Systematic evaluation of quantotypic peptides for targeted analysis of the human kinome

Jonathan D. Worboys[1], **John Sinclair**[1], **Yinyin Yuan**[2], and **Claus Jørgensen**[1,3]

[1]Division of Cancer Biology, The Institute of Cancer Research, London, UK

[2]Division of Molecular Pathology, The Institute of Cancer Research, London, UK

Abstract

In targeted proteomics it is critical that peptides are not only proteotypic, but also accurately represent the level of the protein (quantotypic). Numerous approaches are used to identify proteotypic peptides, but quantotypic properties are rarely assessed. Here, we show that measuring ratios of proteotypic peptides across biological samples can be used to empirically identify peptides with good quantotypic properties, and use this to identify quantotypic peptides for 21% of the human kinome.

Introduction and results

Selected Reaction Monitoring (SRM) is an attractive method for accurate quantification of proteins by mass spectrometry (MS) of complex samples[1,2]. This approach is highly sensitive (low attomole levels), offers a broad dynamic range (five orders of magnitude), as well as excellent analytical reproducibility[3,4]. Individual peptides that are both detectable and unique to the protein of interest (proteotypic peptides[5]) are selected, and combinations of precursor and fragment masses (transitions) are measured on a triple quadrupole mass spectrometer. The identification of proteotypic peptides has been facilitated by proteomics repositories such as PRIDE, PeptideAtlas and GPM[4,6-9]. Where no prior information is available, proteotypic peptides are typically predicted and synthesised[10,11]. While these approaches have been widely used, Stergachis _et. al._ recently showed that optimal proteotypic peptides could only be defined by empirically evaluating all _in silico_ predicted peptides across the entire protein coding sequence[12].

The underlying assumption for protein quantification in bottom up proteomics is that the level of the measured peptide(s) is stoichiometric to the level of the protein (quantotypic). Several factors may impact the quantotypic properties of peptides such as differential post-

translational modification, alternative splicing and the completeness of proteolytic digestion. Selection of optimal quantotypic peptides is crucial to ensure accurate quantification of protein levels. However, due to an incomplete understanding of the elements that impact peptide quantotypic behaviour, there are currently only limited guidelines for predicting quantotypic peptides[4,13]. Importantly, since synthetic peptides and proteins do not recapitulate the complexity of post-transcriptional and translational modifications observed *in vivo*, these may not be optimal when evaluating quantotypic peptide properties.

To develop a high confidence SRM assay for the human kinome, we set out to systematically identify proteotypic peptides and empirically evaluate their quantotypic properties. To ensure our assay covers the complexity of post-transcriptional and translational modifications observed *in vivo*, we enriched and identified endogenously expressed protein kinases using discovery-based MS followed by SRM assessment of all *in silico* predicted non-modified tryptic peptides (see Fig. 1a for workflow). Initially we sought to identify expressed protein kinases across a panel of six cell lines. Using ActivX nucleotide analogues (desthiobiotin-ATP and -ADP), we enriched nucleotide binding proteins and identified isolated proteins by data-dependent analysis on an Orbitrap Velos[14,15]. This led to the accumulated identification of 219 protein kinases (Fig. 1b and Supplementary Fig. 1), covering 42% of the human kinome (Fig. 1d). To evaluate the proteotypic properties of all tryptic peptides for the identified kinases, we *in silico* digested the entire protein coding sequence and evaluated the intensity of all SRM transitions on a triple quadrupole mass spectrometer[12] (Online Methods). To increase sensitivity during this stage of assay development, we evaluated protein kinases from enriched samples where the discovery analysis provided high sequence coverage. In total, we evaluated 35954 transitions (y-series ions) targeting 5806 peptides across 208 protein kinases (Fig. 1c and Online Methods). Due to the high number of peptides that were evaluated, we predicted the retention time of each peptide using SSRCalc 3.0[16], which facilitated the evaluation of close to 1000 transitions in a single MS analysis. In total, this led to the identification of 4375 transitions for 1820 peptides covering 207 protein kinases (Supplementary Table 1). Subsequent filtering for sequence uniqueness further reduced this to 790 proteotypic peptides targeting 196 protein kinases, covering 37% of the human kinome (Fig. 1d), where 132 (25%) were covered by three or more proteotypic peptides. All proteotypic peptides, from the 132 kinases, were subsequently validated with synthetic counterparts where 453 of 466 peptides displayed a Pearson correlation >0.8 and ΔiRT <10 (<1.5 minutes), validating 97% of the peptides (Online Methods, Supplementary Fig. 2 and Supplementary Table 2). To our knowledge this represents the highest coverage of the human kinome by SRM to date.

