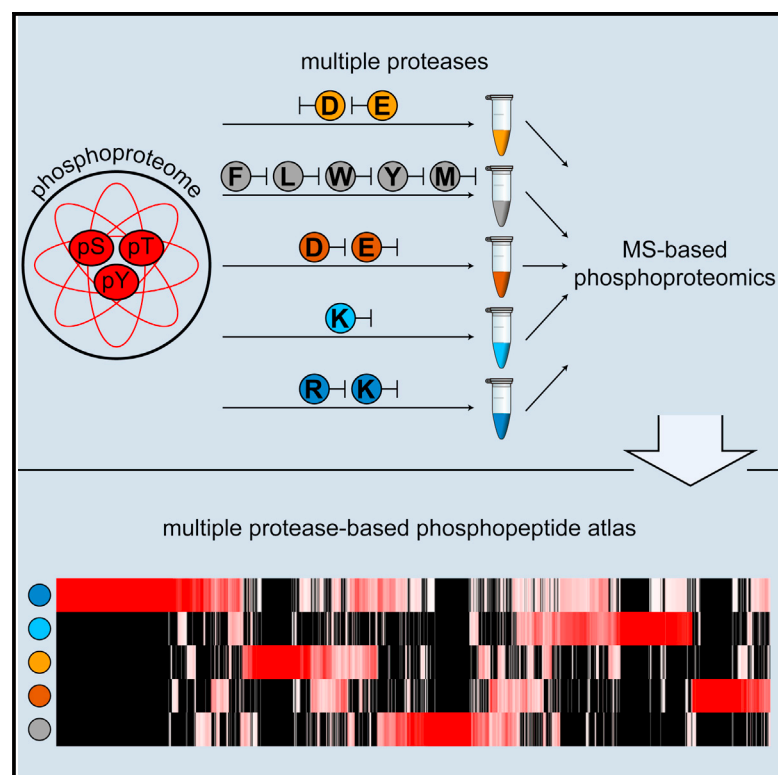# An Augmented Multiple-Protease-Based Human Phosphopeptide Atlas

Graphical Abstract

Authors

Piero Giansanti, Thin Thin Aye, Henk van den Toorn, Mao Peng, Bas van Breukelen, Albert J.R. Heck

Correspondence

a.j.r.heck@uu.nl

In Brief

Giansanti et al. present an augmented human phosphopeptide atlas of 37,771 unique phosphopeptides. Using a multiple protease workflow in combination with highly selective Ti^{4+}-IMAC phosphopeptide enrichment, the authors show that optimal detection of each protein individual phosphorylation event is linked to a preferred protease.

Highlights

- There is a considerable bias in the current deposited phosphoproteome

- Key phosphosites can be occluded by using trypsin alone

- Optimal MS detection of a phosphorylation site is linked to a favorable protease

- An augmented human phosphopeptide atlas is presented

CrossMark

CellPress

An Augmented Multiple-Protease-Based Human Phosphopeptide Atlas

Piero Giansanti,[1,2,3] Thin Thin Aye,[1,2,3] Henk van den Toorn,[1,2,3] Mao Peng,[1,2] Bas van Breukelen,[1,2] and Albert J.R. Heck[1,2,*]

[1]Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Paadualaan 8, 3584 Utrecht, the Netherlands
[2]Netherlands Proteomics Centre, Paadualaan 8, 3584 Utrecht, the Netherlands
[3]Co-first author
*Correspondence: a.j.r.heck@uu.nl
 http://dx.doi.org/10.1016/j.celrep.2015.05.029

SUMMARY

Although mass-spectrometry-based screens enable thousands of protein phosphorylation sites to be monitored simultaneously, they often do not cover important regulatory sites. Here, we hypothesized that this is due to the fact that nearly all large-scale phosphoproteome studies are initiated by trypsin digestion. We tested this hypothesis using multiple proteases for protein digestion prior to Ti^{4+}-IMAC-based enrichment. This approach increases the size of the detectable phosphoproteome substantially and confirms the considerable tryptic bias in public repositories. We define and make available a less biased human phosphopeptide atlas of 37,771 unique phosphopeptides, correlating to 18,430 unique phosphosites, of which fewer than 1/3 were identified in more than one protease data set. We demonstrate that each protein phosphorylation site can be linked to a preferred protease, enhancing its detection by mass spectrometry (MS). For specific sites, this approach increases their detectability by more than 1,000-fold.

INTRODUCTION

Cellular signaling proceeds largely via cascades of post-translational modifications, in which reversible protein phosphorylation provides a key mechanism (Huang and White, 2008; Rigbolt and Blagoev, 2012). Site-specific protein phosphorylation can be monitored by site-specific phospho-antibodies, such as those raised against, for instance, pERK T202/Y204 and pSRC Y419. Although powerful, there are only a limited number of these antibodies available, hardly sufficient to monitor the more than 100,000 unique phosphosites present in a human cell. Other caveats in using these antibodies are their limited specificity, recognizing multiple sites in a single protein, or similar phospho-sequences in other proteins. Moreover, these approaches are difficult to multiplex for use in high-throughput assays.

Mass spectrometry (MS)-based phosphoproteomics has recently surfaced as the method of choice for global and high-throughput protein analysis. Immense progress in both mass spectrometric instrumentation (Hebert et al., 2014; Michalski et al., 2011, 2012) and sample preparation and analysis (Di Palma et al., 2012; Ruprecht and Lemeer, 2014; Wiśniewski et al., 2009; Yates et al., 2014) have allowed this technology to confidently identify as many as thousands of proteins and phosphorylation sites in a single experiment and to accurately quantify changes in protein expression or post-translational modifications (PTMs). Recently, we demonstrated the high potential of a new material for phosphopeptide enrichment (Zhou et al., 2013). This affinity matrix termed titanium (IV)-IMAC (Ti^{4+}-IMAC) was shown to possess high selectivity, sensitivity, and quantification reproducibility, allowing in-depth monitoring of more than 10,000 phosphorylation events by a single-step phosphopeptide enrichment (de Graaf et al., 2014).

Predominantly, large-scale phosphoproteomics analyses are based on peptides derived from the tryptic digestion of proteins in a lysate (Mallick et al., 2007; Sharma et al., 2014). Trypsin represents a valid choice because it is highly specific, very effective and, compared to the other available proteases, generates a higher number of peptides in the preferred mass range suitable for identification by MS (Guo et al., 2014; Swaney et al., 2010). As a result, the vast majority of the reported workflows in proteomics are dominated by using exclusively trypsin (Tsiatsiani and Heck, 2015; Wilhelm et al., 2014). However, potentially interesting sequences, and thus particular relevant phosphosites, will remain occluded by this approach, as the tryptic peptides generated do not always possess appropriate physicochemical properties that make them suitable for detection by LC-MS/MS.

