

Immunoaffinity profiling of tyrosine phosphorylation in cancer cells

John Rush¹, Albrecht Moritz¹, Kimberly A Lee¹, Ailan Guo¹, Valerie L Goss¹, Erik J Spek¹, Hui Zhang^{1,2}, Xiang-Ming Zha^{1,2}, Roberto D Polakiewicz¹ & Michael J Comb¹

Tyrosine kinases play a prominent role in human cancer, yet the oncogenic signaling pathways driving cell proliferation and survival have been difficult to identify, in part because of the complexity of the pathways and in part because of low cellular levels of tyrosine phosphorylation. In general, global phosphoproteomic approaches reveal small numbers of peptides containing phosphotyrosine. We have developed a strategy that emphasizes the phosphotyrosine component of the phosphoproteome and identifies large numbers of tyrosine phosphorylation sites. Peptides containing phosphotyrosine are isolated directly from protease-digested cellular protein extracts with a phosphotyrosine-specific antibody and are identified by tandem mass spectrometry. Applying this approach to several cell systems, including cancer cell lines, shows it can be used to identify activated protein kinases and their phosphorylated substrates without prior knowledge of the signaling networks that are activated, a first step in profiling normal and oncogenic signaling networks.

Unregulated tyrosine kinase activity can drive malignancy and tumor formation by generating inappropriate proliferation and survival signals¹. The ability of researchers to recognize unregulated tyrosine kinases and the pathways driving tumor growth and survival could play a pivotal role in the development of effective drug therapies² and in the development of assays to identify drug-responsive patients. Progress in these areas is limited in part by a need for experimental approaches that can isolate and identify tyrosine-phosphorylated peptides in large numbers, without preconceived biases about where tyrosine phosphorylation sites will be found. Current phosphoproteomic approaches generally reveal only small numbers of tyrosine phosphorylation sites, in keeping with the low level of phosphotyrosine relative to phosphoserine and phosphothreonine residues³.

For example, one current phosphoproteomic approach^{4,5} uses immobilized metal affinity chromatography (IMAC) to purify phosphopeptides in bulk from protease-digested cellular protein extracts, which have been methylated to reduce nonspecific binding of acidic unphosphorylated peptides to IMAC resin. The isolated peptides are then identified by reversed-phase liquid chromatography–tandem mass spectrometry (LC-MS/MS). In a yeast protein extract, this

approach identified 216 phosphopeptides and 324 phosphorylation sites but only 3 tyrosine phosphorylation sites⁴. Similarly this approach identified only 3 tyrosine phosphorylation sites among 17 phosphopeptides purified from capacitated human sperm cells⁵.

A targeted phosphoprotein identification approach uses phosphotyrosine-specific antibodies to immunoprecipitate tyrosine-phosphorylated proteins from cell extracts. Immunoprecipitated proteins are separated by gel electrophoresis, digested in gel bands with a protease and analyzed by MS/MS. Typically, immunoprecipitated proteins are identified without identifying their tyrosine phosphorylation sites. For example, vav-2⁶, STAM2⁷ and Odin⁸ were isolated from epidermal growth factor–treated HeLa cells by immunoprecipitation with a phosphotyrosine-specific antibody. After digestion these three proteins were identified by MS/MS, but no tyrosine phosphorylation sites were found even though all three proteins were known to be tyrosine-phosphorylated.

A third approach combines immunoprecipitation of phosphotyrosine proteins and purification of phosphopeptides by the methyl-esterification/IMAC method to increase the number of identified tyrosine phosphorylation sites^{5,9,10}. This resulted in the identification of 19 tyrosine phosphorylation sites from Jurkat cells stimulated by T-cell receptor ligation⁹, a total of 38 phosphotyrosine sites from two cell lines expressing the Bcr-Abl fusion tyrosine kinase⁹, and after method optimization¹⁰, as many as 66 phosphotyrosine sites from 10⁹ Jurkat cells treated with the tyrosine phosphatase inhibitor pervanadate.

Here we describe a simple strategy that uses a phosphotyrosine-specific antibody to immunoprecipitate phosphotyrosine peptides directly from digested cellular protein extracts and standard LC-MS/MS methods to identify large numbers of tyrosine phosphorylation sites. Our results contradict the prevailing view that phosphospecific antibodies are not generally suitable for purifying phosphopeptides^{3,11,12}. Using this strategy we have identified 688 nonredundant phosphotyrosine-containing peptides and 628 phosphotyrosine sites, the majority of which are novel, from three distinct cell types. We show that this immunoaffinity profiling strategy can be used to identify protein kinases that are abnormally activated in cancer cells as well as their substrates and other downstream elements.

To evaluate the feasibility of our approach for purifying tyrosine-phosphorylated peptides from complex peptide mixtures, we used

¹Cell Signaling Technology Inc., 166B Cummings Center, Beverly, Massachusetts 01915, USA. ²Present addresses: Institute for Systems Biology, 1441 N. 34th St., Seattle, Washington 98103, USA (H.Z.) and Department of Internal Medicine, Howard Hughes Medical Institute, Roy J. and Lucille A. Carver College of Medicine, University of Iowa, Iowa City, Iowa 52242, USA (X.Z.). Correspondence should be addressed to M.J.C. (mcomb@cellsignal.com).