To empirically assess the quantotypic properties of all proteotypic peptides, we devised an easily implemented workflow building on the use of endogenous proteins for assay development. Since quantotypic peptides are stoichiometric to the level of the protein, we reasoned that the relative level between proteotypic peptides from the same protein could be used to empirically assess their quantotypic properties. Any modification of a peptide will result in a net decrease in the level of the unmodified version. Therefore, within each protein, the ratio between any two quantotypic peptides will be constant across multiple

biological conditions. Conversely, if a peptide is differentially modified across samples, the ratio to other proteotypic peptides changes. Since peptides are compared in a pairwise manner within each protein, this approach is tolerant to differences in total protein levels and is more robust when differences exist.

To demonstrate this approach, we determined the relative ratio between all proteotypic peptides within each protein kinase with three or more validated proteotypic peptides, across all six cell lines (Fig. 2a). Using ActivX enriched samples we determined the level of 412 peptides from 109 protein kinases and calculated their pairwise ratios. Subsequently, we determined the correlation between the ratios of individual pairs of peptides across the cell lines using Pearson correlation. This analysis showed a high degree of correlation (P <0.05) between 80% of peptide pairs across these 6 cell lines (Fig. 2b, Group 1). To rigorously test the performance of the identified quantotypic peptides, we further evaluated their behaviour across an independent panel of six additional cell lines (Group 2, Online Methods). Lysates were enriched using ActivX and the relative ratio of all validated proteotypic peptides were compared in a pairwise manner. This analysis identified 80% of peptides to have good quantotypic properties across the 6 cell lines in group 2 (Supplementary Fig. 3a). Importantly, 72% of all quantotypic peptides from group 1 were validated in group 2, with an AUC from a ROC analysis of 0.71 (Supplementary Fig. 3b).

As this analysis identified 80% of the proteotypic peptides to have good quantotypic properties in both groups 1 and 2, we assessed whether increasing the number of samples included in the analysis (in this case cell lines) affects the ability to identify quantotypic peptides. Firstly we fitted a smoothing spline model to the data from each group (Fig. 2b and Supplementary Fig. 3a), which predicts that increasing the number of samples will identify over 90% as quantotypic. Subsequently, we repeated the correlation analysis across both groups (P <0.05, Fig. 2c and Supplementary Fig. 4)**.** This analysis shows that inclusion of 10 or more samples facilitates the identification of ~95% of peptides with good quantotypic properties, thus confirming the previous estimation. In total, this led to the identification of quantotypic peptides for 107 protein kinases (Figure 2d). We hypothesise that the high fractions of identified quantotypic peptides is due to the use of endogenously expressed proteins for evaluating proteotypic properties, thus poor performing peptides will have been excluded in the initial evaluation.

Since protein kinases are regulated by phosphorylation, we subsequently determined whether growth factor stimulation influences the behaviour of assigned quantotypic peptides. To evaluate the effect of early as well as late signalling events, cells were stimulated with IGF or EGF for 5, 10 or 30 minutes and all quantotypic peptides were assessed in a pairwise manner across all samples (Supplementary Fig. 5). Interestingly, we observed that only 4 of the 406 (<1%) measured peptides displayed significantly different levels (described in Online Methods) under these conditions (Supplementary Information). As such, the empirically identified quantotypic peptides are minimally affected by growth factor stimulation.

