MCP Papers in Press. Published on September 21, 2009 as Manuscript M900291-MCP200

In depth qualitative and quantitative profiling of tyrosine phosphorylation using a combination of phosphopeptide immuno-affinity purification and stable isotope dimethyl labeling

Paul J. Boersema[1,2,*], Leong Yan Foong[3,*], Vanessa M.Y. Ding[3,*], Simone Lemeer[1,2,*,†], Bas van Breukelen[1,2], Robin Philp[3], Jos Boekhorst[4], Berend Snel[4], Jeroen den Hertog[5,6], Andre B. H. Choo[3,7#] and Albert J. R. Heck[1,2,8#]

[1] Biomolecular Mass Spectrometry and Proteomics Group, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH Utrecht, the Netherlands

[2] Netherlands Proteomics Centre

[3] Bioprocessing Technology Institute, A*STAR (Agency for Science, Technology and Research), 20 Biopolis Way #06 - 01 Centros, Singapore, 138668, Singapore

[4] Bioinformatics, Department of Biology, Faculty of Science, Utecht University, Padualaan 8, 3584 CH Utrecht, the Netherlands

[5] Hubrecht Institute – KNAW & University Medical Center Utrecht, the Netherlands.

[6] Institute of Biology Leiden, Leiden University, the Netherlands

[7] Division of Bioengineering, Faculty of Engineering, National University of Singapore, Singapore

[8] Centre for Biomedical Genetics, the Netherlands

* These authors authors contributed equally to this work and should be considered as joint first authors

† Current address: Lehrstuhl für Bioanalytik, Technische Universität München, An der Saatzucht 5, 85354 Freising, Germany.

To whom correspondence may be addressed: Albert J.R. Heck, Biomolecular Mass Spectrometry and Proteomics Group, Utrecht University, Padualaan 8, 3584 CH Utrecht, the Netherlands, a.j.r.heck@uu.nl, tel. +31 30 2536797, fax.+31 30 2536919 and Andre B.H. Choo, Stem Cell Group, Bioprocessing Technology Institute, A*STAR (Agency for Science, Technology and Research), 20 Biopolis Way #06 - 01 Centros, Singapore, 138668, Singapore, andre_choo@bti.a-star.edu.sg, tel.: +65-64788856, fax: +65-64789561.

Running title: Quantitative phosphotyrosine profiling by dimethyl labeling

Abbreviations

EGFR – EGF receptor

FDR – false discovery rate

HTP – high throughput

iTRAQ - isobaric tag for relative and absolute quantitation

LTP – low throughput

NSCLC – non-small cell lung cancer

Pfam – protein family

PRIDE – Proteomics Identifications

PTK – protein tyrosine kinase

SILAC – Stable isotope labeling with amino acids in cell culture

XIC – extracted ion chromatogram

Summary

Several mass spectrometry based assays have emerged for the quantitative profiling of cellular tyrosine phosphorylation. Ideally, these methods should reveal the exact sites of tyrosine phosphorylation, be quantitative and not cost-prohibitive. The latter is often an issue as typically several milligrams of (stable isotope labeled) starting protein material is required to enable the detection of low abundant phosphotyrosine peptides. Here, we adopted and refined a peptide centric immuno-affinity purification approach for the quantitative analysis of tyrosine phosphorylation by combining it with a cost-effective stable isotope dimethyl labeling method.

We were able to identify by mass spectrometry, using just two LC-MS/MS runs, more than 1100 unique non-redundant phosphopeptides in HeLa cells from about 4 mg of starting material without requiring any further affinity enrichment as close to 80% of the identified peptides were tyrosine phosphorylated peptides. Stable isotope dimethyl labeling could be incorporated prior to the immuno-affinity purification, even for the used large quantities (mg) of peptide material, enabling the quantification of differences in tyrosine phosphorylation upon pervanadate treatment or EGF stimulation. Analysis of the EGF-stimulated HeLa cells, a frequently used model system for tyrosine phosphorylation, resulted in the quantification of 73 regulated unique phosphotyrosine peptides. The quantitative data was found to be exceptionally consistent with the literature, evidencing that such a targeted quantitative phosphoproteomics approach can provide reproducible results. In general, the combination of immuno-affinity purification of tyrosine phosphorylated peptides with large scale stable isotope dimethyl labeling provides a cost-effective approach that can alleviate variation in sample preparation and analysis as samples can be combined early on. Using this approach a rather complete qualitative and quantitative picture of tyrosine phosphorylation signaling events can be generated.

Introduction

Reversible tyrosine phosphorylation plays an important role in numerous cellular processes like growth, differentiation and migration. Phosphotyrosine signaling is tightly controlled by the balanced action of protein tyrosine kinases (PTKs) and protein tyrosine phosphatases. Aberrant tyrosine phosphorylation has been suggested to be an underlying course in multiple cancers (1). Therefore, the identification of tyrosine phosphorylated proteins and the investigation into their involvement in signaling pathways is important. Several groups have attempted to comprehensively study tyrosine phosphorylation by proteomics means (2-5). However, large scale identification of tyrosine phosphorylation sites by mass spectrometry (MS) can be hindered by the low abundance of tyrosine phosphorylated proteins. Especially, signaling intermediates are usually low abundant proteins that show substoichiometric phosphorylation levels. In addition, the identification by mass spectrometry of phosphopeptides from a complex cellular lysate digest is often complicated by ion suppression effects due to a high background of non-phosphorylated peptides. Enrichment of tyrosine phosphorylated proteins or peptides prior to mass spectrometric detection is therefore essential. Traditionally, antibodies against phosphorylated tyrosine have been used to immunoprecipitate tyrosine phosphorylated proteins from cultured cells (2-4, 6-8). This phosphoprotein immuno-affinity purification method has for example been used to study the global dynamics of phosphotyrosine signaling events after EGF stimulation using stable isotope labeling with amino acids in cell culture (SILAC)(2). This approach led to the identification of known and previously unidentified signaling proteins in the EGF receptor (EGFR) pathway, including their temporal activation profile after stimulation of the EGFR, providing crucial information for modeling signaling events in the cell. However, as the identification and quantification of these phosphorylated proteins in these studies was not necessarily based on tyrosine phosphorylated peptides, but largely on non-phosphorylated

peptides, little information is derived on the exact site(s) of tyrosine phosphorylation. Also, binding partners of tyrosine phosphorylated proteins, that itself are not tyrosine phosphorylated, might be co-precipitated and impair the tyrosine phosphorylation quantification. Immuno-affinity purification of phosphotyrosine peptides -rather than proteins- using anti-phosphotyrosine antibodies (5, 9-16) significantly facilitates the identification of the site(s) of phosphorylation as it greatly alleviates most of the above mentioned problems since the tyrosine phosphorylated site can be directly identified and quantified.

Accurate MS-based quantification is typically performed by stable isotope labeling. The isotopes can be incorporated metabolically during cell culture as in stable isotope labeling by amino acids in cell culture (SILAC)(17) or chemically as in isobaric tag for relative and absolute quantitation (iTRAQ)(18) or stable isotope dimethyl labeling (19-21). Typically, most precise quantification can be obtained by metabolic labeling as the different samples can be combined at the level of intact cells (22). However, metabolic labeling is somewhat limited to biological systems that can be grown in culture and the medium may have an effect on the growth and development of the cells. iTRAQ has been used in conjunction with phosphotyrosine peptide immunoprecipitation (5). As the chemical labeling is performed before immunoprecipitation, the differentially labeled samples can be precipitated together, thereby neutralizing the potentially largest source of variation. However, as this phosphotyrosine peptide immunoprecipitation is typically performed on several hundreds of micrograms to milligrams of protein sample, iTRAQ provides in these cases a rather cost prohibitive means.

Here, we present an optimized immuno-affinity purification approach for the analysis of tyrosine phosphorylation and combine it with stable isotope dimethyl labeling (19-21, 23).

We efficiently enriched and identified by MS 1112 unique phosphopeptides derived from 4 mg of starting protein material without any further affinity chromatographic enrichment, whereby up to 80% of the peptides analyzed in the final LC run were phosphotyrosine peptides. We further advanced the method by introducing triplex stable isotope dimethyl labeling prior to immunoprecipitation. We quantified differences in tyrosine phosphorylation upon pervanadate treatment or EGF stimulation to detect site-specific changes in tyrosine phosphorylation. 128 unique phosphotyrosine peptides were identified and quantified upon pervanadate treatment. By employing an internal standard comprising of both mock and pervanadate treated sample, we could more confidently identify and quantify phosphorylation sites that are strongly regulated and on-off situations. Analysis of EGF-stimulated HeLa cells resulted in the quantification of 73 unique phosphotyrosine peptides. Most of the upregulated phosphotyrosine peptides that were identified have previously been reported to be involved in the EGFR signaling pathway, validating our approach. However, for the first time we found TFG to become also highly tyrosine phosphorylated upon EGF stimulation together with some tyrosine phosphorylation sites on for example IRS2, SgK269 and DLG3 that have not been firmly established earlier to be involved in EGFR signaling.

In general, we show that the combination of immuno-affinity purification of tyrosine phosphorylated peptides with large scale chemical stable isotope dimethyl labeling provides a cost-effective approach that can alleviate variation in immunoprecipitation and LC-MS as samples can be combined before immunoprecipitation and the necessity of performing additional enrichment is removed by an optimization of the protocol. With only a single LC-MS run, already a rather complete qualitative and quantitative picture of a signaling event can be generated.

Materials and Methods

Cell culture, stimulation, digest preparation

HeLa cells were grown to confluence in Dulbecco's modified Eagle's medium containing 10% fetal bovine serum (Invitrogen) and 0.05 mg/ml penicillin/streptomycin (Invitrogen). Cells were placed in serum-free medium 16 h before EGF stimulation. Cells were stimulated with 150 ng/ml EGF for 10 or 30 min, 1mM Pervanadate (prepared by incubating 1 mM orthovanadate with 1 mM hydrogen peroxide for several min) for 10 min or left untreated. Cells were washed with cold phosphate buffered saline and lysed. Before labeling and immunoprecipitation, cells were lysed in 8 M Urea/ 50 mM ammoniumbicarbonate, 5 mM sodium phosphate, 1 mM potassium fluoride, 1 mM sodium orthovanadate and EDTA-free protease inhibitor cocktail (Sigma). Samples were reduced with DTT at a final concentration of 10 mM at 56 ^{o}C, subsequently samples were alkylated with iodoacetamide at a final concentration of 55 mM at RT. The samples were diluted to 2 M Urea/50 mM ammoniumbicarbonate and trypsin (1:100; Promega) was added. Digestion was performed overnight at 37 ^{o}C.