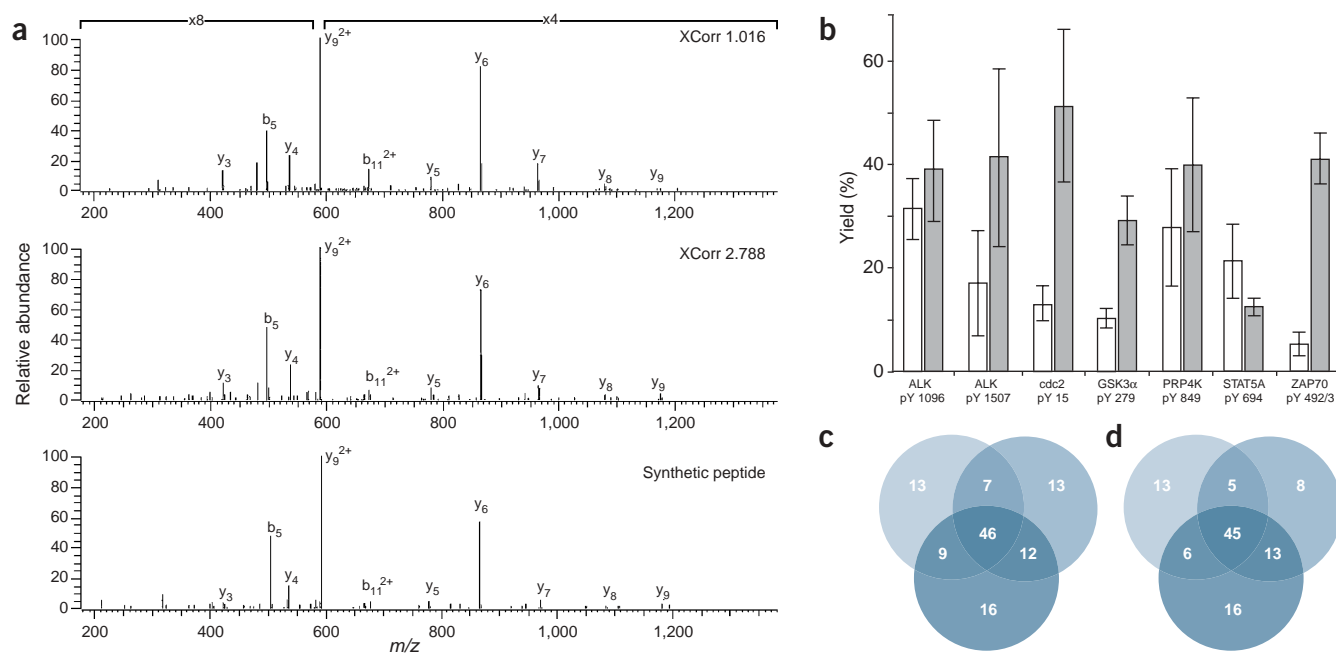


Figure 1 Characteristics of the immunoaffinity profiling strategy. **(a)** The reference library-searching approach. A low-scoring MS/MS spectrum (top) assigned to the GSK3 α phosphopeptide GEPNVSpYKSR was accepted because of its similarity to a high-scoring spectrum (middle) assigned to the same peptide in another study. The spectrum of a synthetic peptide with the assigned sequence (bottom) further validates the assignment. **(b)** Yield of low-abundance phosphopeptides isolated using the immunoaffinity strategy from pervanadate-treated Jurkat cells (white bars) and from a 1:100 dilution of peptides from pervanadate-treated cells into peptides from untreated Jurkat cells (gray bars). Error bars show the standard deviation of four separate measurements of yield by MS peak area. Yield is the peak area for the heavy-isotope peptide added to a peptide fraction before immunoaffinity-isolation. **(c)** Variability associated with immunoaffinity isolation and analysis. Phosphopeptides isolated in triplicate from one solid-phase extraction fraction were analyzed separately. The three batches contained 75 (left circle), 78 (right circle), and 83 (bottom circle) nonredundant phosphopeptide sequences. **(d)** Variability associated with analysis. Phosphopeptides isolated in triplicate from one solid-phase extraction fraction were pooled before analysis. The three batches contained 69 (left circle), 71 (right circle), and 80 (bottom circle) nonredundant phosphopeptide sequences and showed nearly the same degree of overlap as in **c**.

pervanadate-treated Jurkat cells as a model system. Jurkat cells are an established leukemic T-cell line¹³, and pervanadate induces hyperphosphorylation by disabling protein tyrosine phosphatases¹⁴. Proteins were extracted from 2×10^8 cells under denaturing conditions and digested with trypsin, and the resulting complex peptide mixture was separated from nonpeptide components, concentrated and partitioned into three fractions by reversed-phase solid-phase extraction. Each fraction was then treated with the phosphotyrosine-specific antibody P-Tyr-100 immobilized on agarose beads. After thorough washing, peptides were eluted from the immobilized antibody with dilute acid and analyzed by nanoflow LC-MS/MS using an ion trap mass spectrometer. MS/MS spectra were assigned to peptide sequences using the program Sequest¹⁵. Assigned sequences and spectra were imported into a relational database, so that we could filter assignments based on Sequest scores, annotate correct and incorrect assignments, and compare all the spectra assigned to a particular sequence in all the studies we have conducted. Lists of credible phosphopeptide sequence assignments were assembled in a conservative two-step process that used a manual review of spectra to reject false-positive assignments and that simulated a true reference library-searching strategy to allow us to accept low-scoring but correct false-negative assignments. The utility of the reference library-searching strategy is exemplified by the assignment made for Tyr-279 in GSK3 α in **Supplementary Table 1** online. This phosphorylation site would not have been identified by relying on Sequest score-filtering, but its low-scoring MS/MS spectrum (XCorr 1.016) was matched to a high-scoring spectrum observed

from another study (XCorr 2.788) and to the spectrum of a synthetic phosphopeptide (**Fig. 1a**).

We found 194 phosphotyrosine sites in 185 phosphorylated peptides in a single analysis of this sample (see **Supplementary Table 2** online). Most purified peptides were phosphorylated: 90 of the 100 top-scoring Sequest assignments were phosphotyrosine peptides. Phosphopeptides originating from signaling proteins that were detected in Jurkat cells are listed in **Supplementary Table 1** online. About one-fifth (36 of 185) of the phosphotyrosine peptides were from proteins with well-documented roles in T-cell receptor signaling (**Supplementary Table 2** online). Among these were the tyrosine kinases ZAP70 and Lck, which are activated by T-cell receptor stimulation¹⁶. Six tyrosine phosphorylation sites were found in ZAP70, including both Tyr-492 and Tyr-493 in the activation loop, and two sites were found in Lck, including the activation loop site Tyr-394. All the tyrosine phosphorylation sites found in ZAP70 and Lck have been reported previously. A complete list is provided as **Supplementary Table 2** online.

To evaluate the recovery and sensitivity of the immunoaffinity profiling approach, we first constructed two peptide fractions to represent samples with high and low levels of cellular phosphorylation. The high-level sample was a peptide fraction prepared from trypsin-digested, pervanadate-treated Jurkat cells, and the low-level sample was the same peptide fraction diluted 1:100 with peptides from untreated Jurkat cells. We then added seven heavy isotope-labeled synthetic phosphotyrosine peptides at 500 copies per cell (167 fmol

peptide in 2×10^8 cells) to these peptide fractions and isolated phosphotyrosine peptides using the immunoaffinity strategy. Peptides were analyzed as described above except the mass spectrometer was operated in MS-only mode to give better shaped peaks for calculating yield from the extracted ion chromatograms of the heavy-isotope peptides. All seven phosphopeptides were isolated in good yield even though they were present at very low abundance and even when the background level of phosphorylation was high (Fig. 1b).

To evaluate reproducibility, we prepared three peptide fractions from trypsin-digested, pervanadate-treated Jurkat cells by solid-phase extraction, and we then isolated three batches of phosphotyrosine-containing peptides from one of the three solid-phase extraction fractions. Each batch was analyzed separately by LC-MS/MS as described above, except we accepted all assigned sequences with above-threshold Sequest scores to remove any bias associated with spectrum interpretation (for this study only, using thresholds described in Methods). The three phosphopeptide batches showed a high level of overlap (Fig. 1c and Supplementary Table 3 online): 46 nonredundant phosphotyrosine peptide sequences were found in all three batches, and 53, 55 and 58 were common to any two batches. Some assignments made to one batch only were random false-positive assignments that diminish the level of apparent reproducibility and that would have been rejected during review. Thus a sample evaluated in duplicate by the immunoaffinity approach is expected to show reproducibility at a minimum level of 66% (55/83, Fig. 1c), and duplicate analysis gives a reasonably complete phosphorylation profile, which is not improved significantly by further replicate analysis. We repeated this study but pooled the three batches of phosphotyrosine-containing peptides before analysis, to separate variability due to immunoaffinity isolation from variability due to LC-MS/MS analysis conditions, which can be adjusted for higher levels of reproducibility when deemed necessary. The level of variability associated with analysis alone is essentially the same as the level associated with immunoaffinity isolation and analysis (compare Fig. 1c,d). This demonstrates that LC-MS/MS analysis, and not immunoaffinity isolation, is the major determinant of variability in the immunoaffinity approach.