To evaluate whether the SRM assay would be sufficiently sensitive to quantify protein kinases directly from total cell lysates, we prepared lysates from six cell lines, separated

these by SDS-PAGE (Supplementary Fig. 6) and conducted a SRM experiment in biological and technical replicates. In total, we analysed 204 quantotypic peptides from 83 protein kinases across the human kinome and the relative amount of each peptide was calculated (Fig. 3, Supplementary Fig. 7 and Online Methods). Since the most intense proteotypic peptides from individual proteins can be used as representative of the relative protein level[17], we averaged the level of quantified peptides for each kinase and clustered their relative expression level across cell lines (Supplementary Fig. 8). This revealed groups of low and high abundance kinases as well as a third group of kinases with distinct expression levels between individual cell lines. Overall this demonstrates that targeted analysis directly from gel-separated cell lysates is achievable and highly reproducible (Supplementary Fig. 7).

In summary, we empirically evaluate proteotypic and quantotypic behaviour of peptides for targeted analysis across the human kinome and show that endogenously expressed proteins can serve as a practical source for SRM assay generation and optimisation. This approach can be easily implemented across different classes of proteins and importantly facilitates evaluation of the quantotypic behaviour of peptides.

Online Methods

Cell Culture

All cell lines used were obtained from ATCC and were AsPC-1, HPAC, MiaPaCa2, PANC-1, PL45 and PL5 (group 1) and BxPC-3, Capan-2, CFPAC-1, HPAF-II, Panc 10.05 and SW-1990 (group 2). Cells were cultured in Dulbecco's Modified Eagle Medium (DMEM, Life Technology), supplemented with 10% heat-inactivated Fetal Bovine Serum (FBS, Sigma) and 1× Antibiotic Antimycotic Solution (Hyclone) at 37°C, 5% CO2. All cells were grown to 50% confluence before harvesting. All cell lines were mycoplasma negative and periodically checked.

Cell Lysis

Cells were lysed using PLC buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1.5 mM MgCl$_2$, 1 mM EDTA, 10 mM NaPPi, 10% glycerol and 1% Triton X-100, 1 mM PMSF, 1 mM vanadate with protease inhibitor cocktail (Sigma) and phosphatase inhibitor cocktail (Sigma)) on ice. Lysates were collected and vortexed to ensure complete lysis and cleared by centrifugation at 4°C for 15 minutes at 16,000 g. The concentration of all lysates was determined by bicinchoninic assay (BCA, Thermo Scientific).

Western blotting

Equal amounts of total protein were prepared in 1× SDS sample buffer (10 mM Tris-HCl pH 7.4, 10 mM EDTA, 10% glycerol, 1% SDS and 1 mM DTT), and separated on 10% polyacrylamide gels. Proteins were transferred to a nitrocellulose membrane (LI-COR) by a wet transfer system (Bio-Rad), and blocked in 1× Roti-Block (Carl Roth). Primary antibodies used were against GAPDH (Santa Cruz, FL-335, sc-25778) and P-Tyr-1000 (Cell Signalling Technology (CST), #8954). Secondary antibodies used were anti-mouse IgG

DyLight® 680 (CST, #5470) and anti-rabbit IgG DyLight® 800 (CST, #5151). Blots were visualised fluorescently using a LI-COR Odyssey imaging system.