Stable isotope labeling by reductive amination of tryptic peptides

Tryptic peptides were desalted, dried *in-vacuo* and re-suspended in 100 µL of triethylammonium bicarbonate (100 mM). Subsequently, formaldehyde-H_2 (573 µmol) was added, vortexed for 2 min followed by the addition of freshly prepared sodium cyanoborohydride (278 µmol). The resultant mixture was vortexed for 60 min at RT. A total of 60 µl ammonia (25%) was added to consume the excess formaldehyde. Finally, 50 µl formic acid (100%) was added to acidify the solution. For intermediate labels, formaldehyde-D_2 (573 µmol) was used. For the heavy labeling, ^{13}C-D_2-formaldehyde (573 µmol) and freshly prepared cyanoborodeuteride (278 µmol) was used (20, 21). The light, intermediate and

heavy dimethyl labeled samples were mixed in 1:1:1 ratio based on total peptide amount, determined by running an aliquot of the labeled samples on a regular LC-MS run and comparing overall peptide signal intensities.

Immunoprecipitation

Labeled peptides were mixed, desalted, dried down and re-dissolved in immunoprecipitation (IP) buffer (50mM Tris pH 7.4, 150 mM NaCl, 1% n-octyl-beta-d-glucopyranoside and 1x Complete mini (Roche diagnostics)). Prior to IP, pY99 agarose beads (Santa Cruz) were washed in IP buffer. The labeled peptide mixture was added to the pY99 agarose beads and incubation was performed overnight at 4^{o}C. Beads were washed 3 times with IP buffer and 2 times with water. Peptides were eluted by adding 0.15% TFA for 20 min at RT. Eluted peptides were desalted and concentrated on STAGE-tips.

On-line nanoflow liquid chromatography

Nanoflow LC-MS/MS was performed by coupling an Agilent 1100 HPLC system (Agilent Technologies, Waldbronn, Germany) to a LTQ-Orbitrap mass spectrometer (Thermo Electron, Bremen, Germany) as described previously (24). Dried fractions were reconstituted in 10 µL 0.1 M acetic acid and delivered to a trap column (Aqua^{tm} C18, 5 µm, (Phenomenex, Torrance, CA, USA); 20 mm × 100 µm ID, packed in-house) at 5 µL/min in 100% solvent A (0.1 M acetic acid in water). Subsequently, peptides were transferred to an analytical column (ReproSil-Pur C18-AQ, 3 µm, Dr. Maisch GmbH, Ammerbuch, Germany; 40 cm × 50 µm ID, packed in-house) at ~100 nL/min in a 2 or 3 hour gradient from 0 to 40% solvent B (0.1 M acetic acid in 8/2 (v/v) acetonitrile/water). The eluent was sprayed via distal coated emitter tips (New Objective), butt-connected to the analytical column. The mass spectrometer was operated in data dependent mode, automatically switching between MS and MS/MS. Full

scan MS spectra (from *m/z* 300-1500) were acquired in the Orbitrap with a resolution of 60,000 at *m/z* 400 after accumulation to target value of 500,000. The three most intense ions at a threshold above 5000 were selected for collision-induced fragmentation in the linear ion trap at a normalized collision energy of 35% after accumulation to a target value of 10,000.

Data analysis

All MS^2 spectra were converted to single DTA files using Bioworks 3.3 using default settings. Runs were searched using an in-house licensed MASCOT search engine (Mascot (version 2.1.0) software platform (Matrix Science, London, UK)) against the Human IPI database version 3.36 (63012 sequences) with carbamidomethyl cysteine as a fixed modification. Light, intermediate and heavy dimethylation of peptide N-termini and lysine residues, oxidized methionine and phosphorylation of tyrosine, serine and threonine were set as variable modifications. Trypsin was specified as the proteolytic enzyme and up to two missed cleavages were allowed. The mass tolerance of the precursor ion was set to 5 ppm and for fragment ions 0.6 Da. Peptides were assigned to the first protein hit reported by Mascot. The assignment of phosphorylation sites of identified phosphopeptides was performed by the PTM scoring algorithm implemented in MSQuant as described previously (25). Individual MS/MS spectra from phosphopeptides were accepted for a Mascot score ≥20. This threshold was experimentally tested (see Results section) and the FDR at this score for phosphotyrosine peptides only was estimated to be 2% by performing a concatenated decoy database search. All identified phosphopeptides that were found to be differentially phosphorylated were manually validated.

Quantification

Quantification of peptide triplets of which at least one has obtained a Mascot peptide score of 20 was performed using an in-house dimethyl-adapted version of MSQuant (http://msquant.sourceforge.net/), as described previously (20). Briefly, peptide ratios were obtained by calculating the extracted ion chromatograms (XIC) of the "light", "intermediate" and "heavy" forms of the peptide using the monoisotopic peaks only. The total XIC for each of the peptide forms was obtained by summing the XIC in consecutive MS cycles for the duration of their respective LC-MS peaks in the total ion chromatogram using FT-MS scans. This total XIC was then used to compute the peptide ratio. Heavy and light labeled peptides were found to largely co-elute. Quantified proteins were normalized against the Log_2 of the median of all peptides quantified. StatQuant, an in-house developed program (26), was used for normalization, outlier detection and determination of standard deviation. Ratios of phosphotyrosine levels were normalized to the ratios of (non-specifically binding) non-phosphorylated peptides. Ratios derived from different charge states of the peptide and/or missed cleavages with the same phosphorylation sites were log averaged. Clustering of the phosphotyrosine profiles upon EGF stimulation was performed by K-means clustering (Eucledian distance).

Datasets

The HeLa phosphoproteome was compared to other datasets by mapping protein identifiers and phosphosite locations from Phospho.ELM (version 8.2) and Rikova *et al*. (10) to IPI human version 3.36. Only sites that could be mapped unambiguously to a single IPI identifier were included, resulting in 1405 and 3955 mapped phosphotyrosines, respectively. Overlap was determined by counting the number of identical sites in the different combinations of datasets.

Results

To first evaluate our protocols and to chart the potential arsenal of phosphorylation sites in HeLa cells, we performed immuno-affinity enrichment of tyrosine phosphorylated peptides from HeLa cell lysate digests in which hyper-phosphorylation was induced via inhibition of phosphatases by pervanadate. To allow for a comprehensive coverage of tyrosine phosphorylation, HeLa cells that were stimulated with pervanadate or mock treated were lysed and digested by trypsin and mixed 1:1, followed by peptide immuno-affinity purification of 4 mg total starting protein material by an antiphosphotyrosine antibody immobilized on agarose beads. The peptides that eluted from the immuno-affinity resin were then analyzed without any further enrichment in two separate LC-MS runs using a two and a three hour LC gradient, respectively. Peptides were identified by matching the fragmentation spectra against the IPI human database (3.36). In Figure 1A, a part of the base peak chromatogram of the 3 hour run is shown to illustrate the efficiency and specificity of the immunoprecipitation. The resolved peaks nearly all represent tyrosine phosphorylated peptides, while the few non-phosphorylated peptides of abundant proteins that were detected are largely masked by these phosphopeptides. In Figure 1B, the apparent specificity of the immunoprecipitation and the number of identified tyrosine phosphorylated peptides with decreasing Mascot threshold score is plotted. By decreasing the Mascot threshold score the number of identified redundant tyrosine phosphorylated peptides increases considerably. However, by decreasing the threshold the chance of including false positives is also increased substantially (27). This is reflected in the number of identified non-tyrosine phosphorylated peptides which increases exponentially below a score of 20. Consequently, the apparent immunoprecipitation specificity remains above 75% from threshold score 45 to 20 and then suddenly drops. A Mascot score of 20 was therefore taken as a threshold score as this lies above the apparent inflection point. At this score the false discovery rate (FDR) as calculated

by performing a concatenated decoy database search is seemingly large, approximately 9%. However, when only tyrosine phosphorylated peptides are taken into account the FDR is only 2%. Tandem mass spectra are available in the PRIDE (Proteomics Identifications) database under accession number 9779. To create a unique, non-redundant phosphotyrosine peptide library, we filtered the identified peptides. Phosphorylation sites that were also identified with a methionine oxidation or a miss-cleavage were considered redundant, but different states of additional phosphorylation on serine, threonine and/or tyrosine were considered unique. This led to the identification of 729 unique phosphotyrosine peptides in the 2 hour run and 970 unique phosphotyrosine peptides in the 3 hour run. The overlap between these two runs was very large (Supplemental Figure 1A), leading to the identification of 1112 unique phosphotyrosine peptides with a total of 983 unique tyrosine phosphorylation sites. A list of these identified phosphopeptides including the sites of phosphorylation and Mascot scores are listed in Supplementary Table 2. This is one of the largest experimental datasets of phosphotyrosine peptides detected in a single experiment without any further TiO_2 or IMAC enrichment. We compared our dataset with phosphotyrosine peptides and sites reported in the recent update of the Phospho.ELM (28) database and with a recently reported large phosphotyrosine peptide dataset of lung cancer cell lines consisting of the phosphotyrosine peptides cumulatively identified from 20 different cell lines and 30 different tissues samples (10). Approximately 28% of the phosphotyrosine sites that were identified in the current study overlapped with Rikova *et al.*(10) (Supplemental Figure 1B). The Phospho.ELM database reports whether a phosphorylation site is identified by small scale analysis (low throughput, LTP) or large scale, typically LC-MS based, analysis (high throughput, HTP). The overlap of our dataset with the HTP database is small with 14.3% (Supplemental Figure 1B), but still more than three times larger than the overlap with the LTP database (Supplemental Figure 1C).

The size of the dataset allowed us to do a statistically relevant motif analysis of the residues neighboring the sites of phosphorylation in our dataset to identify conserved sequence motifs that can be recognized by protein tyrosine kinases. Using motif-X (29), the occurrence of motifs from our phosphotyrosine dataset is compared with the occurrence of these motifs in the total human proteome. We found 6 highly significantly enriched motifs that could classify two-thirds of our observed phosphotyrosine sites. These motifs are displayed as WebLogos (30) in Figure 2. The motif with the highest occurrence has a serine residue at the P+1 position. Surprisingly, this motif, has not been previously reported and only one rather promiscuous kinase/phosphatase motif and three also promiscuous binding motifs are known in PhosphoMotif finder(38) with a serine at that position. The second motif contains a proline at the P+3 position. Two motifs show a negatively charged amino acid (aspartic or glutamic acid) at the P+1 position. And two already previously identified motifs as obtained from analysis in mouse brain (11), show a negatively charged amino acid at the P-3 position. To obtain an indication of the potential kinases responsible for phosphorylation of the identified phosphotyrosine peptides we used NetworKIN (31, 32). For 925 phosphosites (pY, pS and pT) we could identify the potential upstream kinase. The Insulin receptor group kinases was shown to be the upstream kinase for more than 50% of the identified phosphotyrosine peptides that were identified in our dataset, whereby the insulin receptor and insulin-like growth factor 1 receptor contribute equally. Also the EphA receptors, and in particular EphA4, seem to have a high number of substrates represented in our dataset (see Supplemental Figure 2).