The results obtained with pervanadate-treated Jurkat cells suggested this approach could be used to identify tyrosine kinases that are abnormally activated in cancer cells and their substrates. To investigate this, we used the immunoaffinity profiling approach to identify phosphotyrosine sites in NIH/3T3 cells expressing constitutively active c-Src Y527F tyrosine kinase (Fig. 2a).

Protein extracts from 3T3-Src cells were digested with trypsin, and phosphotyrosine peptides were purified and analyzed as described above. We found 180 phosphopeptides and 185 phosphotyrosine sites in 3T3-Src. Many phosphopeptides were from the activation loops of protein kinases (Supplementary Table 4 online), suggesting the pathways they control are active in the cell. A complete list of the tyrosine-phosphorylated peptides and proteins found in 3T3-Src is provided as Supplementary Table 5 online.

The immunoaffinity approach identified tyrosine-phosphorylated peptides from proteins that are associated with Src activation. As expected, phosphorylation was detected at Tyr-424 in the activation loop of Src (Supplementary Table 4 online). Tyrosine phosphorylation was also found in known substrates of Src, for example, the nonreceptor tyrosine kinase FAK, the adaptor protein p130Cas, the actin-binding protein cortactin, the phospholipid-binding protein annexin A2, and the STAM-interacting protein Hrs^{17–19}. FAK was phosphorylated at Tyr-397, which creates a binding site for Src-family kinases²⁰, and at its activation loop sites Tyr-576 and Tyr-577, which

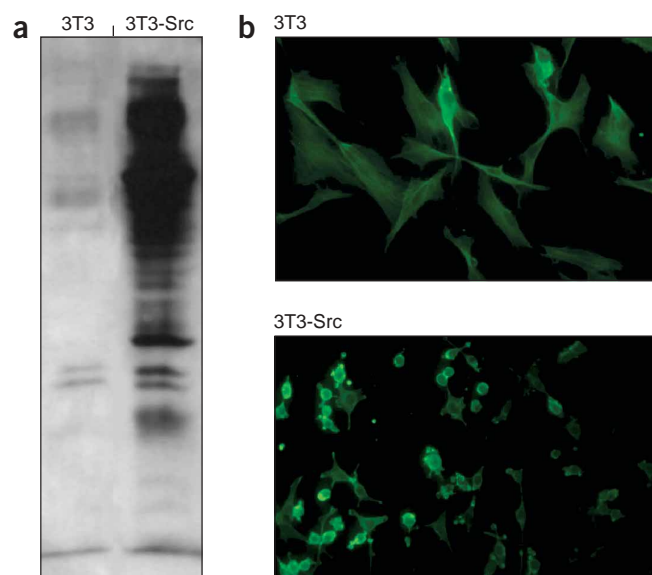


Figure 2 Immunoaffinity profiling of NIH 3T3 cells expressing oncogenic Src Y527F. (a) Western blot using the phosphotyrosine antibody P-Tyr-100 of 3T3 cells and 3T3 cells expressing constitutively active c-Src Y527F. (b) Actin immunofluorescent staining of 3T3 cells and 3T3-Src.

are known to be phosphorylated by Src-family kinases¹⁷. Seven tyrosine phosphorylation sites were found in p130Cas, all in the Src-substrate region of p130Cas and six containing the YXXP motif, which when phosphorylated results in the interaction of p130Cas with several signaling proteins²¹. Some phosphorylation sites map to putative Src substrates. For example, SCAP2 is a homolog of the Src-associated phosphoprotein SKAPP55 and was found to be phosphorylated at Tyr-260, which in SKAPP55 is thought to be phosphorylated by the Src-family kinase Fyn²². We also found tyrosine phosphorylation sites in proteins that are known to be downstream of Src, for example, the actin-binding protein β -catenin and the cytoskeletal proteins paxillin and vinculin²³. Src is known to phosphorylate substrates that induce actin remodeling and cytoskeletal reorganization²⁴. Immunofluorescent staining with actin antibody showed 3T3-Src had a condensed morphology typical of cells undergoing cytoskeletal rearrangement (Fig. 2b). Consistent with this, some tyrosine-phosphorylated peptides originated from cytoskeletal proteins such as the cytoskeletal actins β and γ , the actin binding-protein anillin and the cytoskeletal-associated protein enigma homolog. Applying the immunoaffinity approach to a cell line with a dominant activated kinase resulted in the identification of biologically meaningful phosphorylation sites, including sites from the activated kinase and some of its known substrates and putative substrates.

To further evaluate our approach we investigated two cancer cell lines, Karpas 299 and SU-DHL-1, that are derived from anaplastic large cell lymphomas (ALCL)²⁵. The majority of ALCL is characterized by a chromosomal translocation that fuses the nucleophosmin (NPM) and anaplastic lymphoma kinase (ALK) genes²⁶. Although the two cell lines are derived from different patients, both express the oncogenic fusion kinase NPM-ALK, which possesses constitutive tyrosine kinase activity and can transform nonmalignant cells²⁷.

Tryptic phosphotyrosine peptides were purified and analyzed from extracts of the two ALCL cell lines as described above. As expected,

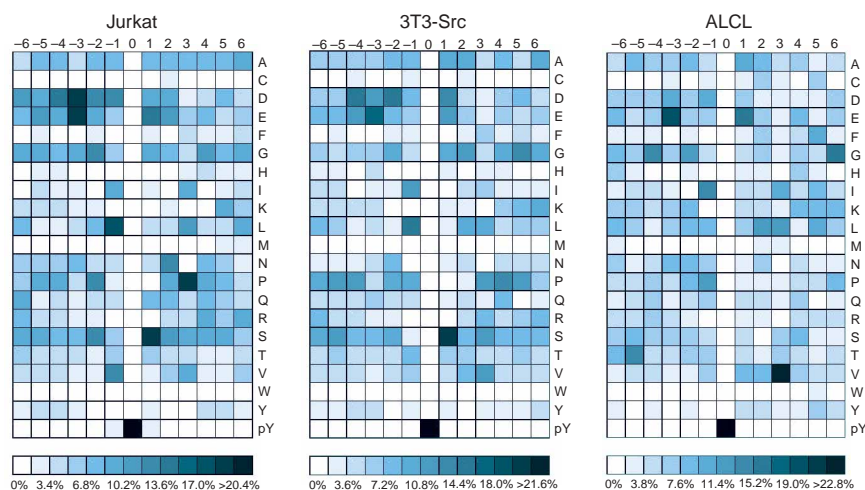


Figure 3 Distribution of amino acid residues surrounding phosphotyrosine among the phosphopeptides identified from three distinct cell types. For each cell type, phosphopeptide sequences were aligned with phosphotyrosine at the residue 0 position, and the frequency of each amino acid at each flanking position was calculated and plotted.

there was large overlap (72%) between the phosphorylation sites found in these two similar cell lines (Table 1 and Supplementary Table 6 and 7 online), and these tyrosine phosphorylation sites were distinct from the ones found in Jurkat cells. Of the 119 tyrosine phosphorylation sites found in SU-DHL-1, 87 were found in Karpas 299 but only 4 were found in Jurkat cells.