Kinase Enrichment with ATP/ADP probes

Cell lysates were enriched for kinases using ActivX Desthiobiotin-ATP and -ADP probes (Thermo Scientific), essentially according to manufacturers' instructions. Briefly, cell lysates were desalted using Zeba spin desalting columns (7K MWCO, 5ml, Thermo Scientific) to remove endogenous ATP. Lysates were eluted with reaction buffer (20 mM HEPES pH7.4, 150 mM NaCl, 0.1% Triton X-100 supplemented with protease inhibitors (Sigma)). Protein concentration was determined using a BCA assay (Thermo Scientific) and further diluted to a final concentration of 2 mg/ml. For labelling with the ActivX probes, 1 mg of cell lysate was adjusted to 2 mM $MgCl_2$ and incubated with 20 μM of ActivX probe in a final volume of 500 μl for 30 minutes, at room temperature. Following, 500 μl Urea lysis buffer (8 M Urea, 5 mM Tris-HCl pH7.4, 150 mM NaCl, 1 mM EDTA, 1% NP-40, 5% glycerol) was added to the lysate to stop the reaction. Samples were then incubated with 25 μl high capacity streptavadin agarose resin (Thermo Scientific) for 1 hour at room temperature. Beads were collected by centrifugation at 3000 rpm for 30 seconds, washed with 800 μl 4 M Urea lysis buffer, three times, and boiled in 3× SDS sample buffer.

In-gel digestion

All gel electrophoresis for Mass Spectrometric analysis was carried out using pre-cast any kD mini-PROTEAN gels (Bio-Rad). Proteins were visualised using GelCode™ blue staining (Thermo Scientific) and the gel was processed for mass spectrometry analysis using in-gel digestion. Specifically, each lane was cut into either 10 slices for discovery analysis, or specific MW regions for SRM, and placed into individual low-binding microcentrifuge tubes (Sigma). Each gel-band was then washed three times in 50% (v/v) Acetonitrile (MeCN) for 10 minutes, and dried under vacuum in a Savant SC250 express speedvac concentrator (Thermo Scientific) for 10 minutes. The dried gel bands were reduced in 10 mM dithiothreitol (DTT), 5 mM ammonium bicarbonate (AmBic) pH8 for 45 minutes at 50°C followed by alkylation in 50 mM iodoacetamide (IAA) in 5 mM AmBic for 1 hour at room temperature, in the dark. Gel pieces were subsequently washed three times with 50% MeCN and dried under vacuum for 10 minutes. Proteins were digested with 100 ng sequence grade trypsin (Promega) in 5 mM AmBic for 18 hours at 37°C. Following this, tubes were briefly centrifuged and peptides were extracted with 100 μl 50% MeCN (v/v), 5% Trifluoroacetic acid (v/v) three times. Extracted peptides were pooled in a new microcentrifuge tube, dried under vacuum, resuspended in 0.1% formic acid (FA) and analysed by Liquid Chromatography – Mass Spectrometry (LC-MS).

Mass spectrometry

Discovery-based analysis was conducted on a LTQ Orbitrap Velos mass spectrometer (Thermo Scientific) coupled to a NanoLC-Ultra 2D with a cHiPLC-Nanoflex chromatography system (Eksigent). Chromatographic separation was carried out on a 200 μm i.d. × 0.5 mm trap column packed with C18 (3 μm bead size, 120 Å, Eksigent), a 75 μm i.d. × 15 cm column packed with C18 (3 μm bead size, 120 Å, Eksigent) with a linear gradient of 5-50% solvent B (Acetonitrile, 0.1% Formic Acid) against solvent A (H_2O, 0.1%

Formic Acid) with a flow rate of 300 nl/min. The mass spectrometer was operated in a data-dependent mode to automatically switch between Orbitrap MS and ion trap MS/MS acquisition. Survey full scan MS spectra (from *m/z* 375-2,000) were acquired in the Orbitrap with a resolution of 60,000 at *m/z* 400 and FT target value of 1×10^6 ions. The 20 most abundant ions were selected for fragmentation and dynamically excluded for 8 sec. The lock mass option was enabled using the polydimethylcyclosiloxane ion (*m/z* 445.120025) as an internal calibrant. For peptide identification, raw data files produced in the Xcalibur software (Thermo Scientific) were processed in Proteome Discoverer V1.3 (Thermo Scientific) and searched using Mascot (v2.2) against Swissprot human database (04/2013, 89601 entries). Searches were performed with a precursor mass tolerance set to 10 ppm, fragment mass tolerance set to 0.8 Da and a maximum number of missed cleavages set to 2. Static modifications were limited to carbamidomethylation of cysteine, and variable modifications searched were oxidation of methionine and deamidation of asparagine and glutamine residues. Peptides were filtered using a mascot significance threshold < 0.05, peptide score > 20 and FDR < 0.01 (evaluated by Percolator[18]). Proteins were assigned by a minimum of one unique peptide. The mass spectrometry proteomics data have been deposited to the ProteomeXchange consortium[19] via the PRIDE partner repository with the dataset identifier PXD001026.