Quantitative phosphoproteomics

To allow for quantification of tyrosine phosphorylation, we next set out to incorporate stable isotope dimethyl labeling (19-21) into the phosphotyrosine peptide immunoprecipitation

workflow. Therefore, HeLa cells were mock treated or treated with 1 mM pervanadate for 10 min. Cells were lysed and proteins digested. Peptides were labeled with light, intermediate or heavy dimethyl, essentially as described previously (20), but with adopted protocols to allow for the labeling of a few milligrams of sample (21). Peptides derived from the untreated cells were labeled with light dimethyl, peptides derived from a 1:1 mixture of untreated and pervanadate treated cells were labeled with intermediate dimethyl (as an internal standard) and peptides derived from the pervanadate treated cells were labeled with heavy dimethyl. Labeled peptides originating from 2 mg of protein sample for each of the three samples were mixed in a 1:1:1 ratio, and from this complex peptide mixture tyrosine phosphorylated peptides were enriched by immunoprecipitation (Figure 3). Using this approach we were able to identify and quantify from a single LC-MS run 128 unique tyrosine phosphorylated triplet peptides originating from 99 phosphoproteins. As expected, most of the detected peptides show an abundance profile whereby the ion signal from the non-stimulated phosphopeptide is very low and the signal from the intermediate labeled phosphopeptide is equal to half of the sum of the untreated peptide signal (light) and treated phosphopeptide signal (heavy) (Figure 4A). On average the heavy labeled phosphopeptides were 2.22 (±0.49) times more intense than the intermediate. A list of all quantified phosphopeptides is available as Supplementary Table 3 and tandem mass spectra are available in the PRIDE database under accession number 9780. Interestingly, 15 tyrosine phosphorylated peptides seemed not to be affected by pervanadate treatment as the abundance profile did not significantly change (Table 1 and Figure 5B). A GO term (33) and motif-X (29) analysis was performed to find within this set of unaffected phosphopeptides, and their corresponding proteins, an enrichment of certain biological processes, molecular function or conserved sequences, but the set was probably too small to detect any significant enrichment. However, an enrichment was found of phosphotyrosine sites that fall within a protein family (Pfam) domain (34) (p=2.7e-6, Fisher's

exact test). 12 of the 15 unaffected phosphotyrosine sites (80%) fall within a Pfam domain, while only 18% of the remainder of phosphotyrosine sites in the quantified database fall within such globular domains which is close to a previously reported 15-17% (35). It remains therefore elusive what is the exact cause of the apparent resistance to phosphatases at these particular phosphotyrosine sites.

With the quantitative method established, we next set out to investigate tyrosine phosphorylation mediated EGFR signaling pathways using stable isotope dimethyl labeling followed by phosphotyrosine immunoprecipitation. In our study, HeLa cells were stimulated with EGF for 0, 10 or 30 min to study the temporal phosphotyrosine signaling pathways after stimulation of the EGF receptor. We used essentially similar methods as described above for the pervanadate treated cells (Figure 3). Briefly, HeLa cells were stimulated with 150 ng/ml EGF for 0, 10 or 30 min. Cells were lysed and proteins were digested. The derived peptide mixtures were stable isotope labeled with light, intermediate or heavy dimethyl labels and mixed in a 1:1:1 ratio after which tyrosine phosphorylated peptides were immunoprecipitated using the antiphosphotyrosine antibody. 73 unique phosphotyrosine peptides, originating from 52 phosphoproteins could be differentially quantified over all three time-points after EGF stimulation in a single LC-MS analysis without any additional affinity enrichment. These associated phosphoproteins including their sites of phosphorylation are listed in Table 2 (For a full table with Mascot and PTM scores, see Supplemental Figure 4) and tandem mass spectra are available in the PRIDE database under accession number 9777. All except one of the phosphotyrosine peptides that were identified, have been reported before and can be found in the PhosphoSitePlus database (www.phosphosite.org)(36). The observed temporal ratio profiles largely clustered into three groups (Figure 5A). Cluster 1 consists of sites that show no or only a small change in phosphorylation levels. Cluster 2 consists of tyrosine

phosphorylation sites that show an immediate upregulation upon EGF stimulation that remains after 30 min. Cluster 3 comprises of phosphorylation sites that show a large and quick (10 min) increase in tyrosine phosphorylation that already diminishes at 30 min. In the latter cluster autophosphorylation sites of EGFR are found along with established direct EGFR interactors such as Gab1, Cbl and SHC1. In cluster 2, are downstream targets like p38a, stat3 and GSK3-β. Some phosphotyrosine sites were found to be regulated that have not been conclusively established to be in involved in EGFR signaling such as sites on SgK269, IRS2, HNRNPA1 and ATP1A1. The tyrosine phosphorylation on TFG has not been reported before. Finally, in cluster 1, phosphotyrosine peptides were detected from CTTN, ENO1 and WASL that showed an apparent downregulation upon EGF stimulation. Representative examples of tyrosine phosphorylated peptides from the three clusters and their temporal profiles are shown in Figure 5B-D.

Comparing the identified phosphotyrosine peptides from the three separate experiments, evidently, a large overlap can be observed between the more comprehensive dataset of 1112 phosphopeptides in the pervanadate blocked cells and the two other quantitative experiments (Supplemental Figure 3). 86% of the phosphotyrosine peptides identified in the quantitative pervanadate study overlapped with the library, while 74% of the EGF study overlapped with the library set. Also, 51% of the phosphotyrosine peptides identified in the EGF experiment overlapped with those identified in the quantitative pervanadate experiment. Ratios of these phosphopeptides were plotted against each other (Supplemental Figure 4). Some of the sites that showed an increase in tyrosine phosphorylation upon phosphatase inhibition showed no increase upon EGF stimulation. However, tyrosine phosphorylation events that showed a large increase upon EGF stimulation also showed a strong increase upon phosphatase inhibition by pervanadate. Interestingly, some of the phosphotyrosine sites that were not

affected by pervanadate treatment, such as CDC2 (pY15) and GSK3β (pY216) did increase in abundance upon EGF stimulation.

Discussion

Analysis of tyrosine phosphorylation in cells or tissue is extremely important to understand critical signaling processes involved in processes like development and human disease. In this work, we explored and further optimized an enrichment method for phosphotyrosine peptides based on immuno-affinity purification using phosphotyrosine specific antibodies. Our protocol allowed a very high level of enrichment efficiency providing an eluted fraction dominated by phosphotyrosine peptides, requiring no further IMAC or TiO_2 enrichment prior to LC-MS/MS analysis. From just two of these eluates we were able to identify 1112 unique, non-redundant phosphotyrosine peptides derived from only 4 mg of starting material. The overlap of our HeLa cells phosphotyrosine sites with the dataset reported in Rikova *et al.* (10) is 28%. Notably, the dataset was of Rikova *et al.* was obtained by analyzing several dozens of non-small cell lung cancer (NSCLC) cell lines and NSCLC tumors. The different constitution of their and our cells, together with a different antibody used for phosphotyrosine peptide enrichment, might explain the relative small overlap in tyrosine phosphopeptides, while the greater protein amount and larger number of LC-MS analyses might have allowed them for an even deeper penetration into the tyrosine phosphoproteome. The overlap between LTP and HTP phosphosites in the Phospho.ELM database is surprisingly low indicating that the tyrosine phosphoproteome has not yet been fully mapped (28). Not unexpectedly, the overlap between our dataset and the HTP dataset is larger than with the smaller LTP dataset. Apparently, a different subset of phosphosites is detected by classical biology means than by LC-MS even though datasets derived by the latter method typically contain hundreds of phosphosites. It is estimated that of the peptides that are identified in a vertebrate cell to be phosphorylated approximately 0.05-2% is phosphorylated on tyrosine residues (25, 37). If these numbers are taken as true, the number of phosphotyrosine sites identified in this study (over 1000) would suggest that there may be concurrently even up to 1,000,000 serine and

threonine phosphorylation sites present as well. Therefore, even with ever improving analytical tools it still remains a great challenge to comprehensively analyze the complete cellular phosphoproteome.

Several stable isotope labeling based quantification methods have been used in combination with phosphoproteomic approaches, including chemical labeling such as iTRAQ (4, 5, 13) and metabolic labeling such as SILAC (2, 39). The advantage of a chemical modification approach over metabolic labeling is the ability to label samples after cell lysis and digestion. This makes the approach more generically applicable as it also allows the quantitative analysis of biological samples that cannot be grown in culture such as human body fluids or tissue. The benefit of peptide level immunoprecipitation combined with stable isotope labeling is that differently labeled samples can be combined prior to immunoprecipitation, thereby neutralizing the potentially largest source of variation. Furthermore, quantification at the peptide level allows the separate analysis of phosphorylation events on the same protein. Here, we introduced stable isotope dimethyl labeling for quantification of immunoprecipitated phosphotyrosine peptides. As the starting material for these immunoprecipitation typically is several milligrams of protein, the use of iTRAQ labeling can be cost prohibitive, while stable isotope dimethyl labeling is performed with inexpensive generic reagents and thereby does not pose financial restrictions to the amount of sample to be labeled (21). Importantly, the chemical labeling does not seem to significantly affect the immunoprecipitation process itself.

We were able to identify and quantify 128 unique phosphotyrosine peptides upon pervanadate induction, all detected by their characteristic peptide triplets that can be easily distinguished in the LC-MS spectra. As expected most of the peptide ion signals of the internal standard intensity were equal to half of the sum of the untreated and treated signal intensities,

confirming that stable isotope dimethyl labeling does not alter phosphorylation states or impair immunoprecipitation. For most phosphotyrosine peptides, the detected ratios were close to the expected 0:1:2, indicating that tyrosine phosphorylation was largely enhanced upon pervanadate treatment. In Figure 4A, an example of such a phosphotyrosine peptide can be seen. No signal is detected at the m/z where the light labeled, non-pervanadate treated, phosphotyrosine peptide would reside. The intermediate labeled phosphotyrosine peptide from the internal standard is half the intensity of the heavy labeled phosphotyrosine peptide. The internal standard facilitates the detection and quantification of such an on-off situation. Without the internal standard only a single isotope envelope would be visible. The internal standard confirms the on-off situation and validates the identified sequence by the number of lysine residues that can be easily determined based on the mass shift between the intermediate and heavy labeled peptides. Surprisingly, about 15 out of 128 of the quantified peptides showed a close to 1:1:1 ratio in the three samples, indicating that these phosphorylation sites were not affected by pervanadate treatment. Pervanadate is thought to inhibit phosphatases by irreversibly oxidizing the catalytic cysteine (40). The fact that all of these phosphopeptides were selected by the mass spectrometer for sequencing indicates that these phosphopeptides are relatively abundant. These sites might therefore be constitutively phosphorylated. However, some of the sites showed an increase in phosphorylation upon EGF stimulation and, assuming that the time scale of 10 min is too low for protein synthesis, this suggests that some non-phosphorylated tyrosine residues are phosphorylated upon EGF stimulation. Explanations for the unaltered phosphorylation levels upon pervanadate treatment might involve specific phosphatases that are not inhibited by pervanadate or phosphorylation sites that are not easily accessible for these phosphates, so that inhibition does not lead to increased phosphorylation under the condition used here. The here observed enrichment of phosphotyrosine sites residing in Pfam domains, which are protein domains with a known function and tertiary

structure, for this class of peptides could be in agreement with the latter hypothesis. It has been suggested that proteins become typically phosphorylated on sites outside these functional domains (35, 41) as phosphorylation sites on the interspersing variable regions might be more exposed for kinases and phosphatases. We hypothesize that the unaffected phosphorylation sites might therefore be structurally essential or concealed in a non-accessible tertiary structural element.