Although it has been shown that proteome coverage can be increased by using multiple alternative proteases (Bian et al., 2012; Gauci et al., 2009; Guo et al., 2014; Peng et al., 2012; Swaney et al., 2010), this approach has not systematically been investigated for the analysis of PTMs. Here, we report a systematic study using five commercially available proteases to extend the coverage of the phosphoproteome. We targeted the phosphoproteome of human Jurkat T cells stimulated by PGE_2 and used Ti^{4+}-IMAC for the enrichment of the phosphopeptides from lysates digested in parallel by AspN, chymotrypsin, GluC,
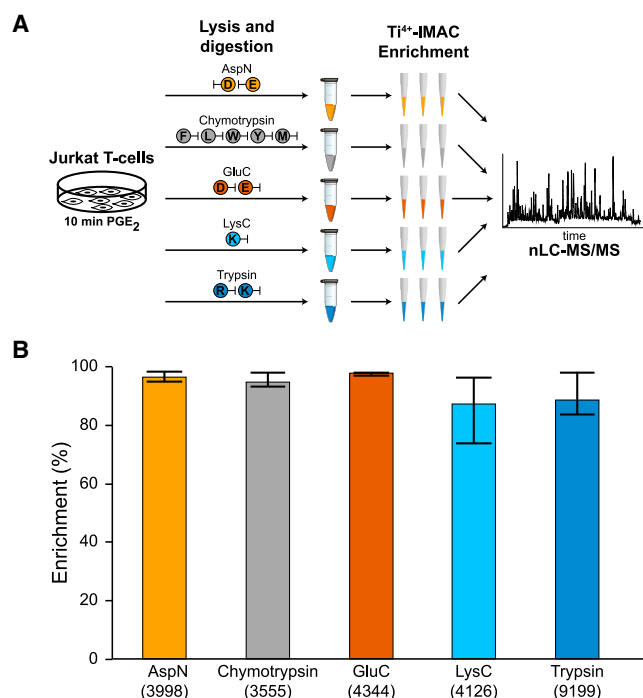
A

B

Figure 1. Generation of the Human Phosphopeptide Atlas

(A) Jurkat T lymphocyte cells were stimulated for 10 min with PGE$_2$, lysed, and digested in parallel with one of the five proteases: AspN, chymotrypsin, GluC, LysC, and trypsin. Each digest was divided over three Ti^{4+}-IMAC phosphopeptide enrichment columns, loading 200 μg per enrichment, resulting in 15 samples that were analyzed by nLC-MS/MS. The color scheme representing the data on each individual protease is kept consistent throughout the manuscript. These colors are AspN, orange; chymotrypsin, gray; GluC, vermillion; LysC, sky blue; and trypsin, blue.

(B) Enrichment efficiency and total number of unique phosphosites detected. The bar chart represents the median enrichment efficiency calculated as the percentage of phosphopeptides in the six MS runs, demonstrating the high selectivity (typically above 90%) of the Ti^{4+}-IMAC enrichment for all used proteases. For each population, the whiskers represent the minimum and maximum selectivity. The number of unique phosphosites identified in the data sets originating from each protease is reported underneath the bars.

LysC, and trypsin, starting with just 600 μg for each digest. Using a stringently filtered cumulative data set of 37,771 unique phosphopeptides, linked to 18,430 unique phosphosites, we were able to answer the following questions: is it beneficial to do phosphoproteomics using multiple enzymes rather than extending the number of trypsin replicates? Do different proteases possess unique phosphopeptides features? Is the enzymatic efficiency affected by the presence of a phosphorylated site? Is label-free quantitative phosphoproteomics possible when using other enzymes than trypsin? Is there at present a detectable bias toward "tryptic" phosphosites in the very large public depositories? The final result of our work is a human phosphopeptide resource, made publicly available, that can be queried by using a simple and intuitive web interface to facilitate the identification of the most-suitable protease for both shotgun as well as targeted MRM/PRM/SWATH-based phosphoproteomics studies.

RESULTS AND DISCUSSION

Generation of the Data for an Augmented Human Phosphopeptide Atlas

All phosphoproteomics data presented here were obtained from human Jurkat T cells that were harvested following 10 min of PGE$_2$ stimulation. PGE$_2$ increases the levels of cAMP, thereby activating among other the cAMP-dependent kinase PKA. We did choose PGE$_2$-activated T cells, as recently a rather large trypsin-based phosphoproteomics data set was made available for this system using a similar experimental approach (de Graaf et al., 2014). The activated T cells were lysed and extracted proteins were digested with each of the five enzymes, i.e., AspN, chymotrypsin, GluC, LysC, and trypsin, using identical protocols as described in detail earlier (Low et al., 2013). Three independent Ti^{4+}-IMAC enrichments were used for each of the five digests to specifically enrich for phosphopeptides, and only 200 μg of digest material were used as input per enrichment (Figure 1A). All 15 samples were analyzed in replicate by nLC-MS/MS using 150-min gradients and a decision-tree-driven peptide fragmentation scheme, selecting the appropriate activation technique being either CID or ETD, depending on the precursor charge state and m/z (Frese et al., 2011; Swaney et al., 2008).

In total, more than 1.1 × 10^6 MS/MS spectra were acquired (Table 1), resulting in 37,771 unique phosphopeptides originating from 5,326 phosphoproteins. A complete list of all the identified peptide to spectrum matches (PSMs), (phospho)peptides, including confidence scores for identification and site localization for each enzyme, can be found in Table S1 and on the web-based resource page (http://phosphodb.hecklab.com). Among the five data sets, trypsin comprised the largest number of unique phosphopeptide identifications (13,476), followed by GluC (7,502), LysC (6,595), AspN (6,407), and chymotrypsin (5,376).

Comparison of Data Sets Originating from Different Proteolytic Digests

As Ti^{4+}-IMAC has been so far only used for the enrichment of phosphopeptides from tryptic digests, we first assessed the feasibility and validity of using this material for the enrichment of non-tryptic digests. The data from three independent Ti^{4+}-IMAC enrichments clearly indicate a high selectivity (Figure 1B), ranging from on average 86% for LysC up to 98% for GluC, similar to the enrichment efficiency obtained for trypsin (90%). Thus, the selectivity for enrichment by Ti^{4+}-IMAC seems protease independent. We next examined the number of phosphorylation sites identified by each protease (Table 1; Figure 1B). Using the same amount of sample (600 μg), we recovered 9,199 unique phosphorylation sites (from 13,476 unique phosphopeptides) in the tryptic digest; whereas, around 4,000 unique phosphosites sites could be recovered from each of the other digests. All these numbers are quite favorable (considering the low amount of sample used) but also clearly show that trypsin seemingly outperforms the other enzymes in number of identified sites. Such a lower efficiency for GluC, AspN, LysC, and chymotrypsin is in agreement with what has been reported at the unmodified peptide level (Low et al., 2013; Swaney et al., 2010). However, we hypothesize that there may be multiple

Table 1. Unique Phosphopeptide and Phosphoprotein Identifications in Each of the Protease Data Sets

Protease	Trypsin	LysC	AspN	GluC	Chymotrypsin	Cumulative
No. MS/MS	248,753	224,227	204,657	208,559	225,911	1,112,107
No. PSMs	87,650	59,104	45,095	67,808	40,064	299,721
Success rate (%)	35.2	26.4	22.0	32.5	17.7	26.9
No. CIDs	64,970	24,523	19,438	28,714	22,270	159,915
No. ETDs	22,680	34,581	25,657	39,094	17,794	139,806
CID/ETD ratio	2.86	0.71	0.76	0.73	1.25	1.14
No. phosphopeptides	13,476	6,595	6,407	7,502	5,376	37,771
No. phosphoproteins	3,519	1,868	2,021	2,225	1,936	5,326

Although the number of MS/MS events and phosphopeptide enrichment efficiency are alike in all digests, the number of unique phosphopeptides detected is twice as high for trypsin as for the other proteases. ETD is highly beneficial and complementary in the identification of phosphopeptides, especially in the digest of LysC, AspN, and GluC.

origins for this observation. At least part of the higher number of identifications for trypsin can be attributed to the higher identification rates (PSMs), likely due to a positive bias of search engines toward trypsin (Granholm et al., 2014). In addition, superior fragmentation of tryptic peptides, especially in CID/HCD, attributes further to the higher number of identification for trypsin (Meyer et al., 2014; Swaney et al., 2010). In line with this notion, our data show that the use of ETD fragmentation complementary to CID/HCD fragmentation is substantially more beneficial for digests of LysC, AspN, and GluC and not so much for trypsin (see Table 1).

Next, we evaluated the unique phosphorylation sites per protease data set and between the five data sets. To compare the data sets, we applied stringent site assignments criteria by using the phosphoRS algorithm (Taus et al., 2011), which allowed to confidentially localize the phosphorylation event with amino acid resolution. The resulting data set reveals that, by using different proteases, we can significantly enhance our phosphopeptide atlas, as the five data sets turned out to be extremely complementary. As illustrated in Figure 2A, we cumulatively identify 18,430 unique phosphosites, albeit that just about 27% of these were identified in more than one protease data set and only a marginal fraction of 0.3% (i.e., 61) could be identified in all five data sets.