To identify more phosphorylation sites, the same SU-DHL-1 cell extract was digested with trypsin, chymotrypsin, endoproteinase GluC or elastase²⁸, and tyrosine-phosphorylated peptides were purified and analyzed as described above, resulting in the identification of 90 phosphotyrosine peptides from the trypsin digest, 59 from chymotrypsin, 46 from endoproteinase GluC and 83 from elastase (see Supplementary Table 8 online). Using a panel of proteases increased the number of distinct tyrosine phosphorylation sites found from 86 using trypsin alone to 180 using all four proteases. Most (36 of 55) phosphorylation sites from the elastase digest were not found in the tryptic digest. Activation loop phosphorylation at Tyr-1282 of ALK was found only in the chymotryptic digest, as a 12-residue peptide; this site would be in a 5-residue tryptic peptide or a 57-residue endoproteinase GluC peptide, both outside the range of peptide lengths amenable to MS/MS-based identification. Furthermore, the protease panel generated overlapping sequences, which confirmed phosphorylation site assignments. The ALK Tyr-1507 site was found in four peptides: two tryptic peptides, one chymotryptic peptide and one elastase peptide.

We found 278 phosphopeptides representing 180 phosphotyrosine sites in SU-DHL-1 cells, and they are grouped according to protein function in Table 1 and Supplementary Table 8 online. In contrast to Jurkat and 3T3-Src where phosphorylation of several tyrosine kinase activation loops was observed, the only tyrosine kinase showing activation-loop phosphorylation in the ALCL cell lines was ALK. Nine sites of ALK phosphorylation, including five new sites, were observed (Supplementary Table 8 online). The four known ALK sites included phosphotyrosine residues that allow NPM-ALK to interact with other signaling proteins such as phospholipase C- γ (via ALK Tyr-1604)²⁹, SHC (via Tyr-1096)²⁷ and IRS-1 (via Tyr-1507)²⁷. Several adaptor proteins were tyrosine-phosphorylated, including

dok2, IRS-1, SHC, Crk and CrkL. SHC was phosphorylated at Tyr-349 and Tyr-350, which are important for protein-protein interactions, and at Tyr-427, which is implicated in Grb2 binding and activation of the Ras/Raf/mitogen-activated protein kinase (MAPK) pathway³⁰.

Among human lymphomas, STAT3 phosphorylation is correlated with ALK expression^{31,32}, suggesting STAT3 phosphorylation may be a secondary marker for this malignancy, and a recent study demonstrated a favorable clinical outcome for the minority of ALCL patients who have tumors expressing ALK but not phosphorylated STAT3³³. We found that both STAT3 isoform 1 and STAT3 isoform 2 were phosphorylated at Tyr-705, which induces dimerization and nuclear translocation of this transcription factor³⁴, in these ALCL cell lines.

The ALCL cell lines differed from other cell lines we examined by containing many tyrosine-phosphorylated peptides from metabolic enzymes such as ATP citrate lyase, G6PDH, IMP dehydrogenase, lactate dehydrogenase A and B, pyruvate kinase-3 and transketolase. We also observed phosphorylation of ribosomal proteins such as L31, P0 and S10 in both cell lines (Supplementary Table 8 online). Nucleophosmin, the ALK fusion partner, is a multifunctional nucleolar protein involved in ribosomal biogenesis³⁵ that can localize the fusion kinase within both the cytoplasmic and nuclear compartments³⁶. It is possible that some ribosomal proteins become phosphorylated because of their affinity for the nucleophosmin domain of the NPM-ALK fusion kinase.

To summarize the tyrosine-phosphorylated peptides that we found in three distinct cell types, we compared the sequences flanking identified phosphotyrosine sites (Fig. 3). The distribution of amino acid residues surrounding phosphotyrosine is distinct for each cell line and likely reflects the substrate recognition motifs of the underlying activated protein kinases. The distribution observed for ALCL is particularly intriguing because it indicates a strong preference for a glutamic acid residue at the -3 position and for a valine residue at +3 and because the substrate recognition motif for ALK is unknown.

In proteomics, protein identification is facilitated by the availability of guidelines for the publication of protein and peptide identification results³⁷ and by the use of statistical methods for validating assigned sequences (listed in ref. 37). Some of these guidelines address specifically the issue of protein identifications that are based on the observation of a single peptide, which are viewed with caution since there is only one experimental result to indicate that the protein is, in fact, present in a sample. However, protein identification proteomics and our immunoaffinity strategy have different expected outcomes. Because the immunoaffinity strategy separates phosphorylated peptides from unphosphorylated peptides, observing just one phosphopeptide from a protein is a common and expected result, since many phosphorylated proteins have only one tyrosine-phosphorylated site.

For this reason, it is appropriate to use additional criteria to validate the results of phosphoproteomic methods that separate phosphorylated peptides from unphosphorylated peptides. Phosphorylation site assignments are probably correct if any of these additional criteria are

Table 1 Phosphotyrosine peptides from signaling proteins in two anaplastic large cell lymphoma cell lines^a