The EGF receptor plays an important role in a variety of cellular processes and has been studied intensively over the years. The EGFR signaling pathway has been investigated by a variety of (phospho-)proteomic approaches (2, 4, 5, 9, 13, 25, 43, 44). Upon activation by a growth factor, the PTK activity of EGFR is enhanced (45). Our quantitative phosphotyrosine peptide immunoprecipitation, therefore, provides an excellent way to study EGFR signaling. For this reason, and as the pathway is very well documented in the literature, we took it here as a model system to further test our new approach. We analyzed HeLa cells that were stimulated for 0, 10 or 30 min by EGF. 73 unique tyrosine phosphorylated triplet peptides were identified and quantified over all time points. Most of the tyrosine phosphopeptides we identified have been reported before also in the context of EGF stimulation. For example, more than 30% of the here identified phosphosites overlap with those identified in Zhang *et al.*(5). In that report, similar EGF stimulation conditions were used and the effect was also measured after 0, 10 and 30 min. Reassuringly, the observed ratios for these overlapping phosphopeptides for 10 min and 30 min after EGF stimulation correlated very well with those found in our study (correlation coefficients 0.87 and 0.82, respectively, see Figure 6). This is especially remarkable considering different cell lines were used, HeLa *versus* 184A1 parental human mammary epithelial cells, suggesting a high conservation of the EGFR signaling pathway between these two cell lines. Moreover, it reveals that quantitative tyrosine

phosphoproteome studies can provide highly reproducible results, even between different laboratories and cell lines used, which is a status not yet readily achieved in global phosphoproteome studies. Most of the identified phosphotyrosine peptides that are increased upon EGF stimulation are known members of the EGF signaling pathway. The EGFR autophosphorylation sites pY1172 and pY1197 showed an increase in phosphorylation after 10 min EGF stimulation whereby this phosphorylation decreased again after 30 min, consistent with previously published data (5, 46). Furthermore, phosphopeptides of Gab1 (pY659), STAM2 (pY192 and pY371), RBCK1 (pY288), Cbl (pY674), SHIP-2 (pY986 and pY1135) and Epsin (pY17) showed an extensive increase in phosphorylation and have shown to be involved in EGF signaling and internalization{Tong, 2008 #5}{Zhang, 2005 #22}{Pesesse, 2001 #47}{Prasad, 2005 #48}{Wolf-Yadlin, 2007 #49}.

Some of the tyrosine phosphorylated peptides that showed an increase in abundance after EGF stimulation have not been reported before or the phosphosite has not been described profoundly to be involved in EGFR signaling. For example, here, we observed IRS2 pY260 to become more phosporylated upon EGF stimulation (47, 48). Also, TFG (pY392) shows high tyrosine phosphorylation in response to EGF treatment, but the phosphosite has not been reported before. Recently, this protein has been described to be a substrate of Src (49) like SgK269, a pseudokinase for which we found two phosphosites (pY531 and pY635) to be upregulated (50, 51). Upregulation of both these phosphotyrosine sites has not been associated with EGF stimulation before, but pY635 was suggested to be involved in Src interaction (51) and was shown to be downregulated upon drug induced EGFR inhibition (9). With these two proteins we might shed new light on a less well established Src-based part of the EGFR signaling pathway. Next, DLG3 (pY673), which was shown to be downregulated upon drug induced EGFR inhibition (9) was here, in agreement, found to be upregulated upon EGF stimulation.

Finally, only a few tyrosine phosphorylation sites were observed to be significantly downregulated upon EGF stimulation: N-WASP (pY256), ENO1 (pY44) and CTTN (pY421). Interestingly, CTTN has been shown to bind N-WASP. pY421 is one of three residues known to be phosphorylated by Src while Erk can phosphorylate CTTN at two other serine residues (52, 53). This Erk and Src phosphorylation of CTTN was suggested to act as a switch on-switch off mechanism respectively for the activation of N-WASP leading to actin polymerization which is important in for example cytokinesis (54). What exactly the outcome of the downregulation of these sites is regarding actin polymerization remains to be elucidated.

Our quantitative phosphotyrosine profiling in HeLa cells enables a direct comparison of the response of individual phosphorylation sites upon EGF stimulation and pervanadate treatment. Some sites show a strong increase after pervanadate treatment, but are seemingly unaffected by EGF stimulation, while others are increased in abundance after both EGF stimulation and pervanadate treatment. The existence of the latter category would suggest that in the pervanadate treatment conditions EGFR stimuli exist or kinases are active that target substrates that are also in the EGFR signaling pathway. This is not surprising as the fetal calf serum on which these cells are grown contain hormones and growth factors. This is reflected in the overrepresentation of insulin receptor, insulin-like growth factor receptor and ephrin receptors in the upstream kinases as predicted by NetworKIN in the larger library phosphotyrosine dataset.

In conclusion, immunoprecipitation of tyrosine phosphorylated stable isotope dimethyl labeled peptides allows the quantitative analysis of tyrosine phosphorylation. As isotope

labeling is performed after cell lysis and enzymatic digestion, the method is applicable to virtually any sample type, including human tissue. As demonstrated for the EGFR signaling pathway, by these means protein tyrosine kinase signaling pathways can be studied in a relatively quick and inexpensive manner. Several phosphotyrosine sites could be newly identified or further substantiated to be involved in EGFR signaling. By performing quantification on the peptide, rather than protein levels, different phosphorylation events on the same protein can be readily monitored.

Acknowledgements

This work was supported by the Netherlands Proteomics Centre (http://www.netherlandsproteomicscentre.nl), a program embedded in the Netherlands Genomics Initiative. We thank Dr. Rune Linding for his help with the NetworKIN analysis.

References

1. Blume-Jensen, P., and Hunter, T. (2001) Oncogenic kinase signalling. *Nature* **411**, 355-365.
2. Blagoev, B., Ong, S. E., Kratchmarova, I., and Mann, M. (2004) Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics. *Nat. Biotechnol.* **22**, 1139-1145.
3. Steen, H., Kuster, B., Fernandez, M., Pandey, A., and Mann, M. (2002) Tyrosine phosphorylation mapping of the epidermal growth factor receptor signaling pathway. *J. Biol. Chem.* **277**, 1031-1039.
4. Thelemann, A., Petti, F., Griffin, G., Iwata, K., Hunt, T., Settinari, T., Fenyo, D., Gibson, N., and Haley, J. D. (2005) Phosphotyrosine signaling networks in epidermal growth factor receptor overexpressing squamous carcinoma cells. *Mol. Cell. Proteomics* **4**, 356-376.
5. Zhang, Y., Wolf-Yadlin, A., Ross, P. L., Pappin, D. J., Rush, J., Lauffenburger, D. A., and White, F. M. (2005) Time-resolved mass spectrometry of tyrosine phosphorylation sites in the epidermal growth factor receptor signaling network reveals dynamic modules. *Mol. Cell. Proteomics* **4**, 1240-1250.
6. Pandey, A., Podtelejnikov, A. V., Blagoev, B., Bustelo, X. R., Mann, M., and Lodish, H. F. (2000) Analysis of receptor signaling pathways by mass spectrometry: Identification of Vav-2 as a substrate of the epidermal and platelet-derived growth factor receptors. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 179-184.
7. Amanchy, R., Kalume, D. E., Iwahori, A., Zhong, J., and Pandey, A. (2005) Phosphoproteome analysis of HeLa cells using stable isotope labeling with amino acids in cell culture (SILAC). *J. Proteome Res.* **4**, 1661-1671.
8. Hinsby, A. M., Olsen, J. V., Bennettt, K. L., and Mann, M. (2003) Signaling initiated by overexpression of the fibroblast growth factor receptor-1 investigated by mass spectrometry. *Mol. Cell. Proteomics* **2**, 29-36.
9. Guo, A., Villen, J., Kornhauser, J., Lee, K. A., Stokes, M. P., Rikova, K., Possemato, A., Nardone, J., Innocenti, G., Wetzel, R., Wang, Y., MacNeill, J., Mitchell, J., Gygi, S. P., Rush, J., Polakiewicz, R. D., and Comb, M. J. (2008) Signaling networks assembled by oncogenic EGFR and c-Met. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 692-697.
10. Rikova, K., Guo, A., Zeng, Q., Possemato, A., Yu, J., Haack, H., Nardone, J., Lee, K., Reeves, C., Li, Y., Hu, Y., Tan, Z., Stokes, M., Sullivan, L., Mitchell, J., Wetzel, R., Macneill, J., Ren, J. M., Yuan, J., Bakalarski, C. E., Villen, J., Kornhauser, J. M., Smith, B., Li, D., Zhou, X., Gygi, S. P., Gu, T. L., Polakiewicz, R. D., Rush, J., and Comb, M. J. (2007) Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. *Cell* **131**, 1190-1203.
11. Ballif, B. A., Carey, G. R., Sunyaev, S. R., and Gygi, S. P. (2008) Large-scale identification and evolution indexing of tyrosine phosphorylation sites from murine brain. *J. Proteome Res.* **7**, 311-318.
12. Wolf-Yadlin, A., Hautaniemi, S., Lauffenburger, D. A., and White, F. M. (2007) Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 5860-5865.
13. Wolf-Yadlin, A., Kumar, N., Zhang, Y., Hautaniemi, S., Zaman, M., Kim, H. D., Grantcharova, V., Lauffenburger, D. A., and White, F. M. (2006) Effects of HER2 overexpression on cell signaling networks governing proliferation and migration. *Mol. Syst. Biol.* **2**, 15.
14. Zheng, H., Hu, P., Quinn, D. F., and Wang, Y. K. (2005) Phosphotyrosine proteomic study of interferon alpha signaling pathway using a combination of immunoprecipitation and immobilized metal affinity chromatography. *Mol. Cell. Proteomics* **4**, 721-730.