Characterization of the Identified Phosphopeptides Reveals that Phosphorylation Induces Widespread Protease Missed Cleavages

To gain more insight into the complementarity observed among the five data sets, we first evaluated the global physico-chemical characteristics of the phosphopeptides. First, we examined the length distribution of the identified phosphopeptides (Figure S1A), which revealed that they possess on average similar length distributions, with most of them having a length in between 14 and 25 amino acids. Not surprisingly, identified phosphopeptides generated by trypsin digestion are on average relatively shorter (7–20 amino acids long) compared to phosphopeptides identified after GluC digestion, which generates longer sequences (14–30 amino acids long). Although the results generally correlated with data reported in other studies for unmodified peptides (Biringer et al., 2006; Kalli and Håkansson, 2010; Swaney et al., 2010), we noticed

that there was a substantial shift in the observed average peptide length for trypsin, from ~10 for unmodified peptides to ~18 amino acids for phosphopeptides. This behavior has been attributed by the presence of a phosphorylation site in close proximity to the sites of proteolytic cleavage, which prevent trypsin and LysC cleavage and neighboring sites (Molina et al., 2007). To explore whether a similar phenomenon exists in the other protease data sets, we investigated the frequency of missed cleavage events in all five data sets. As depicted by the violin plots in Figure S1B, relative high number of missed cleavages is present in the phosphopeptide data sets from each proteolytic digest. If the enzymatic digestions were inefficient, one would expect unmodified peptides to show the same distribution. On the contrary, the analysis on the unmodified peptides, co-purified by the Ti^{4+}-IMAC material, reveal a more efficient digestion, similar to the one commonly observed for the same enzymes in large-scale proteome studies (Low et al., 2013). Thus, the here-generated data set reiterates that the number of missed cleavages is significantly higher for phosphorylated peptides, not only in digest generated by trypsin and LysC but also in AspN, GluC, and chymotrypsin digests. To gain closer insights into the influence of phosphorylation on digestion, we probed the composition of the amino acid sequence surrounding the missed cleavage site for the phosphopeptides detected in each data set (Figure S1C). For each data set, the residue corresponding to an uncleaved site was centered and the distance to the nearest localized phosphoresidue, within ±5 positions, was calculated. If phosphorylation has no influence on the cleavage, an even distribution of the phosphoresidue should be observed around the missed cleaved bond. In accordance with previous studies (Dickhut et al., 2014; Gershon, 2014), this analysis shows for trypsin and LysC a phosphoresidue in the positions +1, +2, and +3 selectively hampers cleavage. For AspN and GluC, phosphorylation impedes cleavage mostly when located at position −2, whereas for chymotrypsin, a phosphorylation at the +1 position hampers protease activity. The negative correlation between cleavage efficiency and nearby phosphorylation events is thus generally protease independent, albeit that specific protease-dependent rules govern the most-affecting sites.

We also determined the charge distribution of the phosphopeptides in each data set. As illustrated in Figure S1D, trypsin
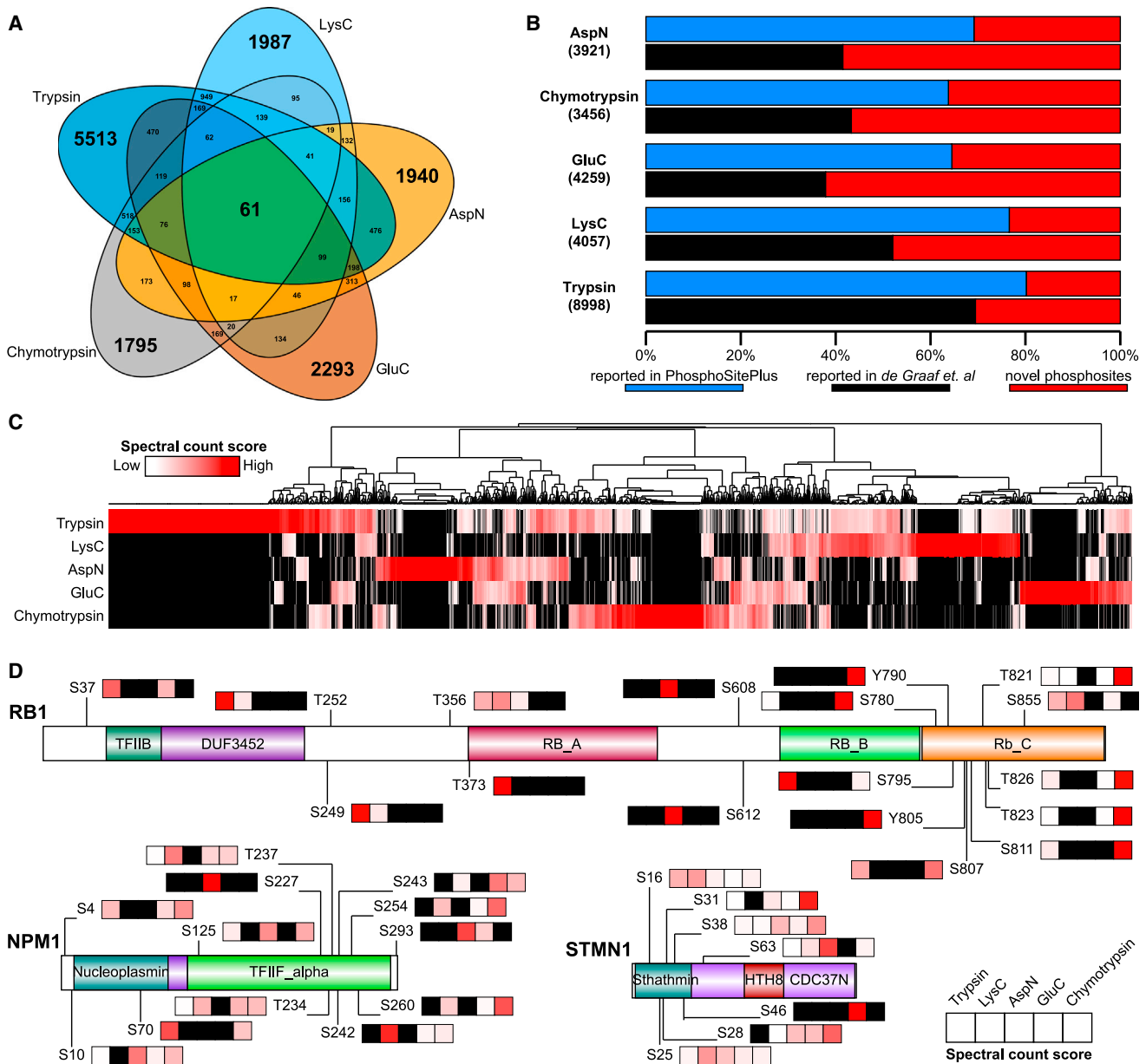
Figure 2. Qualitative and Quantitative Benefit of Using Multiple Proteases in Phosphoproteomics

(A) Venn diagrams displaying the overlap in detected unique phosphosites between the data sets generated by trypsin, LysC, AspN, GluC, and chymotrypsin. Nearly 3/4 of the phosphosites (i.e., ~13,500) were not detected by more than one protease, indicative of the high orthogonality of the multi-protease strategy.

(B) Benchmarking the unique and distinct phosphorylation sites detected in each digest to the human phosphorylation sites reported in the comprehensive PhosphoSitePlus database depository (~153,900 entries) and the related large Jurkat (phospho)proteome data set reported earlier by de Graaf et al. (2014; ~16,200 entries). The here-identified phosphopeptide sites in the AspN, GluC, and chymotrypsin digests are clearly underrepresented both in the public depository as well as in the largest reported Jurkat cell (tryptic) phosphoproteome. These comparisons clearly reveal a substantial tryptic bias in public depositories.