Targeted analysis was conducted on a TSQ Vantage triple quadrupole mass spectrometer (Thermo Scientific) coupled to a NanoLC-Ultra 1D with a cHiPLC-Nanoflex chromatography system (Eksigent). Reversed-phase chromatographic separation was carried out as for discovery-based analysis. The mass spectrometer was operated with a Q1 unit resolution of 0.4 Th and a Q3 0.7 Th. Q2 was operated at 1.5 mTorr with predicted collision energies for each peptide[20]. Each transition had a minimum dwell time of 20 ms, with cycle times of 2.2 s. The raw data files produced in Xcalibur software (Thermo Scientific) were analysed using Skyline[21]. We used the extracted ion chromatograms for the 2 most intense transitions (primary transitions) to determine the peptide abundance. These were summed together to get an area per peptide, and these areas were summed with all peptides per protein to acquire final protein areas. The data have been deposited to the PeptideAtlas SRM Experiment Library (PASSEL)[22] with the dataset identifier PASS00531.

In silico digestions and empirical evaluation of SRM peptides

Proteins identified from discovery analysis were filtered to exclude those above 350 kDa, and keeping only the longest isoforms where multiple isoforms were identified. In total, 208 protein kinases were digested *in silico* using the Skyline software[21]. For our SRM analysis, we monitored all possible fully tryptic, doubly charged peptides that were between 6 and 20 amino acids in length. Peptides that contained a methionine were excluded and all cysteine residues were considered to be carbamidomethylated. For each peptide we monitored all singly charged fragment ions from y2 to the last y ion −1, with a m/z ratio between 300-1,500. Retention times for each peptide were predicted using SSRCalc 3.0 (ref. 16). Peptides were considered detected when all fragment ions co-eluted, with at least 3 data points across the peak, and a signal to noise of at least 7 (ref. 3). We were unable to identify unique peptides with sufficiently high number of transitions and a high signal/noise for 12

protein kinases. This is due, mainly, to their shorter coding region and high homology within the protein kinase family.

Validation of proteotypic peptides

Synthetic versions were acquired for 466 of the proteotypic peptides (Thermo Scientific Biopolymers), and the relative ion distribution was used for validation with a Pearson correlation > 0.8 as cut-off. In addition, relative elution time was assigned for all peptides using a standard peptide mix (Thermo Scientific, #88320, iRT values[23]) for calibration. Overall the measured iRT for synthetic peptides and peptides from ActivX enriched samples was offset from 0 to 3.5 iRT values, which likely reflects a difference in the sample matrix. As such, peptides deviating below 10 iRT values from the mean (equivalent to a 3 minute window) were considered validated. The use of a commercially available standard to determine iRT values facilitates the use of the presented peptides across different chromatographic setups. All results are provided (Supplementary Table 2).