15. Tong, J., Taylor, P., Jovceva, E., St-Germain, J. R., Jin, L. L., Nikolic, A., Gu, X., Li, Z. H., Trudel, S., and Moran, M. F. (2008) Tandem immunoprecipitation of phosphotyrosine-mass spectrometry (TIPY-MS) indicates C19ORF19 becomes tyrosine-phosphorylated and associated with activated epidermal growth factor receptor. *J. Proteome Res.* **7**, 1067-1077.
16. 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.
17. Ong, S.-E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., and Mann, M. (2002) Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Mol. Cell. Proteomics* **1**, 376-386.
18. Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A., and Pappin, D. J. (2004) Multiplexed Protein Quantitation in Saccharomyces cerevisiae Using Amine-reactive Isobaric Tagging Reagents. *Mol. Cell. Proteomics* **3**, 1154-1169.
19. Hsu, J. L., Huang, S. Y., Chow, N. H., and Chen, S. H. (2003) Stable-isotope dimethyl labeling for quantitative proteomics. *Anal. Chem.* **75**, 6843-6852.
20. Boersema, P. J., Aye, T. T., van Veen, T. A. B., Heck, A. J. R., and Mohammed, S. (2008) Triplex protein quantification based on stable isotope labeling by peptide dimethylation applied to cell and tissue lysates. *Proteomics* **8**, 4624-4632.
21. Boersema, P. J., Raijmakers, R., Lemeer, S., Mohammed, S., and Heck, A. J. R. (2009) Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nat. Protocols* **4**, 484-494.
22. Bantscheff, M., Schirle, M., Sweetman, G., Rick, J., and Kuster, B. (2007) Quantitative mass spectrometry in proteomics: a critical review. *Anal. Bioanal. Chem.* **389**, 1017-1031.
23. Lemeer, S., Jopling, C., Gouw, J., Mohammed, S., Heck, A. J. R., Slijper, M., and den Hertog, J. (2008) Comparative Phosphoproteomics of Zebrafish Fyn/Yes Morpholino Knockdown Embryos. *Mol. Cell. Proteomics* **7**, 2176-2187.
24. Raijmakers, R., Berkers, C. R., de Jong, A., Ovaa, H., Heck, A. J. R., and Mohammed, S. (2008) Automated Online Sequential Isotope Labeling for Protein Quantitation Applied to Proteasome Tissue-specific Diversity. *Mol. Cell. Proteomics* **7**, 1755-1762.
25. 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.
26. van Breukelen, B., van den Toorn, H. W. P., Drugan, M. M., and Heck, A. J. R. (2009) StatQuant: A post quantification analysis toolbox for improving quantitative mass spectrometry. *Bioinformatics* **25**, 1472-1473.
27. Käll, L., Storey, J. D., MacCoss, M. J., and Noble, W. S. (2008) Assigning Significance to Peptides Identified by Tandem Mass Spectrometry Using Decoy Databases. *J. Proteome Res.* **7**, 29-34.
28. Diella, F., Gould, C. M., Chica, C., Via, A., and Gibson, T. J. (2008) Phospho.ELM: a database of phosphorylation sites update 2008. *Nucl. Acids Res.* **36**, D240-244.
29. 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.
30. Crooks, G. E., Hon, G., Chandonia, J.-M., and Brenner, S. E. (2004) WebLogo: A Sequence Logo Generator. *Genome Res.* **14**, 1188-1190.
31. Linding, R., Jensen, L. J., Ostheimer, G. J., van Vugt, M., Jorgensen, C., Miron, I. M., Diella, F., Colwill, K., Taylor, L., Elder, K., Metalnikov, P., Nguyen, V., Pasculescu, A., Jin,

J., Park, J. G., Samson, L. D., Woodgett, J. R., Russell, R. B., Bork, P., Yaffe, M. B., and Pawson, T. (2007) Systematic discovery of in vivo phosphorylation networks. *Cell* **129**, 1415-1426.
32. Linding, R., Jensen, L. J., Pasculescu, A., Olhovsky, M., Colwill, K., Bork, P., Yaffe, M. B., and Pawson, T. (2008) NetworKIN: a resource for exploring cellular phosphorylation networks. *Nucl. Acids Res.* **36**, D695-D699.
33. Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G. (2000) Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25-29.
34. Sonnhammer, E. L. L., Eddy, S. R., and Durbin, R. (1997) Pfam: A comprehensive database of protein domain families based on seed alignments. *Proteins* **28**, 405-420.
35. Boekhorst, J., van Breukelen, B., Heck, A., and Snel, B. (2008) Comparative phosphoproteomics reveals evolutionary and functional conservation of phosphorylation across eukaryotes. *Genome Biol.* **9**, R144.
36. Hornbeck, P. V., Chabra, I., Kornhauser, J. M., Skrzypek, E., and Zhang, B. (2004) Phosphosite: A bioinformatics resource dedicated to physiological protein phosphorylation. *Proteomics* **4**, 1551-1561.
37. Mann, M., Ong, S.-E., Grønborg, M., Steen, H., Jensen, O. N., and Pandey, A. (2002) Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome. *Trends Biotechnol.* **20**, 261-268.
38. 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.
39. Kratchmarova, I., Blagoev, B., Haack-Sorensen, M., Kassem, M., and Mann, M. (2005) Mechanism of divergent growth factor effects in mesenchymal stem cell differentiation. *Science* **308**, 1472-1477.
40. Huyer, G., Liu, S., Kelly, J., Moffat, J., Payette, P., Kennedy, B., Tsaprailis, G., Gresser, M. J., and Ramachandran, C. (1997) Mechanism of inhibition of protein-tyrosine phosphatases by vanadate and pervanadate. *J. Biol. Chem.* **272**, 843-851.
41. Nuhse, T. S., Stensballe, A., Jensen, O. N., and Peck, S. C. (2004) Phosphoproteomics of the Arabidopsis plasma membrane and a new phosphorylation site database. *Plant Cell* **16**, 2394 - 2405.
42. Yarden, Y. (2001) The EGFR family and its ligands in human cancer: signalling mechanisms and therapeutic opportunities. *Eur. J. Cancer* **37**, 3-8.
43. Morandell, S., Stasyk, T., Skvortsov, S., Ascher, S., and Huber, L. A. (2008) Quantitative proteomics and phosphoproteomics reveal novel insights into complexity and dynamics of the EGFR signaling network. *Proteomics* **8**, 4383-4401.
44. Huang, P. H., Mukasa, A., Bonavia, R., Flynn, R. A., Brewer, Z. E., Cavenee, W. K., Furnari, F. B., and White, F. M. (2007) Quantitative analysis of EGFRvIII cellular signaling networks reveals a combinatorial therapeutic strategy for glioblastoma. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 12867-12872.
45. Schlessinger, J. (2002) Ligand-Induced, Receptor-Mediated Dimerization and Activation of EGF Receptor. *Cell* **110**, 669-672.
46. Schulze, W. X., Deng, L., and Mann, M. (2005) Phosphotyrosine interactome of the ErbB-receptor kinase family. *Mol. Syst. Biol.* **1**, 13.
47. Yamauchi, T., Ueki, K., Tobe, K., Tamemoto, H., Sekine, N., Wada, M., Honjo, M., Takahashi, M., Takahashi, T., Hirai, H., Tushima, T., Akanuma, Y., Fujita, T., Komuro, I., Yazaki, Y., and Kadowaki, T. (1997) Tyrosine phosphorylation of the EGF receptor by the kinase Jak2 is induced by growth hormone. *Nature* **390**, 91-96.

48. Gogg, S., and Smith, U. (2002) Epidermal Growth Factor and Transforming Growth Factor alpha Mimic the Effects of Insulin in Human Fat Cells and Augment Downstream Signaling in Insulin Resistance. *J. Biol. Chem.* **277**, 36045-36051.
49. Amanchy, R., Zhong, J., Molina, H., Chaerkady, R., Iwahori, A., Kalume, D. E., GrÃ¸nborg, M., Joore, J., Cope, L., and Pandey, A. (2008) Identification of c-Src Tyrosine Kinase Substrates Using Mass Spectrometry and Peptide Microarrays. *J. Proteome Res.* **7**, 3900-3910.
50. Leroy, C., Fialin, C., Sirvent, A., Simon, V., Urbach, S., Poncet, J., Robert, B., Jouin, P., and Roche, S. (2009) Quantitative Phosphoproteomics Reveals a Cluster of Tyrosine Kinases That Mediates Src Invasive Activity in Advanced Colon Carcinoma Cells. *Cancer Res.* **69**, 2279-2286.
51. Luo, W., Slebos, R. J., Hill, S., Li, M., BraÌ bek, J., Amanchy, R., Chaerkady, R., Pandey, A., Ham, A.-J. L., and Hanks, S. K. (2008) Global Impact of Oncogenic Src on a Phosphotyrosine Proteome. *J. Proteome Res.* **7**, 3447-3460.
52. Head, J. A., Jiang, D. Y., Li, M., Zorn, L. J., Schaefer, E. M., Parsons, J. T., and Weed, S. A. (2003) Cortactin tyrosine phosphorylation requires Rac1 activity and association with the cortical actin cytoskeleton. *Mol. Biol. Cell* **14**, 3216-3229.
53. Nieto-Pelegrin, E., and Martinez-Quiles, N. (2009) Distinct phosphorylation requirements regulate cortactin activation by TirEPEC and its binding to N-WASP. *Cell Commun. Signal.* **7**, 11.
54. Martinez-Quiles, N., Ho, H.-Y. H., Kirschner, M. W., Ramesh, N., and Geha, R. S. (2004) Erk/Src Phosphorylation of Cortactin Acts as a Switch On-Switch Off Mechanism That Controls Its Ability To Activate N-WASP. *Mol. Cell. Biol.* **24**, 5269-5280.

Figure legends

Figure 1. A) LC-MS base peak chromatogram of the eluate from the phosphotyrosine immunoprecipitated pervanadate treated HeLa digest. The numbers indicate intense peaks representing phosphotyrosine containing peptides. Only a few non-phosphorylated peptides were detected in this eluate. A numbered list of these abundant phosphopeptides is available as Supplemental Table 1. * represents a peak of an ion that could not be identified and most likely is a non-peptide species. B) Number of identified phosphotyrosine and non-phosphorylated peptides as a function of the applied Mascot threshold score revealing the apparent specificity of the immunopurification. With decreasing Mascot score threshold the number of non-modified peptides in the dataset increases exponentially below 20. The inflection point at score 20 indicates that at an even lower score, relatively more false positive identifications are introduced in the dataset.

Figure 2. Motif-X (29) analysis of the here acquired HeLa phosphotyrosine dataset identified six conserved motifs, here displayed as WebLogos (30). The numbers indicate the number of peptides exhibiting this motif in the full dataset.

Figure 3. Experimental scheme for the quantitative phosphotyrosine proteome studies. Cell cultures were mock treated or stimulated with pervanadate or EGF. After lysis and enzymatic digestion, peptides were differentially stable isotope dimethyl labeled and combined before immunoprecipitation with a phosphotyrosine specific antibody. The precipitate was analyzed by LC-MS followed by quantification using the triplet peaks originating from the different isotopes.