(C) Heatmap, based on spectral count scores, illustrating the contribution of each protease in the detection of a particular phosphosite. Black color means not detected. Phosphosites were further grouped using hierarchical clustering (distance metric was Euclidean correlation and linkage method was average).

(D) Domain structures for three representative phosphoproteins (RB1, STMN1, and NPM1), bearing multiple phosphosites, whereby by using bars is presented how well they are detectable in the different digests. Black color means not detected. Residue numbers of the identified phosphorylation sites are indicated.

digestion generates, predominantly, doubly and triply protonated peptides. In contrast, LysC, AspN and GluC generated substantially more triply and higher charge state phosphopeptides. Notably, even though the charge and length properties of the

chymotryptic peptides are very similar to the ones generated by trypsin, the success rate in identification for chymotrypsin was a factor 2 lower (Table 1). This can be partly explained by the low specificity of chymotrypsin, which may cleave at five

different amino acids, posing a bigger challenge to current search engines. Despite this lower efficiency, the added value of using chymotrypsin is illustrated by the 1,795 uniquely identified phosphosites (Figure 1A). We think this can be explained by the rather unique capability of chymotrypsin to cleave peptide bonds consisting of amino acids with hydrophobic (L and M) and aromatic large side chains (F, W, and Y). Therefore, we hypothesize that, by using chymotrypsin, more peptides from hydrophobic protein regions can be retrieved, often missed in tryptic digests.

Tryptic Bias in Public Phosphopeptide Depositories

We evaluated the here-identified 18,430 phosphorylation sites against two very large public data depositories (Figure 2B). First, we matched the phosphosites identified in the data sets of each protease against one of the largest manually curated database repositories: PhosphoSitePlus (Hornbeck et al., 2012; ~153,900 human phosphorylation sites), which includes results from large-scale phosphoproteomics experiments but also from lower-throughput assays based on immune purification using phospho-specific antibodies, cumulating data from different cellular and/or tissue origins. We observed that our multiple-protease-based data sets confidently identified 6,032 phosphorylation sites not yet reported in PhosphoSitePlus. Although our trypsin data set yielded also a number of not yet reported sites, more than 70% of the novel sites came from the data sets using the other proteases.

Based on the fact that they have similar sizes, a more-fair benchmark we hypothesized would be to compare the current data sets to the recently (by us) reported exhaustive trypsin-based phosphoproteomics data set acquired from the same Jurkat T lymphocyte cells by using the same enrichment strategy, albeit by using 54 distinctive Ti^{4+}-IMAC-based enrichments (~16,200 phosphorylation sites; de Graaf et al., 2014). In that earlier experiment, we likely approached the maximum in detection of tryptic phosphosites by doing 54 replicates across six time points, instead of the three used here. Strikingly, this analysis (Figure 2B) revealed that many phosphosites found by using the alternative proteases were not detected earlier in the exhaustive trypsin-based experiments. In some of these data sets, this accumulated to close to 2/3 of the data set, i.e., AspN (59%), chymotrypsin (57%), and GluC (62%). From these data, we conclude that, although trypsin performs very well, it provides a biased representation of the phosphoproteome. This bias can be substantially reduced, making use of the orthogonality of the alternative proteases applied here.

Quantitative Assessment of Protease Bias for Individual Phosphoproteins: Impact on Studying Cellular Signaling

In absence of any protease bias, the five proteases should equally contribute to the detection of a particular phosphosite on a given protein, especially for the ~3,300 phosphosites for which high numbers of spectra ($\geq$10) were successfully and unambiguously matched to a peptide sequence. However, when we evaluated each protease contribution by using a spectral count score (see Data Analysis section), we observed massive differences (Figures 2C and S2). Many phosphosites

were either highly overrepresented or underrepresented in particular protease data sets. Consequently, certain protein phosphorylation events are considerably better detectable by using one or two proteases over the others. To further illustrate this phenomenon, we focused on a group of key signaling phosphoproteins, e.g., a list of reported well-known oncogenes (Table S2). In signal transduction, reversible protein phosphorylation often represents an activating or deactivating switch for protein activity, which can play an important role in the pathogenesis of human cancers (Chong et al., 2008). Therefore, the identification and characterization of the phosphorylation events associated to oncogenic signaling are particularly important. Our data clearly show that, even though trypsin is an efficient and robust protease that generates the highest number of (phospho)peptides detectable by MS experiments, it may not always be the optimal choice for specific important regulatory sites. We further zoom in on three illustrative examples from the full list of proteins given in Figure S2, namely, the retinoblastoma protein (RB1), stathmin (STMN1), and nucleophosmin (NPM1) (Figures 2D and S3). Complementary, the web-based database set up to make our data publicly available (http://phosphodb. hecklab.com) provides similar analysis on all the here-reported phosphoproteins.

The retinoblastoma protein is an important tumor suppressor protein that is dysfunctional in several major cancers (Murphree and Benedict, 1984). The hypo-phosphorylated form interacts with and sequesters the E2F1 transcription factor, leading to cell cycle arrest. On the contrary, the hyper-phosphorylated form is unable to interact with E2F1 and, therefore, unable to restrict progression from the G1 to the S phase of the cell cycle. The main regulators of RB1 activity are several members of the cyclin-dependent kinases (CDKs) family, which can phosphorylate RB1 at several different sites, such as S249, S252, S807, S811, T821, and T826 (Knudsen and Wang, 1997). Evidently, in ideal phosphoproteomics experiments, all these sites should be monitored. For three of these sites, our data clearly reveal that chymotrypsin is the protease that would substantially facilitate their detection by MS (Figures 2D and S3).

Another example is provided by STMN1, a protein involved in the biogenesis and remodeling of the cellular microtubule cytoskeleton. STMN1 activity depends on phosphorylation of at least four key serine residues (S16, S25, S38, and S63) by different kinases (Santamaría et al., 2009). Our data show that, whereas S16, S25, and S38 could be detected by each protease, S63 clearly possesses a strong preference in detection by using AspN. This suggests that S63 would be underrepresented in conventional tryptic-based phosphoproteomics approaches. Indeed, when we evaluated these four sites in the PhosphoSitePlus database, we discovered that S16, S25, and S38 were detected by over 100 large-scale mass-spectrometry-based experiments, whereas S63 has been reported in just 50 data sets. Moreover, in the latter case, these observations originated mainly from experiments in which a targeted low-throughput approach had been used, i.e., by using motif-specific antibodies directed against kinase motifs (e.g., PKA, PKC, and PKD). Also, STMN1 S31 phosphorylation was readily detected with over hundreds of PSMs in the chymotrypsin and AspN data sets,
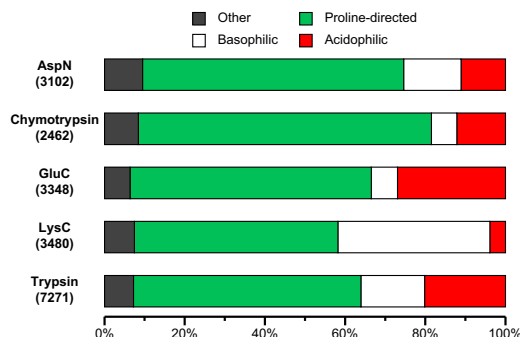
Figure 3. Distribution of the Sequence Motifs Extracted from the Protease-Linked Phosphorylation Data Sets

The motif-x algorithm revealed more than 130 unique phosphorylation motifs in the phosphopeptides data sets. They could generally be classified into proline-directed, basophilic, and acidophilic motifs (http://www.hprd.org). The contribution of each class of motifs is given, and the total number of phosphosites that could be assigned to these motifs is shown in brackets. Most notably, whereas for trypsin, AspN, and chymotrypsin the ratio between acidophilic and basophilic motifs is close to 1:1, GluC and LysC show a clear bias toward the acidophilic and the basophilic motifs, respectively. Additionally, compared to the other proteases, chymotrypsin seems to be biased toward proline-directed motifs.

but only six PSMs were detected in the much-larger tryptic data set. Validating our trypsin bias hypothesis, this site turned out to be also underrepresented in PhosphoSitePlus.