| Protein name ^b | Residue | XCorr ^c | Charge ^d | K ^e | S ^e | Sequence |
|-------------------------------|--------------|--------------------|---------------------|----------------|----------------|---|
| Tyrosine kinase | | | | | | |
| activated p21cdc42Hs kinase | 518 | 5.539 | 3 | | 1 | KPTpYDPVSEDQDPLSSDFK |
| anaplastic lymphoma kinase | 1,078 | 3.712 | 3 | | 1 | HQELQAMQMEQLQSPeYK |
| anaplastic lymphoma kinase | 1,092, 1,096 | 3.018 | 2 | 1 | 1 | TSTIMTDpYNPpYCFAGK |
| anaplastic lymphoma kinase | 1,096 | 4.041 | 3‡ | 3 | 3 | TSTIMTDYNPpYCFAGK |
| anaplastic lymphoma kinase | 1,131 | 4.382 | 3 | 1 | | GLGHGAFGEVpYEGQVSGMPNDPSPLQVAVK |
| anaplastic lymphoma kinase | 1,507 | 3.560 | 3 | 1 | 1 | NKPTSLWNPTpYGSWFTEK |
| anaplastic lymphoma kinase | 1,507 | 4.340 | 3 | | 1 | PTSLWNPTpYGSWFTEKPTK |
| anaplastic lymphoma kinase | 1,507 | 4.258 | 3‡ | 3 | 4 | NKPTSLWNPTpYGSWFTEKPTK |
| anaplastic lymphoma kinase | 1,584, 1,604 | 5.146 | 3 | 3 | 1 | HFPCGNVNPpYGYQQGLPLEAATAPGAGHpYEDTILK |
| anaplastic lymphoma kinase | 1,604 | 4.507 | 3 | 2 | 1 | HFPCGNVNYGYQQGLPLEAATAPGAGHpYEDTILK |
| Janus kinase 3 | 785 | 3.917 | 3 | 1 | 1 | DLNLSISSDpYELLSDPTGALAPR |
| Serine/threonine kinase | | | | | | |
| cdc2 | 15 | 2.805 | 2 | 1 | 3 | IGEGTpYGVVYK |
| cdc2 | 19 | 2.399 | 2 | 1 | 1 | IGEGTYGWPpYK |
| DYRK1A | 145 | 4.698 | 2 | 1 | 1 | VYNDGYDDDpYDYIVK |
| *DYRK1A | 273 | 2.711 | 2 | 2 | 4 | IYQpYQSR |
| DYRK3 | 174 | 1.778 | 3 | | 1 | pYEVLKIIKGSFGQVAR |
| *ERK2 | 187 | 2.828 | 3 | | 1 | VADPDHDTGFLTEpYVATR |
| *GSK3 alpha | 279 | 2.328 | 2‡ | 4 | 5 | GEPNVSpYICSR |
| *HIPK 1 | 352 | 2.847 | 2 | | 1 | AVCSTpYLQSR |
| *p38 alpha MAPK | 182 | 4.080 | 2 | 1 | 1 | HTDDEMTGpYVATR |
| *PRP4K | 849 | 5.008 | 3‡ | 3 | 3 | LCDFGSASHVADNDITPpYLVSr |
| Adaptor | | | | | | |
| CD2-associated protein | 88 | 2.515 | 3 | 1 | | ISTpYGLPAGGIQHPQTK |
| colony-enhancing fact. iso. a | 188 | 4.169 | 2 | 2 | 2 | YLLETSGNLDGLEpYK |
| dok2 | 299 | 3.420 | 2 | 1 | | GQEGEpYAVPFDAVAR |
| HGF reg. tyr. kinase subs. | 216 | 2.999 | 2 | 1 | | VCEPCpYEQLNR |
| insulin receptor substrate 1 | 46 | 2.527 | 2‡ | | 2 | LEpYYENEK |
| insulin receptor substrate 1 | 662 | 2.434 | 3 | | 1 | VDPNGpYMMMSpSGGCSPDIGGGPSSSSSSNAVPSGTSYGK |
| intersectin 2 isoform 1 | 952 | 3.945 | 2‡ | | 2 | REEPEALpYAAVVK |
| Oncogene CBL2 | 674 | 3.234 | 2 | 1 | 1 | IKPSSSANAIpYSLAAR |
| SHC | 349 | 4.242 | 3 | 1 | | MAGFDGSAWDEEEEEPPDHQpYNDFFPGK |
| SHC | 349, 350 | 4.916 | 3 | 2 | 3 | MAGFDGSAWDEEEEEPPDHQpYpYNDFFPGK |
| SHC | 350 | 5.423 | 3 | | 2 | MAGFDGSAWDEEEEEPPDHQpYpYNDFFPGK |
| SHC | 427 | 4.371 | 2 | 2 | 2 | ELFDDPSPpYVNVQNLDK |
| T lymphocyte adaptor | 39 | 3.517 | 3 | 1 | | SCQNLGpYTAASPQAPEAASSTGNAER |
| T lymphocyte adaptor | 216 | 3.141 | 2 | 1 | | SQDPNPQpYSPIIK |
| T lymphocyte adaptor | 305 | 2.927 | 3 | 1 | | GSPGEAPSNIpYVEVEDEGLPATLGHpVLR |
| Wiskott-Aldrich syn. protein | 291 | 5.724 | 2 | 1 | | LIpYDFIEDQGGLEAVR |

^aFor a complete listing of phosphotyrosine peptides found in these samples, see **Supplementary Tables 6 and 7** online. ^b*, activation loop peptides. ^cThe maximum Sequest XCorr value for this assignment in this one analysis. ^dThe charge state associated with the maximum XCorr value for this assignment. ^e‡, in this one analysis the assignment was made to the same sequence in more than one charge state. ^fThe number of times the assignment was made in this one analysis of Karpas 299 cells (K) or SU-DHL-1 cells (S).

met: (i) the same sequence is assigned to coeluting ions with different charge states³⁸; (ii) the site is found in more than one peptide sequence because of overlaps from incomplete proteolysis or use of a panel of proteases (see, e.g., **Supplementary Table 8** online); (iii) the site is found in more than one peptide sequence due to homologous but not identical protein isoforms, for example, in **Supplementary Table 6** online, STAT3 isoform 1 Tyr-705 versus STAT3 isoform 2 Tyr-704 and lactate dehydrogenase A Tyr-238 versus lactate dehydrogenase B Tyr-239; (iv) the site is found in more than one peptide sequence because of homologous but not identical proteins among species, for example, human DEAD-box protein 3 Tyr-103 in **Supplementary Table 2** online versus mouse DEAD-box protein 3 Tyr-103 in **Supplementary Table 5** online; and (v) a synthetic phosphopeptide corresponding to the assigned sequence produces an MS/MS spectrum identical to the one that was used to make the sequence assignment.

We routinely use this last criterion to confirm novel site assignments of particular interest. In practice, we have found that use of a protease panel is the most effective way to increase confidence in phosphorylation site assignments. For example, 48 of the 180 (27%) unique phosphorylation sites reported in **Supplementary Table 8** online were observed in more than one digest.

Our results demonstrate the ability of the immunoaffinity strategy to enhance our understanding of phosphotyrosine-dependent signaling pathways. Simple, sensitive and reproducible methods are established to isolate and identify large sets of tyrosine-phosphorylated peptides from digested cellular extracts. The phosphorylation sites described here are annotated further at PhosphoSite database (<http://www.phosphosite.org/>), a comprehensive resource of human and mouse *in vivo* phosphorylation sites³⁹. Using the PhosphoSite database, we estimate that about 70% of these sites are novel.

Generating pathways and networks from lists of identified tyrosine phosphorylation sites requires linking the sites to upstream and downstream signaling events by further bioinformatics analysis. The immunoaffinity strategy can identify activated tyrosine kinases and phosphorylated substrates in cancer cells and can help identify and elucidate the oncogenic pathways driving disease. When merged with quantitative mass spectrometry methods, the immunoaffinity approach can identify and quantify changes in phosphorylation levels that arise in response to treatment with specific kinase or signaling pathway inhibitors. Finally, the immunoaffinity approach can be extended to antibodies that recognize post-translational modifications other than phosphorylation and to antibodies that recognize phosphorylated-serine and phosphorylated-threonine peptides such as the serine-proline directed MAPK and CDK motifs or basic motifs phosphorylated by the AGC family of protein kinases, allowing focused phosphoproteomic analysis of selected signaling pathways.