Figure 4. Representative examples of mass spectra and fragmentation spectra of peptides identified and quantified after pervanadate treatment. A) The abundance of SHIP-2 peptide TLSEVDpYAPAGPAR (fragmentation spectrum shown of m/z 781.8848, +2, heavy dimethyl labeled) is dramatically increased by pervanadate treatment. B) The abundance of peptide IGEGTpYGVVYK (fragmentation spectrum shown of m/z 665.3498, +2, intermediate dimethyl labeled) is not affected by pervanadate treatment. * indicates the site of stable isotope dimethyl labeling.

Figure 5. A) Clustering of tyrosine phosphorylation profiles. Cluster 1 (dark grey): no or only a small change in phosphorylation levels; cluster 2 (light grey): an immediate upregulation that remains after 30 min; cluster 3 (black): strong and quick increase in phosphorylation. B, C, D) Representative examples of mass spectra and fragmentation spectra of peptides identified and quantified after EGF stimulation from each of the clusters. B) BCAR1 peptide HLLAPGPQDIpYDVPPVR (fragmentation spectrum shown of m/z 1002.0319, +2, heavy dimethyl labeled) is from cluster 1 with only a slight increase in phosphorylation upon EGF stimulation. C) SGK269 peptide SSAIRpYQEVWTSSTSPR (fragmentation spectrum shown of m/z 689.6682, +3, intermediate dimethyl labeled) is from cluster 2 with a larger increase after EGF stimulation. D) RBCK1 peptide NSQEAEVSCPFIDNTpYSCSGK (fragmentation spectrum shown of m/z 1273.0631, +2, heavy dimethyl labeled) is from cluster 3 showing an extensive increase in abundance after EGF stimulation. * indicates the site of stable isotope dimethyl labeling.

Figure 6. Ratios of changes in tyrosine phosphorylation A) 10 min and B) 30 min after EGF stimulation detected in the current study plotted against ratios of overlapping phosphotyrosine sites found in Zhang *et al.*(5).

Table 1. Phosphotyrosine peptides whose abundance did not change significantly upon ervanadate treatment of HeLa cells. L, light labeled peptide (mock treated), I, intermediate labeled peptide (1:1 mock and pervanadate treated), H, heavy labeled (pervanadate treated).

IPI acc. nr.	Gene	Peptide	Site 1	Site 2	ratio I/L	STDEV I/L	ratio H/L	STDEV H/L
IPI00016932	SHIP-2	NSFNNPApYpYVLEGVPHQLLPPEPPSPAR	Y986	Y987	1.04	-	2.38	-
IPI00008530	RPLP0	IIQLLDDpYPK	Y24		0.81	-	0.65	-
IPI00000352	DYRK1B	IYQpYIQSR	Y273		0.92	-	1.03	-
IPI00291175	VCL	SFLDSGpYR	Y822		0.93	-	0.73	-
IPI00396485	EEF1A1	EHALLApYTLGVK	Y141		1.02	-	0.93	-
IPI00026689	CDC2	IEKIGEGTpYGVVYK	Y15		1.06	0.05	1.12	0.10
IPI00028570	GSK3B	QLVRGEPNVSpYICSR	Y216		1.10	0.01	1.68	0.15
IPI00654623	TNS3	KLSLGQpYDNDAGGQLPFSK	Y540		1.15	-	1.19	-
IPI00023503	CDK3	VEKIGEGTpYGVVYK	Y43		1.16	-	1.41	-
IPI00026689	CDC2	IGEGpTpYGVVYK	T14	Y15	1.27	0.04	1.49	0.08
IPI00013721	PRPF4B	LCDFGSASHVADNDITPpYLVSR	Y849		1.29	0.08	1.33	0.08
IPI00012885	PTK2	Y(ox)MEDSTpYYKASK	Y576		1.30	-	1.42	-
IPI00008438	RPS10	IAIpYELLFK	Y12		1.41	-	1.70	-
IPI00013981	YES1	LIEDNEpYTAR	Y426		1.47	-	2.22	-
IPI00166578	PRAGMIN	EATQPEPIpYAESTKR	Y411		1.66	0.03	2.44	0.23

Table 2. Identified and quantified tyrosine phosphopeptides after EGF stimulation.

IPI acc. nr.	Gene	Peptide	Site 1	Site 2	log2 ratio 10/0min	stdev	log2 ratio 30/0min	stdev
IPI00011676	N-WASP	VIpYDFIEK	Y256		-1.49	0.66	-1.86	0.74
IPI00465248	ENO1	AAVPSGASTGIpYEALELR	Y44		-0.90	0.25	-0.08	0.25
IPI00029601	CTTN	TQTPPVSPAPQPTEERLPSSPVpYEDAASFK	Y421		-0.73	-	-1.47	-
IPI00220644	PKM2	TATESFASDPILpYRPVAVALDTK	Y105		-0.68	-	-0.13	-
IPI00012885	PTK2	GSIDREDGSLQGPIGNQHIpYQPVGKPDPAAPPK	Y861		-0.52	0.45	0.43	0.09
IPI00216423	ITSN2	REEPEALpYAAVNK	Y967		-0.41	-	-0.79	-
IPI00021439	ACTB	IWHHTFpYNELR	Y91		-0.32	-	0.57	-
IPI00008530	RPLP0	IIQLLDDpYPK	Y24		-0.14	-	-0.16	-
IPI00215949	HIPK2	AVCSTpYLQSR	Y361		-0.13	-	-0.34	-
IPI00013721	PRPF4B	LCDFGSASHVADNDITPpYLVSR	Y849		-0.10	-	0.09	-
IPI00000352	DYRK1B	IYQpYIQSR	Y273		-0.06	-	-0.04	-
IPI00217966	LDHA	QVVESApYEVIK	Y239		0.02	-	0.30	-
IPI00641339	BCAR1	GLPPSNHHAVpYDVPPSVSK	Y306		0.05	-	0.58	-
IPI00552750	TNK2	KPTpYDPVSEDQDPLSSDFKR	Y574		0.08	0.12	-1.30	0.25
IPI00641339	BCAR1	AQQGLpYQVPGPSPQFQSPPAK	Y128		0.16	-	0.77	-
IPI00641339	BCAR1	HLLAPGPQDIpYDVPPVR	Y249		0.20	0.15	0.73	0.22
IPI00301058	VASP	VQIpYHNPTANSFR	Y39		0.21	0.12	0.26	0.06
IPI00654623	TNS3	LSLGQpYDNDAGGQLPFSK	Y540		0.24	-	0.40	-
IPI00021076	PKP4	NNYALNTTATpYAEPYRPIQYR	Y478		0.32	-	0.77	-
IPI00306959	KRT7	LSSARPGGLGSSSLpYGLGASRPR	Y40		0.32	0.10	0.82	0.01
IPI00641339	BCAR1	RPGPGTLpYDVPRER	Y387		0.33	0.15	1.00	0.07
IPI00021267	EPHA2	TYVDPHTpYEDPNQAVLK	Y594		0.35	-	0.69	-
IPI00166578	PRAGMIN	EATQPEPIpYAESTK	Y411		0.39	0.14	0.20	0.23
IPI00291175	VCL	SFLDSGpYR	Y822		0.40	-	0.35	-
IPI00220030	PXN	VGEEEHVpYSFPNK	Y118		0.43	0.19	0.50	0.11
IPI00013981	Src	LIEDNEpYTAR	Y426		0.44	-	0.72	-
IPI00418471	VIM	SLYASSPGGVpYATR	Y61		0.45	-	0.65	-
IPI00220030	PXN	FIHQQPQSSSPVpYGSSAK	Y88		0.47	-	0.50	-
IPI00396485	EEF1A1	STTTGHLIpYK	Y29		0.50	-	0.32	-
IPI00298347	PTPN11	IQNTGDpYYDLYGGEK	Y62		0.50	-	0.75	-
IPI00026689	CDC2	IGEGpTpYGVVYK	T14	Y15	0.58	0.00	1.22	0.00
IPI00026689	CDC2	IGEGTYGVVpYKGR	Y19		0.59	0.15	0.49	0.03
IPI00016932	SHIP-2	ERLpYEWISIDKDEAGAK	Y886		0.63	-	1.55	-

IPI00021267	EPHA2	VLEDDPEATpYTTSGGKIPIR	Y772		0.65	-	1.04	-
IPI00220030	PXN	FIHQQPQSSpSPVpYGSSAK	S85	Y88	0.74	0.00	0.91	0.00
IPI00396485	EEF1A1	EHALLApYTLGVK	Y141		0.80	-	0.61	-
IPI00182469	CTNND1	SLDNNpYSTPNER	Y898		0.90	-	0.33	-
IPI00012885	PTK2	YMEDSTpYYK	Y576		0.91	-	0.71	-
IPI00182469	CTNND1	LNGPQDHSHLLpYSTIPR	Y96		0.98	-	1.03	-
IPI00737545	SGK269	NAIKVPIVINPNApYDNLAIYK	Y635		1.17	-	0.98	-
IPI00022353	TYK2	LLAQAEGEPCpYIR	Y292		1.19	-	1.01	-
IPI00215965	HNRNPA1	SSGPpYGGGGQYFAKPR	Y341		1.22	-	0.45	-
IPI00028570	GSK3B	GEPNVSpYICSR	Y216		1.34	0.74	0.95	0.39
IPI00176903	PTRF	SFTPDHVVpYAR	Y308		1.44	-	1.15	-
IPI00014454	RIN1	EKPAQDPLpYDVPNASGGQAGGPQRPGR	Y36		1.53	0.05	1.81	0.11
IPI00412752	STAT3	YCRPESQEHPEADPGSAAPpYLK	Y705		1.56	-	0.96	-
IPI00029601	CTTN	GPVSGTEPEPVpYSMEAADYR	Y446		1.57	-	1.21	-
IPI00026689	CDC2	IEKIGEGTpYGVVYK	Y15		1.66	-	0.71	-
IPI00334715	GRLF1	NEEENIpYSVPHDSTQGK	Y1105		1.69	0.00	1.64	0.04
IPI00552750	TNK2	VSSTHpYYLLPERPSYLER	Y913		1.87	-	0.89	-
IPI00216423	ITSN2	LIpYLVPEK	Y552		1.93	-	1.26	-
IPI00022462	TFRC	SAFSNLFGGEPLSpYTR	Y20		2.04	-	1.41	-
IPI00737545	SGK269	SSAIRpYQEVWTSSTSPR	Y531		2.04	-	1.63	-
IPI00464978	IRS2	SYKAPYTCGGDSDQpYVLMSSPVGR	Y825		2.16	-	1.61	-
IPI00006482	ATP1A1	GIVVpYTGDR	Y260		2.28	-	1.88	-
IPI00031068	GAB1	APSASVDSSLpYNLPR	Y259		2.42	-	2.19	-
IPI00002857	p38a	HTDDEMTGpYVATR	Y182		2.98	-	0.36	-
IPI00294619	TFG	NRPPFGQGpYTQPGPGYR	Y392		4.13	-	1.04	-
IPI00003479	Erk1	VADPDHDHTGFLTEpYVATR	Y187		3.98	0.29	2.68	0.25
IPI00003479	Erk1	VADPDHDHTGFLpTEpYVATR	T185	Y187	4.31	-	2.98	-
IPI00016932	SHIP-2	NSFNNPApYYVLEGVPHQLLPPEPPSPAR	Y986		4.49	0.27	3.29	0.48
IPI00023343	DLG3	RDNEVDGQDpYHFVVSR	Y673		4.74	-	4.15	-
IPI00018274	EGFR	GSHQISLDNPDpYQQDFFPK	Y1172		5.49	-	4.27	-
IPI00016932	SHIP-2	TLSEVDpYAPAGPAR	Y1135		5.70	-	3.35	-
IPI00020178	STAM	QQSTTLSTLpYPSTSSLLTNHQHEGR	Y198		5.84	0.24	3.64	0.25
IPI00513796	SHC1	ELFDDPSpYVNVQNLDKAR	Y317		5.97	0.69	4.71	0.61
IPI00290542	STAM2	LVNEAPVpYSVYSK	Y371		6.12	-	2.78	-