As third example, we highlight nucleophosmin. NPM1 is overexpressed, mutated, and chromosomally translocated in many tumor types (Falini et al., 2007). During the different phases of the cell cycle NPM1 can be phosphorylated at multiple sites by PLK or CDK kinases, modifications that have been proposed to trigger or abolish NMP1 functions. In our data, we detect a high number of PSMs for S4, S10, S70, S125, T234, T237, and S242 and lower levels of PSMs for S260, S227, S293, S243, and S254 (Table S2). However, of these S10, T234, T237, and S242 are hardly detectable in the tryptic data set, whereas these sites have high numbers of PSMs in the AspN and chymotrypsin data sets (Figures 2D and S3). GluC digestion seems to be really beneficial for T234 and T237, whereas S70 is only detected with high PSMs in the data from the tryptic and chymotryptic digests. In summary, our data on nucleophosmin reveal that each site in this protein has a preference in detectability correlated with different proteases (Figure S3). Moreover, only by using all five proteases, we are able to map all known important regulatory phosphosites in this protein.

These three proteins illustrate the protease complementarity, which holds true for all the other phosphoproteins detected and listed in Table S2. Clearly, when the phosphorylation status of a given protein needs to be monitored, the use of complementary proteases is highly beneficial. Therefore, targeted proteomics studies should be expanded to include non-tryptic peptides, as intensities likely increase by factors between 10 and 10,000.

Presently, a number of factors limit the usefulness of these non-tryptic phosphopeptides for systems biology studies. First,

the public available repositories are nearly exclusively based on tryptic peptides, underrepresenting many potential interesting sites. Second, spectra are not always accessible to assess the quality of phosphopeptide identification and the accuracy of the phosphorylation site assignment. Moreover, the data are mostly presented as lists of phosphosites, whereas information at the phosphopeptide level might provide more useful information for further analysis such as targeted MRM/PRM/SWATH. To enhance the utility of our augmented phosphopeptide atlas, we provide our data as a searchable resource for targeted phosphoproteomics experiments (http://phosphodb.hecklab.com). The users can browse through the phosphorylation sites on their proteins of interest, apply custom filters, visually identify which protease would be the most suitable for the detection of the sites of interest, and export the necessary parameters to setting up the MRM/PRM-like assays.

Our resource provides a first draft of a less-biased human phosphopeptide atlas that may be further expanded by importing data from other proteases, other cell lines, and tissue. Moreover, we believe that this resource, which also includes all raw data and MS/MS spectra interpretations, may help bioinformaticians in their efforts to develop and improve computational prediction and analysis tools for non-tryptic peptides in phosphoproteomics analysis, which we here identified as being one of the current likely bottlenecks.

Computational Analysis on the Identified Phosphorylation Sites

To assess a potential preference for certain kinase motifs in the five phosphopeptides data sets, we subjected the here-identified phosphorylation sites first to the motif-x algorithm (Schwartz and Gygi, 2005). A total of 132 distinct motifs could be defined among all the five proteases data sets (Table S3). By convention, we could classify these kinase motifs into four well-defined groups: acidic, basophilic, proline-directed, and "other" motifs (Figure 3). This analysis revealed that the sequences bearing the [pS/pT]P motif, target of the very large family of proline-directed kinases including, among others, CDKs and MAPKs (Lu et al., 2002), were almost evenly spread across all the five proteases data sets, with just AspN and chymotrypsin possibly displaying a slight overrepresentation. Considering next the substrates of basophilic and acidophilic kinases, a clear bias is observed in between the data sets, with especially LysC and GluC being specific outliers (Figure 3).

Motifs in the basophilic group convolute around the PKA/Akt/PKC [R/K]X[R/K]XX[pS/pT] consensus motif (Pearce et al., 2010). Many of the arginine-containing sequences were retrieved from the tryptic digests, whereas lysine-rich motifs were mainly present in the LysC data set, making these two proteases the most suitable when the interest lays in the analysis of basophilic kinases substrates. Likewise, we hypothesized that AspN and GluC would be the protease of choice in detecting substrates of acidophilic kinases (e.g., CK1, CK2, and PLKs), which mostly target serine and threonine residues flanked by acidic residues (Amanchy et al., 2007). Our analysis

revealed that only GluC showed a clear bias toward detecting peptides bearing an acidophilic motif.

All the identified phosphorylation sites were further subjected to the NetworKIN algorithm (Horn et al., 2014) to predict the kinases involved (Table S3). This analysis predicted that phosphorylation of the acidophilic motifs was mostly associated to the kinases CK1, CK2, and SGK1. These three kinases were mainly overrepresented in the trypsin data set. In agreement with above basophilic kinases, members of the AGC and CAMK groups were predominantly enriched in the trypsin and LysC data sets. Interestingly, the NetworKIN analysis also revealed a bias toward LysC for almost all members of the GRK family (Figure S4). GRKs are activated downstream of GPCRs, which we here activated by PGE$_2$, and function to phosphorylate and control the activity of these receptors. Unlike most AGC kinases, for GRKs phosphorylation, the consensus motifs are not clearly defined yet. It has been reported that some members prefer acidic residues flanking the phospho-acceptor sites whereas other members favor basic residues (Pitcher et al., 1998). Indeed, a closer inspection of the phosphosites detected in our data sets confirmed this binary characteristic, with lysine, glutamic acid, or both residues in close proximity to the phosphorylated serine/threonine.

A Protease Bias in Label-free Quantification of Phosphorylation Sites

Due to their highly dynamic spatial regulation, it is crucial to perform quantitative studies on protein phosphorylation to fully understand the signaling networks controlling cellular fate. Many different strategies have been developed to accomplish this task (Macek et al., 2009). Recently, others and we demonstrated the possibility of performing in-depth and high-throughput reproducible label-free phosphoproteome quantification (Courcelles et al., 2013; de Graaf et al., 2014; Montoya et al., 2011; Soderblom et al., 2011). Here, we assessed the feasibility of obtaining similar quantitative data when proteases other than trypsin are used, making it possible to profile phosphosites that would be easily missed in a conventional trypsin-based workflow. Such an analysis would provide insight into the reproducibility of the enrichment protocol but also the reproducibility in digestion by the different proteases. To minimize any potential bias of the different algorithms used to perform identification and quantification, we processed the here-obtained data through the same computational pipeline as previously described (de Graaf et al., 2014).

The data from three independent enrichments within each of the distinct protease data sets show a high quantitative reproducibility across all phosphosites, with a correlation typically around 0.85 (Figures 4A and S5). Therefore, we conclude that quantitative phosphoproteomics, including label-free quantification, is possible, not only by using trypsin, as shown before (de Graaf et al., 2014; Montoya et al., 2011), but also by using any of the other four alternative proteases used here. However, when we compare data generated by using more than one protease, the correlation is low (r ∼0.25–0.55). These lower correlations further endorse the previously demonstrated substantial protease bias in data sets generated in yeast on non-modified peptides (Peng et al., 2012). In Figure 4B, a

zoom in of some of these correlation plots is given, highlighting a few illustrative phosphosites in which one protease clearly outperforms the other, resulting sometimes in a more than 1,000-fold higher phospho-site bearing peptide intensity. As an example, S37 on *RB1* is easily detected as a high abundant peptide following trypsin digestion. In the GluC digest, its intensity is much lower, affecting the accuracy of the peak detection. Similarly, the tryptic peptide bearing S10 on NPM1 is low abundant, whereas the AspN peptide bearing this S10 site is about 4-fold more intense. We believe that these results support the observation that the best peptides are not necessarily tryptic (Peng et al., 2012), highlighting the valuable contribution of non-tryptic peptides also in quantitative phosphoproteomics analysis.