METHODS

Cell culture, stimulation and lysis. Cells were grown in a 5% CO₂ incubator at 37 °C. Jurkat, Karpas 299 and SU-DHL-1 cells were cultured in RPMI 1640 medium supplemented with 10% fetal bovine serum (Jurkat) or calf serum (Karpas 299, SU-DHL-1) and penicillin/streptomycin. Jurkat cells cultured to a density of 1.2×10^6 cells/ml were washed in PBS at 20–25 °C and resuspended in PBS at 7×10^7 cells/ml. After preincubation at 37 °C for 20 min, calyculin A and sodium pervanadate were added to final concentrations of 50 ng/ml and 1 mM, respectively, and cells were incubated for 20 min at 37 °C. After centrifugation at 20–25 °C, cells were resuspended at 1.25×10^8 cells/ml in lysis buffer (20 mM HEPES, pH 8.0, 9 M urea, 1 mM sodium vanadate) and sonicated. Karpas 299 and SU-DHL-1 cells cultured to a density of $0.5\text{--}0.8 \times 10^6$ cells/ml were washed with PBS at 4 °C, resuspended in lysis buffer as above, and sonicated. NIH/3T3 cells expressing constitutively active Src were cultured in DMEM medium supplemented with 10% bovine serum and penicillin/streptomycin under selection (1.5 µg/ml puromycin). Cells at about 80% confluency were starved in medium without serum for 3 h. After complete aspiration of medium from the plates, cells were scraped off the plate in 10 ml lysis buffer per 2×10^8 cells (as above, supplemented with 2.5 mM sodium pyrophosphate, 1 mM β-glycerol-phosphate) and sonicated.

Protein digestion. Sonicated cell lysates were cleared by centrifugation at 20,000g, and proteins were reduced with DTT at a final concentration of 4.1 mM and alkylated with iodoacetamide at 8.3 mM. For digestion with trypsin, protein extracts were diluted in 20 mM HEPES pH 8.0 to a final concentration of 2 M urea and immobilized TLCK-trypsin (Pierce) was added at 1–2.5 ml beads (200 TAME units trypsin/ml) per 10^9 cells. For digestion with chymotrypsin, endoproteinase GluC and elastase, lysates were diluted in 20 mM HEPES pH 8.0 to a final concentration of 1 M urea, and GluC (Worthington Biochemicals) or elastase (Roche) was added at 0.5 mg per 10^9 cells. Chymotrypsin (Worthington Biochemicals) was added at 10 mg per 10^9 cells. Digestion was performed for 1–2 d at 20–25 °C.

Reversed-phase solid-phase extraction of digests. Trifluoroacetic acid (TFA) was added to protein digests to a final concentration of 1%, precipitate was removed by centrifugation at 2,000g for 5 min and digests were loaded onto Sep-Pak C18 columns (Waters) equilibrated with 0.1% TFA. A column volume of 0.7–1.0 ml was used per 2×10^8 cells. Columns were washed with 15 volumes of 0.1% TFA, followed by 4 volumes of 5% acetonitrile in 0.1% TFA. Peptide fraction I was obtained by eluting columns with 2 volumes each of 8, 12, and 15% acetonitrile in 0.1% TFA and combining the eluates. Fractions II and III were a combination of eluates after eluting columns with 18%, 22%, 25% acetonitrile in 0.1% TFA and with 30%, 35%, 40% acetonitrile in 0.1% TFA, respectively. All peptide fractions were lyophilized.

Immunoaffinity purification (IAP) of phosphopeptides. Peptides from each fraction corresponding to 2×10^8 cells were dissolved in 1 ml of IAP buffer (20 mM Tris/HCl pH 7.2, 10 mM sodium phosphate, 50 mM NaCl) and

insoluble matter (mainly in peptide fraction III) was removed by centrifugation at 2,000g for 5 min. IAP was performed on each peptide fraction separately. The phosphotyrosine monoclonal antibody P-Tyr-100 (Cell Signaling Technology) from ascites fluid was coupled noncovalently to protein G agarose (Roche) at 4 mg/ml beads by incubation overnight at 4 °C with gentle shaking. After coupling, antibody resin was washed twice with PBS and three times with IAP buffer (5–10 bead volumes of buffer for each wash). Efficient coupling was verified by boiling an aliquot of antibody resin in SDS-PAGE sample buffer for 5 min and evaluating the yield of released antibody against a standard of purified antibody by SDS-PAGE, followed by staining with Coomassie blue. Immobilized antibody (15 µl, 60 µg) was added as 1:1 slurry in IAP buffer to 1 ml of each peptide fraction, and the mixture was incubated overnight at 4 °C with gentle shaking. The immobilized antibody beads were washed three times with 1 ml IAP buffer and twice with 1 ml water, all at 4 °C. Peptides were eluted from beads by incubation with 75 µl of 0.1% TFA at 20–25 °C for 10 min. The antibody resin is not reusable since this elution step also releases the antibody from protein G.

Analysis by LC-MS/MS. Peptides in the IAP eluate (40 µl) were concentrated and separated from eluted antibody using 0.2 µl StageTips⁴⁰. Peptides were eluted from the microcolumns with 1 µl of 40% acetonitrile, 0.1% TFA (fractions I and II) or 1 µl of 60% acetonitrile, 0.1% TFA (fraction III) into 7.6 µl of 0.4% acetic acid/0.005% heptafluorobutyric acid (HFBA). This sample was loaded onto a 10 cm × 75 µm PicoFrit capillary column (New Objective) packed with Magic C18 AQ reversed-phase resin (Michrom Biosources) using a Famos autosampler with an inert sample injection valve (Dionex). The column was then developed with a 45-min linear gradient of acetonitrile in 0.4% acetic acid, 0.005% HFBA delivered at 280 nl/min (Ultimate, Dionex). Tandem mass spectra were collected in a data-dependent manner with an LCQ Deca XP Plus ion trap mass spectrometer (ThermoFinnigan), using a top-four method, a dynamic exclusion repeat count of 1 and a repeat duration of 0.5 min.

Assigning peptide sequences using Sequest. MS/MS spectra were evaluated using TurboSequest in the Sequest Browser package (v. 27, rev. 12) supplied as part of BioWorks 3.0 (ThermoFinnigan). Individual MS/MS spectra were extracted from the raw data file using the Sequest Browser program CreateDta, with the following settings: bottom MW, 700; top MW, 4,500; minimum number of ions, 20; minimum total ion current, 4×10^5 ; and precursor charge state, unspecified. Spectra were extracted from the beginning of the raw data file before sample injection to the end of the eluting gradient. The IonQuest and VuDta programs were not used to further select MS/MS spectra for Sequest analysis. MS/MS spectra were evaluated with the following TurboSequest parameters: peptide mass tolerance, 2.5; fragment ion tolerance, 0.0; maximum number of differential amino acids per modification, 4; mass type parent, average; mass type fragment, average; maximum number of internal cleavage sites, 10; neutral losses of water and ammonia from b and y ions were considered in the correlation analysis. Proteolytic enzyme was specified except for spectra collected from elastase digests, since elastase was found to cleave after many amino acid residues including phosphotyrosine (see **Supplementary Table 8** online). Searches were performed against the NCBI mouse protein database (for 3T3 cells) (released on February 25, 2003 and containing 35,726 protein sequences) or against the NCBI human protein database (for all other studies) (released on April 29, 2003 and containing 37,490 protein sequences). Cysteine carboxamidomethylation was specified as a static modification, and phosphorylation was allowed as a variable modification on serine, threonine, and tyrosine residues or on tyrosine residues alone. We found that restricting phosphorylation to tyrosine residues had little effect on the number of phosphorylation sites assigned. For example, for **Supplementary Table 2**, 292 redundant phosphotyrosine-containing sequence assignments were made when phosphorylation was restricted to tyrosine, and 287 of these were still top-ranked when phosphorylation was allowed on serine, threonine and tyrosine residues; of the missing 5 assignments, 3 were second-ranked (DeltaCn of 0.015–0.063), 1 was third-ranked (DeltaCn of 0.039) and 1 was fourth-ranked (DeltaCn of 0.104).