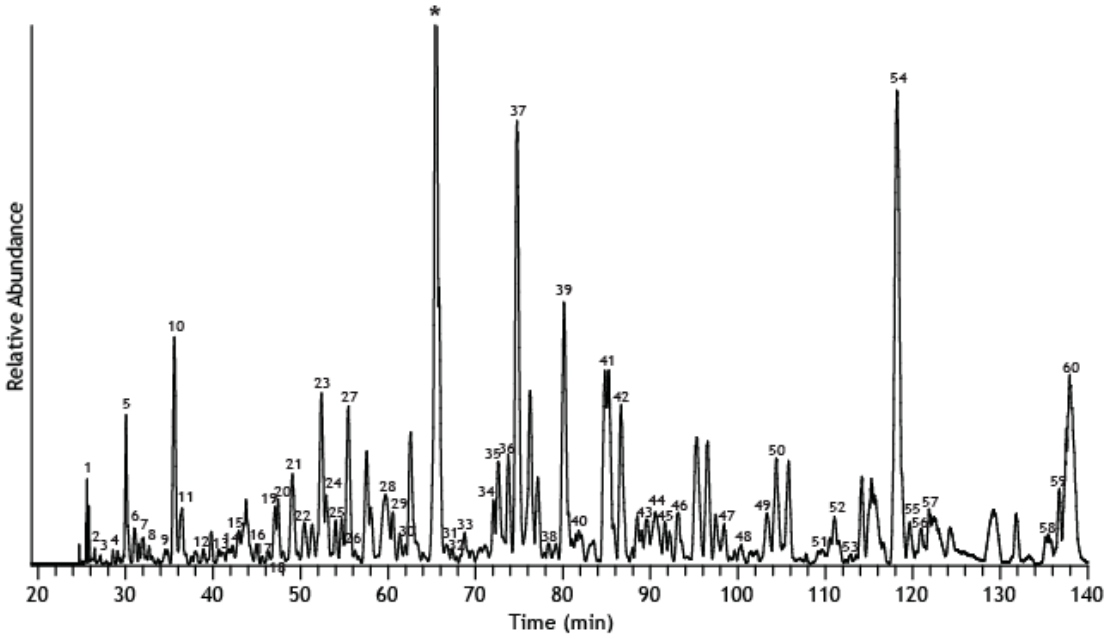
IPI00002495	EPN1	NIVHNpYSEAEIK	Y17	6.45	-	5.19	-
IPI00018274	EGFR	GSTAENAEpYLR	Y1197	6.68	-	4.07	-
IPI00027269	CBL	IKPSSSANAIpYSLAARPLPVPK	Y674	6.81	0.05	3.04	0.35
IPI00031068	GAB1	SSGSGSSVADERVDpYVVVDQQK	Y659	8.59	-	7.76	-
IPI00010748	RBCK1	NSQEAEVSCPFIDNTpYSCSGK	Y288	8.93	-	6.95	-
IPI00290542	STAM2	SLpYPSSEIQLNNK	Y192	9.57	-	4.95	-

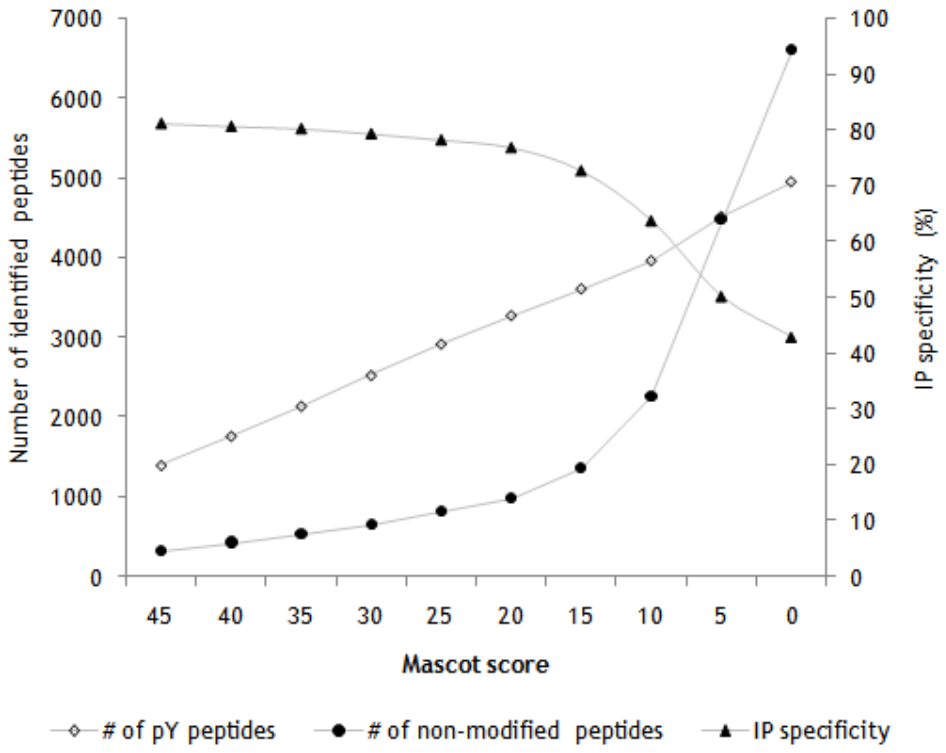
Table 3. Ratios of changes in tyrosine phosphorylation upon EGF stimulation of phosphorylation sites detected in the current study and Zhang *et al.* (5).

		Zhang et al. (2005)		Boersema et al. (2009)	
Peptide Sequence/phosphosite	pY-site	log2 ratio, 10/0 min	log2 ratio, 30/0 min	log2 ratio, 10/0 min	log2 ratio, 30/0 min
AAVPSGASTGIpYEALELR	Y44	0.47	0.42	-0.90	-0.08
LCDFGSASHVADNDITPpYLVSR	Y849	0.44	0.07	-0.10	0.09
EATQPEPIpYAESTKR	Y411	0.06	0.34	0.39	0.20
VGEEEHVpYSFPNK	Y118	0.52	0.59	0.43	0.50
TYVDPHTpYEDPNQAVLK	Y594	1.07	0.86	0.35	0.69
YMEDSTpYYK	Y576	1.16	1.35	0.91	0.71
IGEGTpYGVVYK	Y15	0.79	0.30	1.66	0.71
HLLAPGPQDIpYDVPPVR	Y249	1.03	1.33	0.20	0.73
IQNTGDpYYDLYGGEK	Y62	0.42	0.25	0.50	0.75
VSSTHpYYLLPERPSYLER	Y857	2.91	1.31	1.87	0.89
YCRPESQEHPEADPGSAAPpYLK	Y705	1.94	0.23	1.56	0.96
GEPNVSpYICSR	Y279	0.39	0.14	1.34	0.95
VLEDDPEATpYTTSGGKIPIR	Y772	0.87	0.82	0.65	1.04
GPVSGTEPEPVpYSMEAADYR	Y446	0.95	0.97	1.57	1.21
SAFSNLFGGEPLSpYTR	Y20	1.44	0.85	2.04	1.41
EKPAQDPLpYDVPNASGGQAGGPQR PGR	Y36	1.84	1.55	1.53	1.81
VADPDHDHTGFLTEpYVATR	Y204	3.42	3.07	3.98	2.68
LVNEAPVYSVpYSK	Y374	3.66	2.06	6.12	2.78
NSFNNPApYYVLEGVPHQLLPPEPPpS PARAPVPSATK	Y986/ S1003	5.10	3.83	4.49	3.29
GSHQISLDNPDpYQQDFFPK	Y1172	4.16	3.32	5.49	4.27
ELFDDPSpYVNVQNLDK	Y317	5.26	4.52	5.97	4.71
NSQEAEVSCPFIDNTpYSCSGK	Y288	6.21	4.07	8.93	6.95