To further demonstrate that some phosphorylation sites can be hardly observable in tryptic digests, we conducted a targeted proteomics experiment making use of a well-defined inclusion list (Jaffe et al., 2008; Schmidt et al., 2008). With this approach, the mass spectrometer is specifically instructed to sequence only pre-selected ion species, overcoming the stochastic undersampling that may occur during LC-MS/MS analysis. As a proof of principle, we selected two phosphorylation sites with a well-known biological significance, S780 on RB1, which is one of the residues modulating the activity of RB1 during the cell cycle, and T638 on PRKCA, one of three residues needed to be phosphorylated for full activation of this kinase. We hypothesized that the two phosphosites would be easier to assay using chymotrypsin and AspN, respectively, rather than trypsin because of the observed differences in spectral count values for the candidate phosphopeptides. Equal starting amounts of Jurkat cells lysate were digested using AspN (five replicates), chymotrypsin (five replicates), or trypsin (seven replicates). Each digest was subjected to phosphopeptide enrichment as described above, and resulting phosphopeptides were then analyzed by using the inclusion-list-based approach. As depicted in Figure 4C, a difference in peak area of ∼200-fold was observed for S780 on RB1. Alike for T638 on PRKCA, the use of AspN led to an ∼20-fold more-intense phospho-site-bearing peptide when compared to its tryptic counterpart. To demonstrate that this difference is not caused by different amount of sample loaded on the column during the nLC-MS/MS analysis, we compared the total ion current (TIC) of all the inclusion-list-based acquisitions (Figure S6). This analysis revealed that similar intensity of the TICs was measured across the different digests, demonstrating that equal amounts were loaded of the phosphopeptides. We conclude thus that the biases revealed by our spectral-counting-based methods are also detectable and present when adopting a targeted proteomics approach.

Conclusions

Although high-throughput phosphoproteomics approaches can nowadays identify several thousands of unique phosphosites, it has become apparent that a large portion of regulatory important phosphorylation events remain elusive. These issues are partly due to the fact that many sites are simply inaccessible or very hard to detect following digestion with trypsin, hampering MS-based studies. In this work, we demonstrated that the use
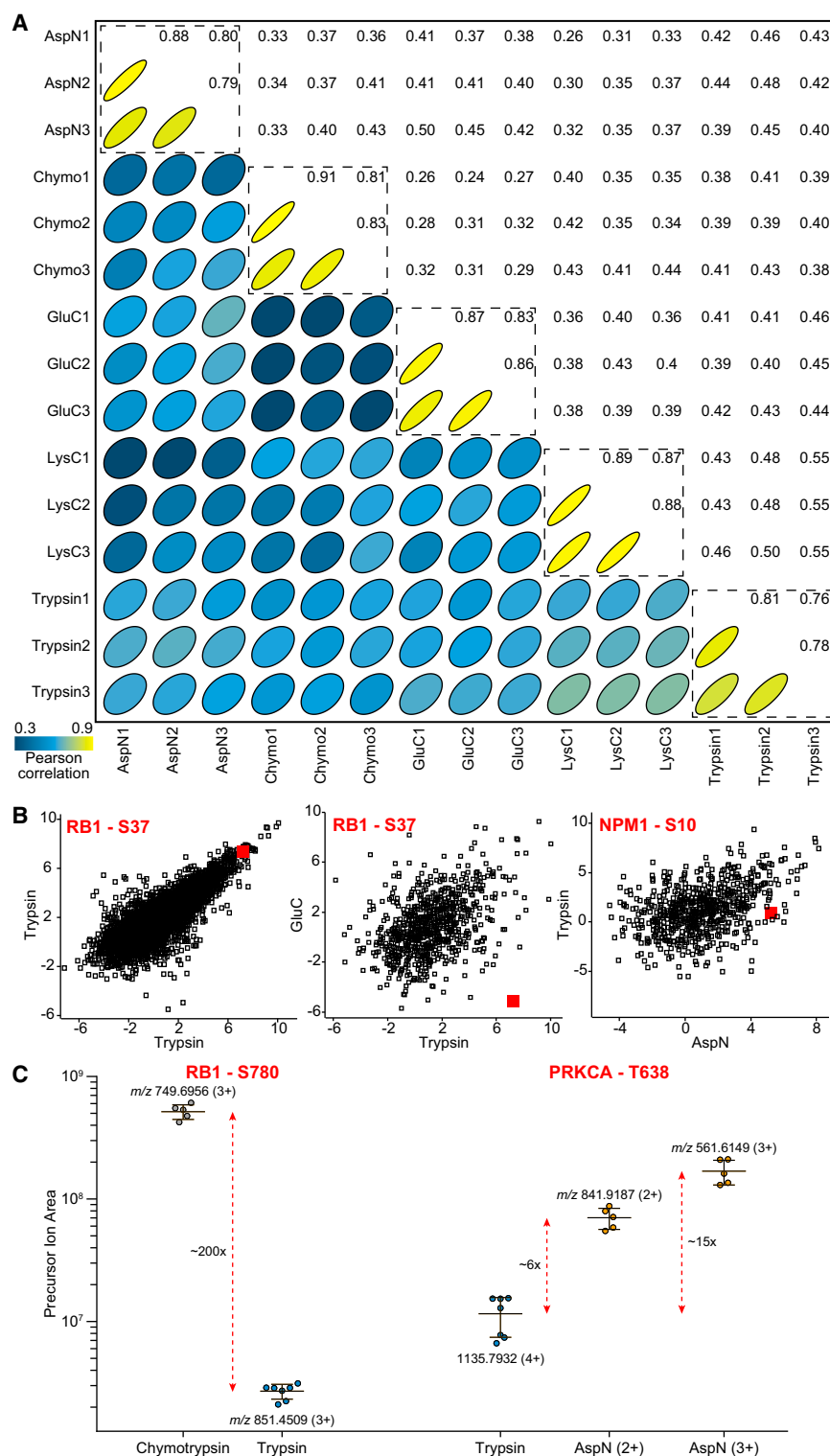
Figure 4. Label-free and Targeted Phosphoproteomics Using Complementary Proteases

(A) Pearson correlation matrix of all performed experiments reveal a high correlation in phosphosite intensity when data sets are obtained following digestion by the same protease (dashed squares; r > 0.8) but a low correlation between data sets originating from different proteases (r ~0.25–0.55).

(B) Illustrative scatter plots depicting phosphosite-bearing peptide intensities (log base 2) in digests obtained by different proteases. The plots are extracted from Figure S5. Two phosphosites are highlighted in red, showing a clear bias in intensity toward detection by trypsin (RB1-S37) or chymotrypsin (NPM1-S10) in this label-free intensity-based analysis.

(C) Also, a targeted phosphoproteomics experiments on specific phosphorylation sites shows a clear bias in average precursor intensity (error bars, ±SD) toward detection by chymotrypsin (RB1-S608) and AspN (PRKCA-T638).

phosphoproteome coverage, ultimately showing that this workflow enables reproducible, quantitative, and complementary phosphoproteomic analysis. As a result, we make publicly available a human phosphopeptide atlas of more than 37,771 unique phosphopeptides, correlating to over 18,000 unique phosphosites. Researchers interested in particular phosphorylation events on targeted proteins may use the here-described web tool to pick the appropriate protease that will lead to successful detection of their sites of interest and extract the necessary parameters to construct targeted assays for phosphoproteomics studies.

Our extensive resource also allows us to conclude several important technical issues, notably, (1) Ti^{4+}-IMAC phosphopeptide enrichment is equally efficient and selective independent of the protease used; (2) cumulative evidence from all protease data sets demonstrates unambiguously that protein phosphorylation reduces the cleavage efficiency near those sites for each protease used; (3) label-free quantitative phosphoproteomics is possible, however, only when comparing data generated by one protease; and (4) there is a clear bias toward trypsin in the currently available and widely used phosphopeptide and phosphosite databases.