Reviewing assigned peptide sequences. All spectra and all sequence assignments made by Sequest were imported into a relational database written in

FileMaker Pro and MySQL. The database relates records according to the sequence assignment that was top-ranked by Sequest, so that all spectra associated with a particular sequence assignment can be found and compared in a pairwise manner, regardless of their Sequest scores. The spectra are referenced in the database by the same URL links generated by Sequest. Assigned sequences were accepted or rejected following a conservative, two-step process. In the first step, a subset of high-scoring sequence assignments was selected by filtering for XCorr values of at least 1.5 for a charge state of +1, 2.2 for +2, and 3.3 for +3, allowing a maximum RSp value of 10. Each and every spectrum in this subset was manually reviewed, and assignments in this subset were rejected if any of the following criteria were satisfied: (i) the spectrum contained at least one major peak (at least 10% as intense as the most intense ion in the spectrum) that could not be mapped to the assigned sequence as an *a*, *b* or *y* ion, as an ion arising from neutral-loss of water or ammonia from a *b* or *y* ion, or as a multiply protonated ion; (ii) the spectrum did not contain an uninterrupted series of *b* or *y* ions equivalent to at least five residues; or (iii) the sequence was not observed at least three times in all the studies we have conducted (except for overlapping sequences due to incomplete proteolysis or use of proteases other than trypsin). In the second step, we accepted assignments with below-threshold scores if the low-scoring spectrum showed a high degree of similarity to a high-scoring spectrum collected in another study (see **Supplementary Fig. 1** online), which simulates a true reference library-searching strategy. For **Supplementary Table 2**, from an initial subset of 235 high-scoring assignments, step 1 resulted in the rejection of 21 false-positive assignments and step 2 allowed us to accept 78 false-negative assignments, giving a final tally of 292 redundant sequences. All spectra supporting the final list of assigned sequences used to build the tables shown here were reviewed by at least three people to establish their credibility. To evaluate the false-positive rate associated with the scoring thresholds used in the first step of our two-step process, we inverted the sequences in the NCBI human protein database and the NCBI mouse protein database using the perl script db_reverse.pl, written by Roger Moore (available as part of the qscore software package, see <http://www.cityofhope.org/microseq/downloads/qscore.zip>), and used these inverted sequence databases in Sequest searches. The number of above-threshold assignments to unphosphorylated peptides and phosphorylated peptides made against the forward sequence databases and the reverse sequence databases is given as **Supplementary Table 9** online. The overall false-positive assignment rate was 13%, but the majority of false-positive assignments were to unphosphorylated peptides. The overall false-positive assignment rate for phosphorylated peptides was only 1.7% (range of 0.90 to 2.7%).

Synthetic peptides. Peptides labeled with heavy-isotope residues were synthesized at Cell Signaling Technology as described⁴¹. The peptide sequences were: ALK pY 1096, TSTIMTDYNPpYcCFAGK; ALK pY 1507, NKPTSLWNpTYGSwFTEK; cdc2 pY 15, IGEgTpYGVVYK; GSK3 α pY 279, GEPNVSpYcCSR; PRP4K pY 849, LcCDFGSASHVADNDITpYLVSR; STA-T5A pY 694, AVDGpYVKPQIK; and ZAP70 pY 492/493, ALGADDSpYpYTAR; where pY indicates phosphotyrosine, cC is carboxamidomethyl-cysteine and the heavy-isotope residue is underlined. Heavy-isotope residues increased peptide mass by 6 (P, V) or 4 (A). Cysteine-containing synthetic peptides were purified immediately after alkylation with iodoacetamide. All synthetic peptides were purified by reversed-phase high-performance liquid chromatography, and the purified synthetic peptide stocks were quantified by amino acid analysis using a PicoTag station (Waters) for acid hydrolysis and an AccQ-Fluor reagent kit (Waters) for amino acid derivatization.

Note: Supplementary information is available on the Nature Biotechnology website.

ACKNOWLEDGMENTS

We are grateful to Jeffrey Knott, Jason Reynolds and Matthew Prokop for providing synthetic peptides, Sandra Schieferl for the 3T3-*Src* cell line and David Mason for advice on the use of ALCL cell lines. We thank Ross M. Tomaino, Scott A. Gerber and Steven P. Gygi for helping us collect tandem mass spectra in the early stages of this project, and Daniel Schwartz, Sean A. Beausoleil, Rob Duarte and Steven P. Gygi for producing heat maps for the phosphotyrosine peptides we identified. This work was supported by a Small Business Innovation Research phase I grant 1R43CA101106 of

the Innovative Molecular Analysis Technologies program from the National Cancer Institute (J.R.).

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 15 July; accepted 7 October 2004