Figure 1

A





B

Figure 2

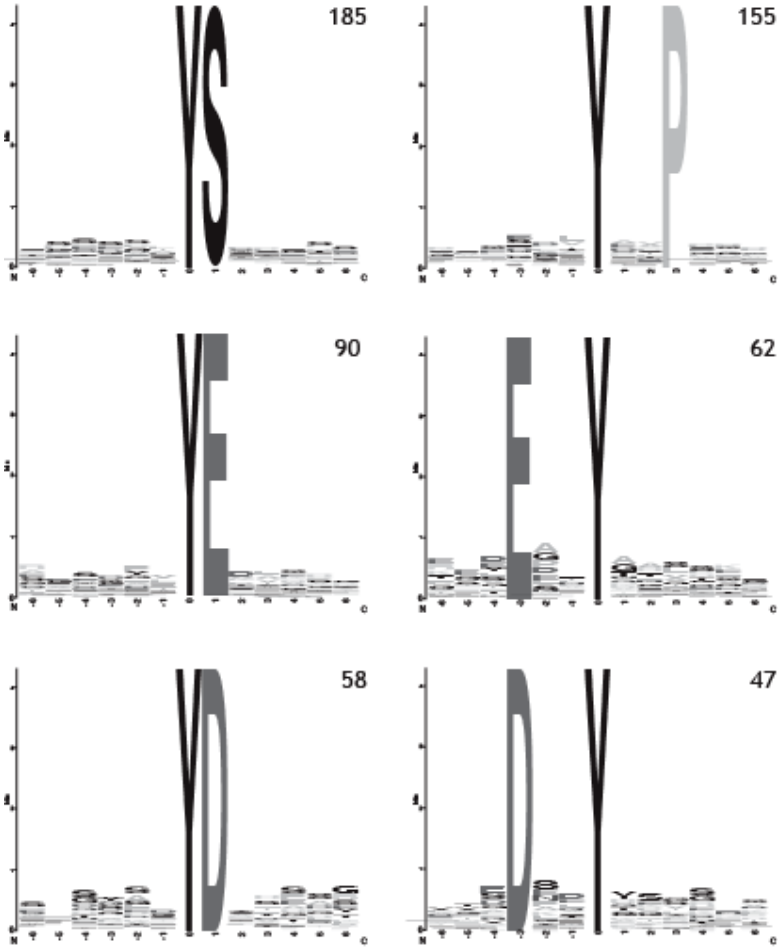


Figure 3

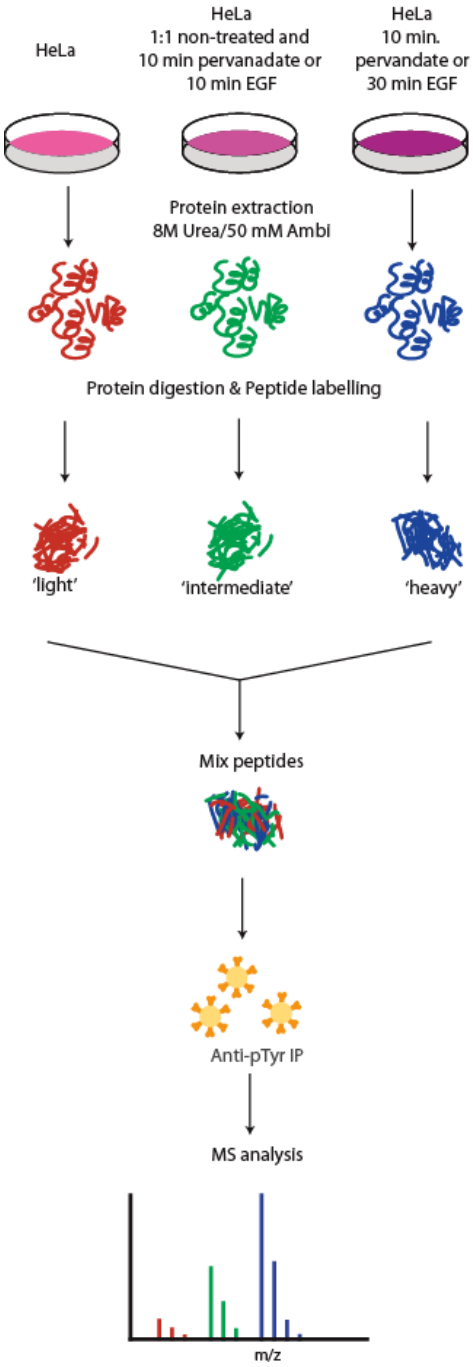
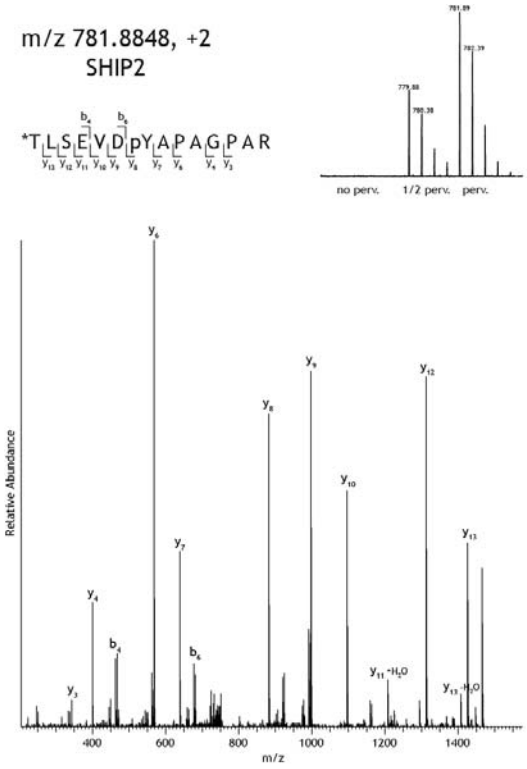


Figure 4

A



B

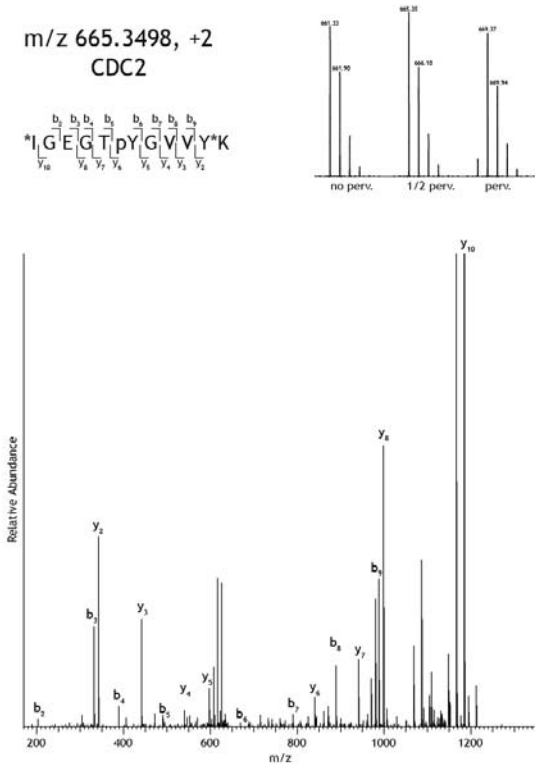
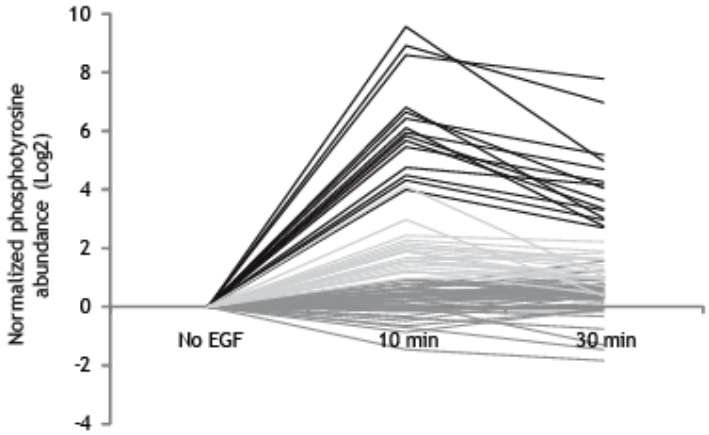


Figure 5

A



B

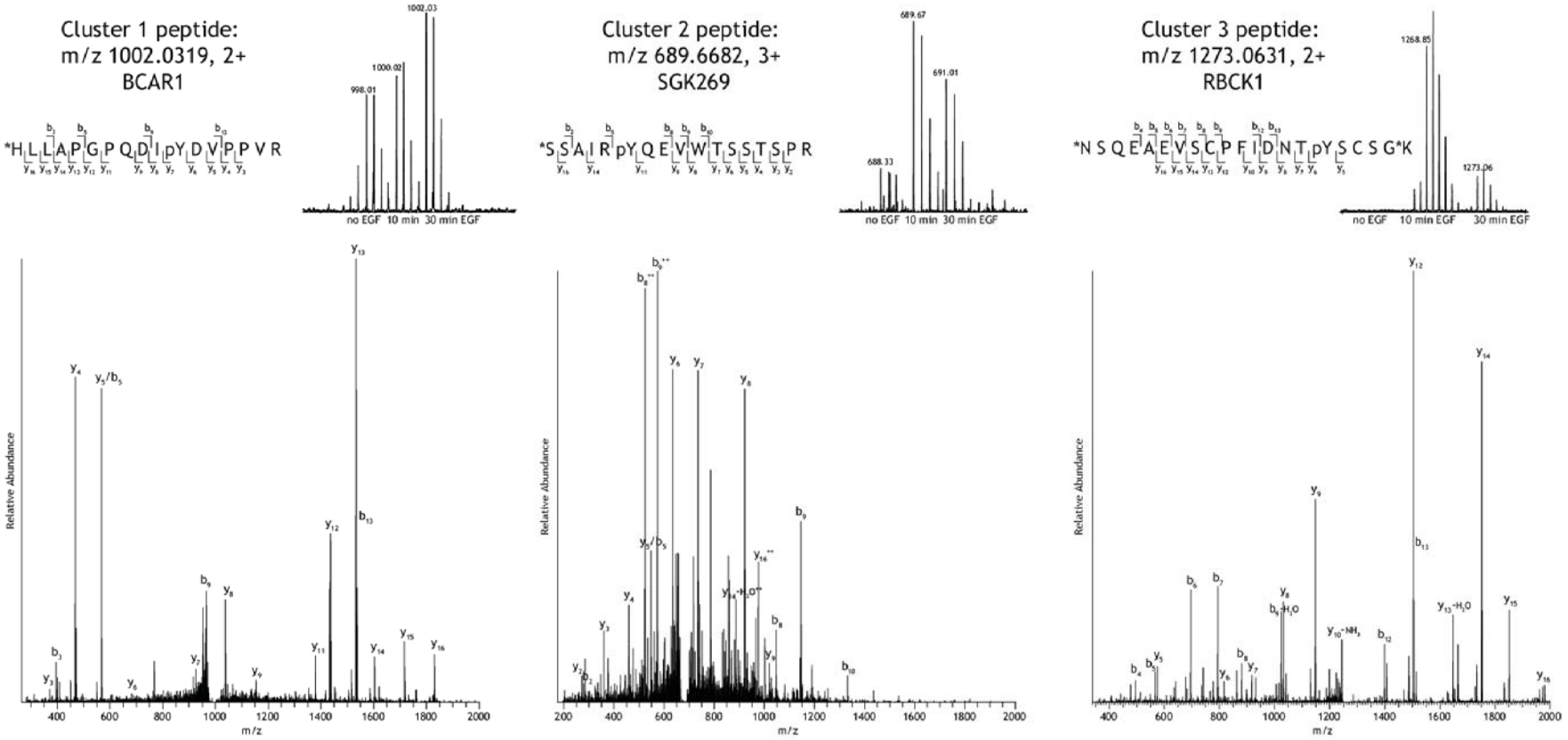
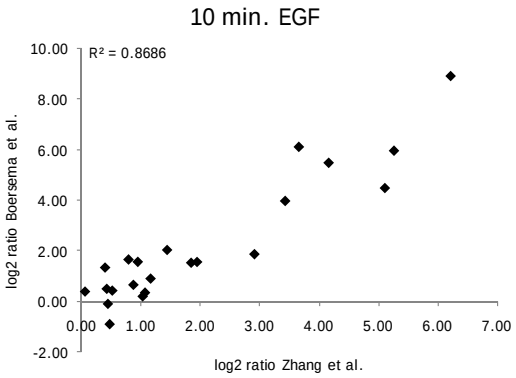


Figure 6

A



B