We expect our data will be a valuable resource for researcher in the signaling and proteomics field and may assist in encouraging

of multiple proteases is beneficial for large-scale phosphoproteomics analysis, exposing many regulatory relevant phosphosites on key signaling proteins, which would be occluded when only using trypsin. Notably, we also substantially increased the

people to perform, next to trypsin-based experiments, additional experiments with one or more complementary proteases.

EXPERIMENTAL PROCEDURES

Sample Preparation and MS Analysis

Jurkat T lymphoma cells were resuspended at a final concentration of 1 or 2×10^6 cells/ml with 10 μM PGE$_2$ in RPMI and incubated for 10 min. After treatment, Jurkat cells were lysed in 50 mM ammonium bicarbonate (pH 8.0), 8 M urea, 1 mM sodium orthovanadate, complete EDTA-free protease inhibitor mixture, and phosSTOP phosphatase inhibitor mixture (both Roche) lysis buffer. Digested proteins were subjected to phosphopeptide enrichment using Ti4+-IMAC beads (Zhou et al., 2013), and enriched phosphopeptides were identified on a LTQ-Orbitrap Elite (Thermo Scientific) or Orbitrap Fusion (Thermo Scientific) using a decision-tree-based ion trap CID or ETD fragmentation. Further details are given in the Supplemental Experimental Procedures.

Data Analysis

The MS data were processed using Proteome Discoverer 1.4 (Thermo Scientific) and searched with the MS-GF+ search tool (Kim and Pevzner, 2014; Kim et al., 2010) against a Swissprot *Homo sapiens* database. The phosphorylation site localization of the identified phosphopeptides was performed using the phosphoRS algorithm 3.1 (Taus et al., 2011). For label-free analysis, raw data were processed with MaxQuant version 1.3.0.5 (Cox and Mann, 2008). For inclusion-list-based experiments, the data analysis was done manually. Extracted ion chromatograms (XICs) for a given parent mass were extracted with Xcalibur 3.0.63 (Thermo Scientific) with a mass tolerance of ±20 ppm. To evaluate the detectability of a given phosphopeptide, we made use of a spectral counting score (SCS) (Old et al., 2005; Zhang et al., 2006). The SCS for a given phosphorylation site was calculated as follows: (1) the amount of PSMs of that phosphosite in a given protease data set was divided by the total spectral counts obtained by that protease and then (2) the obtained value was normalized to 100% on the sum of the five values obtained from each protease. Further details are given in the Supplemental Experimental Procedures.

ACCESSION NUMBERS

The MS proteomics data have been deposited in the ProteomeXchange Consortium via the PRIDE partner repository (Vizcaíno et al., 2013) with the data set identifier PXD001428.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, six figures, and three tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2015.05.029.

AUTHOR CONTRIBUTIONS

P.G. and A.J.R.H. conceived the idea for this study. P.G. and T.T.A. performed the phosphoproteomics experiments. P.G analyzed and interpreted the data supported by T.T.A., M.P., B.v.B., and A.J.R.H. H.v.d.T. and B.v.B. developed the tools for phosphosite localization by the phosphoRS algorithm and built the web-based phosphopeptide atlas depository (http://phosphodb.hecklab.com/). P.G. and A.J.R.H. wrote the manuscript. All authors read and approved the manuscript.

ACKNOWLEDGMENTS

REFERENCES

Amanchy, R., Periaswamy, B., Mathivanan, S., Reddy, R., Tattikota, S.G., and Pandey, A. (2007). A curated compendium of phosphorylation motifs. Nat. Biotechnol. *25*, 285–286.

Bian, Y., Ye, M., Song, C., Cheng, K., Wang, C., Wei, X., Zhu, J., Chen, R., Wang, F., and Zou, H. (2012). Improve the coverage for the analysis of phosphoproteome of HeLa cells by a tandem digestion approach. J. Proteome Res. *11*, 2828–2837.

Biringer, R.G., Amato, H., Harrington, M.G., Fonteh, A.N., Riggins, J.N., and Hühmer, A.F. (2006). Enhanced sequence coverage of proteins in human cerebrospinal fluid using multiple enzymatic digestion and linear ion trap LC-MS/MS. Brief. Funct. Genomics Proteomics *5*, 144–153.

Chong, P.K., Lee, H., Kong, J.W., Loh, M.C., Wong, C.H., and Lim, Y.P. (2008). Phosphoproteomics, oncogenic signaling and cancer research. Proteomics *8*, 4370–4382.

Courcelles, M., Frémin, C., Voisin, L., Lemieux, S., Meloche, S., and Thibault, P. (2013). Phosphoproteome dynamics reveal novel ERK1/2 MAP kinase substrates with broad spectrum of functions. Mol. Syst. Biol. *9*, 669.

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.

de Graaf, E.L., Giansanti, P., Altelaar, A.F.M., and Heck, A.J.R. (2014). Single-step enrichment by Ti4+-IMAC and label-free quantitation enables in-depth monitoring of phosphorylation dynamics with high reproducibility and temporal resolution. Mol. Cell. Proteomics *13*, 2426–2434.

Di Palma, S., Hennrich, M.L., Heck, A.J., and Mohammed, S. (2012). Recent advances in peptide separation by multidimensional liquid chromatography for proteome analysis. J. Proteomics *75*, 3791–3813.

Dickhut, C., Feldmann, I., Lambert, J., and Zahedi, R.P. (2014). Impact of digestion conditions on phosphoproteomics. J. Proteome Res. *13*, 2761–2770.

Falini, B., Nicoletti, I., Bolli, N., Martelli, M.P., Liso, A., Gorello, P., Mandelli, F., Mecucci, C., and Martelli, M.F. (2007). Translocations and mutations involving the nucleophosmin (NPM1) gene in lymphomas and leukemias. Haematologica *92*, 519–532.

Frese, C.K., Altelaar, A.F.M., Hennrich, M.L., Nolting, D., Zeller, M., Griep-Raming, J., Heck, A.J.R., and Mohammed, S. (2011). Improved peptide identification by targeted fragmentation using CID, HCD and ETD on an LTQ-Orbitrap Velos. J. Proteome Res. *10*, 2377–2388.

Gauci, S., Helbig, A.O., Slijper, M., Krijgsveld, J., Heck, A.J.R., and Mohammed, S. (2009). Lys-N and trypsin cover complementary parts of the phosphoproteome in a refined SCX-based approach. Anal. Chem. *81*, 4493–4501.

Gershon, P.D. (2014). Cleaved and missed sites for trypsin, lys-C, and lys-N can be predicted with high confidence on the basis of sequence context. J. Proteome Res. *13*, 702–709.

Granholm, V., Kim, S., Navarro, J.C.F., Sjölund, E., Smith, R.D., and Käll, L. (2014). Fast and accurate database searches with MS-GF+Percolator. J. Proteome Res. *13*, 890–897.

Guo, X., Trudgian, D.C., Lemoff, A., Yadavalli, S., and Mirzaei, H. (2014). Confetti: a multiprotease map of the HeLa proteome for comprehensive proteomics. Mol. Cell. Proteomics *13*, 1573–1584.

Hebert, A.S., Richards, A.L., Bailey, D.J., Ulbrich, A., Coughlin, E.E., Westphall, M.S., and Coon, J.J. (2014). The one hour yeast proteome. Mol. Cell. Proteomics *13*, 339–347.

Horn, H., Schoof, E.M., Kim, J., Robin, X., Miller, M.L., Diella, F., Palma, A., Cesareni, G., Jensen, L.J., and Linding, R. (2014). KinomeXplorer: an integrated platform for kinome biology studies. Nat. Methods *11*, 603–604.

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.

Huang, P.H., and White, F.M. (2008). Phosphoproteomics: unraveling the signaling web. Mol. Cell 31, 777–781.

Jaffe, J.D., Keshishian, H., Chang, B., Addona, T.A., Gillette, M.A., and Carr, S.A. (2008). Accurate inclusion mass screening: a bridge from unbiased discovery to targeted assay development for biomarker verification. Mol. Cell. Proteomics 7, 1952–1962.