Published online at <http://www.nature.com/naturebiotechnology/>

- Blume-Jensen, P. & Hunter, T. Oncogenic kinase signalling. *Nature* **411**, 355–365 (2001).
- Sawyers, C.L. Rational therapeutic intervention in cancer: kinases as drug targets. *Curr. Opin. Genet. Dev.* **12**, 111–115 (2002).
- Mann, M. *et al.* Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome. *Trends Biotechnol.* **20**, 261–268 (2002).
- Ficarro, S.B. *et al.* Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. *Nat. Biotechnol.* **20**, 301–305 (2002).
- Ficarro, S. *et al.* Phosphoproteome analysis of capacitated human sperm. Evidence of tyrosine phosphorylation of a kinase-anchoring protein 3 and valosin-containing protein/p97 during capacitation. *J. Biol. Chem.* **278**, 11579–11589 (2003).
- Pandey, A. *et al.* 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. USA* **97**, 179–184 (2000).
- Pandey, A. *et al.* Identification of a novel immunoreceptor tyrosine-based activation motif-containing molecule, STAM2, by mass spectrometry and its involvement in growth factor and cytokine receptor signaling pathways. *J. Biol. Chem.* **275**, 38633–38639 (2000).
- Pandey, A. *et al.* Cloning of a novel phosphotyrosine binding domain containing molecule, Odin, involved in signaling by receptor tyrosine kinases. *Oncogene* **21**, 8029–8036 (2002).
- Salomon, A.R. *et al.* Profiling of tyrosine phosphorylation pathways in human cells using mass spectrometry. *Proc. Natl. Acad. Sci. USA* **100**, 443–448 (2003).
- Brill, L.M. *et al.* Robust phosphoproteomic profiling of tyrosine phosphorylation sites from human T cells using immobilized metal affinity chromatography and tandem mass spectrometry. *Anal. Chem.* **76**, 2763–2772 (2004).
- Marcus, K., Immler, D., Sternberger, J. & Meyer, H.E. Identification of platelet proteins separated by two-dimensional gel electrophoresis and analyzed by matrix assisted laser desorption/ionization-time of flight-mass spectrometry and detection of tyrosine-phosphorylated proteins. *Electrophoresis* **21**, 2622–2636 (2000).
- Quadroni, M. & James, P. Phosphopeptide analysis. *EXS* **88**, 199–213 (2000).
- Schneider, U., Schwenk, H.U. & Bornkamm, G. Characterization of EBV-genome negative “null” and “T” cell lines derived from children with acute lymphoblastic leukemia and leukemic transformed non-Hodgkin lymphoma. *Int. J. Cancer* **19**, 621–626 (1977).
- Srivastava, A.K. & St-Louis, J. Smooth muscle contractility and protein tyrosine phosphorylation. *Mol. Cell. Biochem.* **176**, 47–51 (1997).
- Eng, J.K., McCormack, A.L. & Yates, J.R., III. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5**, 976–989 (1994).
- Kane, L.P., Lin, J. & Weiss, A. Signal transduction by the TCR for antigen. *Curr. Opin. Immunol.* **12**, 242–249 (2000).
- Calalb, M.B., Polte, T.R. & Hanks, S.K. Tyrosine phosphorylation of focal adhesion kinase at sites in the catalytic domain regulates kinase activity: a role for *Src* family kinases. *Mol. Cell. Biol.* **15**, 954–963 (1995).
- Belsches, A.P., Haskell, M.D. & Parsons, S.J. Role of c-*Src* tyrosine kinase in EGF-induced mitogenesis. *Front. Biosci.* **2**, d501–d518 (1997).
- Bache, K.G., Raiborg, C., Mehlum, A., Madhus, I.H. & Stenmark, H. Phosphorylation of Hrs downstream of the epidermal growth factor receptor. *Eur. J. Biochem.* **269**, 3881–3887 (2002).
- Schaller, M.D. *et al.* Autophosphorylation of the focal adhesion kinase, pp125FAK, directs SH2-dependent binding of pp60src. *Mol. Cell. Biol.* **14**, 1680–1688 (1994).
- Bouton, A.H., Riggins, R.B. & Bruce-Staskal, P.J. Functions of the adapter protein Cas: signal convergence and the determination of cellular responses. *Oncogene* **20**, 6448–6458 (2001).
- Curtis, D.J. *et al.* Adaptor protein SKAP55R is associated with myeloid differentiation and growth arrest. *Exp. Hematol.* **28**, 1250–1259 (2000).
- Wang, Y.H., Li, F., Schwartz, J.H., Flint, P.J. & Borkan, S.C. c-*Src* and HSP72 interact in ATP-depleted renal epithelial cells. *Am. J. Physiol. Cell Physiol.* **281**, C1667–C1675 (2001).
- Frame, M.C., Fincham, V.J., Carragher, N.O. & Wyke, J.A. v-*Src*'s hold over actin and cell adhesions. *Nat. Rev. Mol. Cell Biol.* **3**, 233–245 (2002).
- Pulford, K. *et al.* Detection of anaplastic lymphoma kinase (ALK) and nucleolar protein nucleophosmin (NPM)-ALK proteins in normal and neoplastic cells with the monoclonal antibody ALK1. *Blood* **89**, 1394–1404 (1997).
- Morris, S.W. *et al.* Fusion of a kinase gene, ALK, to a nucleolar protein gene, NPM, in non-Hodgkin's lymphoma. *Science* **263**, 1281–1284 (1994).
- Fujimoto, J. *et al.* Characterization of the transforming activity of p80, a hyperphosphorylated protein in a Ki-1 lymphoma cell line with chromosomal translocation t(2;5). *Proc. Natl. Acad. Sci. USA* **93**, 4181–4186 (1996).

28. MacCoss, M.J. *et al.* Shotgun identification of protein modifications from protein complexes and lens tissue. *Proc. Natl. Acad. Sci. USA* **99**, 7900–7905 (2002).
29. Bai, R.Y., Dieter, P., Peschel, C., Morris, S.W. & Duyster, J. Nucleophosmin-anaplastic lymphoma kinase of large-cell anaplastic lymphoma is a constitutively active tyrosine kinase that utilizes phospholipase C- γ to mediate its mitogenicity. *Mol. Cell. Biol.* **18**, 6951–6961 (1998).
30. van der Geer, P., Wiley, S., Gish, G.D. & Pawson, T. The Shc adaptor protein is highly phosphorylated at conserved, twin tyrosine residues (Y239/240) that mediate protein-protein interactions. *Curr. Biol.* **6**, 1435–1444 (1996).
31. Zamo, A. *et al.* Anaplastic lymphoma kinase (ALK) activates Stat3 and protects hematopoietic cells from cell death. *Oncogene* **21**, 1038–1047 (2002).
32. Zhang, Q. *et al.* Multilevel dysregulation of STAT3 activation in anaplastic lymphoma kinase-positive T/null-cell lymphoma. *J. Immunol.* **168**, 466–474 (2002).
33. Khoury, J.D. *et al.* Differential expression and clinical significance of tyrosine-phosphorylated STAT3 in ALK(+) and ALK(-) anaplastic large cell lymphoma. *Clin. Cancer Res.* **9**, 3692–3699 (2003).
34. Darnell, J.E. Jr., Kerr, I.M. & Stark, G.R. Jak-STAT pathways and transcriptional activation in response to IFNs and other extracellular signaling proteins. *Science* **264**, 1415–1421 (1994).
35. Szebeni, A., Hingorani, K., Negi, S. & Olson, M.O. Role of protein kinase CK2 phosphorylation in the molecular chaperone activity of nucleolar protein b23. *J. Biol. Chem.* **278**, 9107–9115 (2003).
36. Bischof, D., Pulford, K., Mason, D.Y. & Morris, S.W. Role of the nucleophosmin (NPM) portion of the non-Hodgkin's lymphoma-associated NPM-anaplastic lymphoma kinase fusion protein in oncogenesis. *Mol. Cell. Biol.* **17**, 2312–2325 (1997).
37. Carr, S. *et al.* The need for guidelines in publication of peptide and protein identification data: working group on publication guidelines for peptide and protein identification data. *Mol. Cell. Proteomics* **3**, 531–533 (2004).
38. Nesvizhskii, A.I., Keller, A., Kolker, E. & Aebersold, R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* **75**, 4646–4658 (2003).
39. Hornbeck, P.V., Chabra, I., Kornhauser, J.M., Skrzypek, E. & Zhang, B. PhosphoSite: A bioinformatics resource dedicated to physiological protein phosphorylation. *Proteomics* **4**, 1551–1561 (2004).
40. Rappsilber, J., Ishihama, Y. & Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **75**, 663–670 (2003).
41. Gerber, S.A., Rush, J., Stemman, O., Kirschner, M.W. & Gygi, S.P. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc. Natl. Acad. Sci. USA* **100**, 6940–6945 (2003).