Kalli, A., and Håkansson, K. (2010). Electron capture dissociation of highly charged proteolytic peptides from Lys N, Lys C and Glu C digestion. Mol. Biosyst. 6, 1668–1681.

Kim, S., and Pevzner, P.A. (2014). MS-GF+ makes progress towards a universal database search tool for proteomics. Nat. Commun. 5, 5277.

Kim, S., Mischerikow, N., Bandeira, N., Navarro, J.D., Wich, L., Mohammed, S., Heck, A.J., and Pevzner, P.A. (2010). The generating function of CID, ETD, and CID/ETD pairs of tandem mass spectra: applications to database search. Mol. Cell. Proteomics 9, 2840–2852.

Knudsen, E.S., and Wang, J.Y. (1997). Dual mechanisms for the inhibition of E2F binding to RB by cyclin-dependent kinase-mediated RB phosphorylation. Mol. Cell. Biol. 17, 5771–5783.

Low, T.Y., van Heesch, S., van den Toorn, H., Giansanti, P., Cristobal, A., Toonen, P., Schafer, S., Hübner, N., van Breukelen, B., Mohammed, S., et al. (2013). Quantitative and qualitative proteome characteristics extracted from in-depth integrated genomics and proteomics analysis. Cell Rep. 5, 1469–1478.

Lu, K.P., Liou, Y.C., and Zhou, X.Z. (2002). Pinning down proline-directed phosphorylation signaling. Trends Cell Biol. 12, 164–172.

Macek, B., Mann, M., and Olsen, J.V. (2009). Global and site-specific quantitative phosphoproteomics: principles and applications. Annu. Rev. Pharmacol. Toxicol. 49, 199–221.

Mallick, P., Schirle, M., Chen, S.S., Flory, M.R., Lee, H., Martin, D., Ranish, J., Raught, B., Schmitt, R., Werner, T., et al. (2007). Computational prediction of proteotypic peptides for quantitative proteomics. Nat. Biotechnol. 25, 125–131.

Meyer, J.G., Kim, S., Maltby, D.A., Ghassemian, M., Bandeira, N., and Komives, E.A. (2014). Expanding proteome coverage with orthogonal-specificity α-lytic proteases. Mol. Cell. Proteomics 13, 823–835.

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, M111.011015.

Michalski, A., Damoc, E., Lange, O., Denisov, E., Nolting, D., Muller, M., Viner, R., Schwartz, J., Remes, P., Belford, M., et al. (2012). Ultra high resolution linear ion trap Orbitrap mass spectrometer (Orbitrap Elite) facilitates top down LC MS/MS and versatile peptide fragmentation modes. Mol. Cell. Proteomics 11, O111.013698.

Molina, H., Horn, D.M., Tang, N., Mathivanan, S., and Pandey, A. (2007). Global proteomic profiling of phosphopeptides using electron transfer dissociation tandem mass spectrometry. Proc. Natl. Acad. Sci. USA 104, 2199–2204.

Montoya, A., Beltran, L., Casado, P., Rodríguez-Prados, J.C., and Cutillas, P.R. (2011). Characterization of a TiO₂ enrichment method for label-free quantitative phosphoproteomics. Methods 54, 370–378.

Murphree, A.L., and Benedict, W.F. (1984). Retinoblastoma: clues to human oncogenesis. Science 223, 1028–1033.

Old, W.M., Meyer-Arendt, K., Aveline-Wolf, L., Pierce, K.G., Mendoza, A., Sevinsky, J.R., Resing, K.A., and Ahn, N.G. (2005). Comparison of label-free methods for quantifying human proteins by shotgun proteomics. Mol. Cell. Proteomics 4, 1487–1502.

Pearce, L.R., Komander, D., and Alessi, D.R. (2010). The nuts and bolts of AGC protein kinases. Nat. Rev. Mol. Cell Biol. 11, 9–22.

Peng, M., Taouatas, N., Cappadona, S., van Breukelen, B., Mohammed, S., Scholten, A., and Heck, A.J. (2012). Protease bias in absolute protein quantitation. Nat. Methods 9, 524–525.

Pitcher, J.A., Freedman, N.J., and Lefkowitz, R.J. (1998). G protein-coupled receptor kinases. Annu. Rev. Biochem. 67, 653–692.

Rigbolt, K.T., and Blagoev, B. (2012). Quantitative phosphoproteomics to characterize signaling networks. Semin. Cell Dev. Biol. 23, 863–871.

Ruprecht, B., and Lemeer, S. (2014). Proteomic analysis of phosphorylation in cancer. Expert Rev. Proteomics 11, 259–267.

Santamaría, E., Mora, M.I., Muñoz, J., Sánchez-Quiles, V., Fernández-Irigoyen, J., Prieto, J., and Corrales, F.J. (2009). Regulation of stathmin phosphorylation in mouse liver progenitor-29 cells during proteasome inhibition. Proteomics 9, 4495–4506.

Schmidt, A., Gehlenborg, N., Bodenmiller, B., Mueller, L.N., Campbell, D., Mueller, M., Aebersold, R., and Domon, B. (2008). An integrated, directed mass spectrometric approach for in-depth characterization of complex peptide mixtures. Mol. Cell. Proteomics 7, 2138–2150.

Schwartz, D., and Gygi, S.P. (2005). An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. Nat. Biotechnol. 23, 1391–1398.

Sharma, K., D'Souza, R.C., Tyanova, S., Schaab, C., Wiśniewski, J.R., Cox, J., and Mann, M. (2014). Ultradeep human phosphoproteome reveals a distinct regulatory nature of Tyr and Ser/Thr-based signaling. Cell Rep. 8, 1583–1594.

Soderblom, E.J., Philipp, M., Thompson, J.W., Caron, M.G., and Moseley, M.A. (2011). Quantitative label-free phosphoproteomics strategy for multifaceted experimental designs. Anal. Chem. 83, 3758–3764.

Swaney, D.L., McAlister, G.C., and Coon, J.J. (2008). Decision tree-driven tandem mass spectrometry for shotgun proteomics. Nat. Methods 5, 959–964.

Swaney, D.L., Wenger, C.D., and Coon, J.J. (2010). Value of using multiple proteases for large-scale mass spectrometry-based proteomics. J. Proteome Res. 9, 1323–1329.

Taus, T., Köcher, T., Pichler, P., Paschke, C., Schmidt, A., Henrich, C., and Mechtler, K. (2011). Universal and confident phosphorylation site localization using phosphoRS. J. Proteome Res. 10, 5354–5362.

Tsiatsiani, L., and Heck, A.J.R. (2015). Proteomics beyond trypsin. FEBS J., Published online March 30, 2015. http://dx.doi.org/10.1111/febs.13287.

Vizcaíno, J.A., Côté, R.G., Csordas, A., Dianes, J.A., Fabregat, A., Foster, J.M., Griss, J., Alpi, E., Birim, M., Contell, J., et al. (2013). The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013. Nucleic Acids Res. 41, D1063–D1069.

Wilhelm, M., Schlegl, J., Hahne, H., Moghaddas Gholami, A., Lieberenz, M., Savitski, M.M., Ziegler, E., Butzmann, L., Gessulat, S., Marx, H., et al. (2014). Mass-spectrometry-based draft of the human proteome. Nature 509, 582–587.

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

Yates, J.R., 3rd, Mohammed, S., and Heck, A.J. (2014). Phosphoproteomics. Anal. Chem. 86, 1313.

Zhang, B., VerBerkmoes, N.C., Langston, M.A., Uberbacher, E., Hettich, R.L., and Samatova, N.F. (2006). Detecting differential and correlated protein expression in label-free shotgun proteomics. J. Proteome Res. 5, 2909–2918.

Zhou, H., Ye, M., Dong, J., Corradini, E., Cristobal, A., Heck, A.J.R., Zou, H., and Mohammed, S. (2013). Robust phosphoproteome enrichment using monodisperse microsphere-based immobilized titanium (IV) ion affinity chromatography. Nat. Protoc. 8, 461–480.