Statistical analysis of quantotypic peptides

Correlation analysis was performed with two-sided Pearson's product moment correlation on total, log-normalised, area of peptides. *P*-values were estimated from the correlation coefficients with a t-distribution with degrees of freedom $n - 2$, where n denotes the number of cell lines used for analysis. Correlation was termed significant if $P < 0.05$, that is, absolute correlation with a critical value > 0.707 in the case of 6 cell lines. For the extrapolation analysis, we first calculated the fraction of quantotypic peptides based on the correlation when number of cell lines $n = 3\text{-}6$ and $n = 3\text{-}12$ (Fig. 2b-c, respectively), each with five resampling without replacement. Next, a cubic smoothing spline model was fitted to the median of data. The degree of freedom for the smoothing spline was chosen to minimise the penalised cross validation error of the model and avoid overfitting (see Supplementary Software). Finally, a smoothing spline with the optimal degree of freedom was fitted to the data and subsequently predicted values for $n = 7\text{-}10$, in the case of 6 cell lines, and $n = 13\text{-}16$, in the case of 12 cell lines.

To reflect differences in numbers of peptides analysed per protein and differences in peptide performance within each protein we assigned a protein confidence score and ranked peptides. Protein confidence scores were calculated by dividing the number of correlated peptides cubed by the total number of peptides monitored (see Supplementary Software). As such, the score is greater where more peptides were monitored, but penalised when more peptides do not correlate. Where no peptides correlate zeroes are assigned. Peptide ranking was assigned based on the multiple intra-protein peptide correlations. As such, all calculated correlations for each peptide were averaged and used for ranking (i.e. peptides that correlate better with other peptides from the same protein are ranked higher). All scores and ranks are provided (Supplementary Tables 3-5).

All analysis and graphics are produced with the statistical package R (R Development Core Team, 2012. http://www.R-project.org/). We provide the full R-script for the analysis of quantotypic properties of peptides, with the full dataset. An associated PDF file, which is

fully annotated to explain and illustrate the full details of the analysis performed, is included and can also be found online (http://yuanlab.org/software/QP/).

Growth factor stimulation

MiaPaCa2 cells were grown to 50% confluence and either left untreated or stimulated with 100 ng/ml EGF or IGF for 5, 10 and 30 minutes (Supplementary Fig. 5). Samples were then enriched with ActivX and all quantotypic peptides analysed. As the multiple correlation analysis for identification of quantotypic peptides relies on differences in protein levels and this analysis was only conducted across a single cell line, peptide quantotypic behaviour cannot be calculated as described above. Therefore, to identify peptides affected by growth factor stimulation, a multiple linear regression analysis was performed across all experimental conditions. The Cook's distance for each peptide was calculated and used to identify outliers. We used a threshold where the Cook's distance (D) is bigger than the 10th percentile of an F distribution with p and n-p degrees, where p is the number of parameters and n is the number of observations[24], in our case 0.37. This analysis identified 4 peptides (from 406) that provide significant leverage on the regression analysis to be considered as outliers, equating to less than 1%.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

1. Picotti P, Bodenmiller B, Aebersold R. Nat. Methods. 2013; 10:24–27.

2. Gillette MA, Carr SA. Nat. Methods. 2013; 10:28–34. [PubMed: 23269374]

3. Domon B, Aebersold R. Nat. Biotechnol. 2010; 28:710–721. [PubMed: 20622845]

4. Lange V, Picotti P, Domon B, Aebersold R. Mol. Syst. Biol. 2008; 4:222. [PubMed: 18854821]

5. Kuster B, Schirle M, Mallick P, Aebersold R. Nat. Rev. Mol. Cell Biol. 2005; 6:577–583. [PubMed: 15957003]

6. Vizcaino JA, et al. Nucleic Acids Research. 2012; 41:D1063–D1069. [PubMed: 23203882]

7. Craig R, Cortens JP, Beavis RC. J. Proteome Res. 2004; 3:1234–1242. [PubMed: 15595733]

8. Deutsch EW, Lam H, Aebersold R. EMBO Rep. 2008; 9:429–434. [PubMed: 18451766]

9. Picotti P, et al. Nat. Methods. 2008; 5:913–914. [PubMed: 18974732]

10. Mallick P, et al. Nat. Biotechnol. 2006; 25:125–131. [PubMed: 17195840]

11. Picotti P, et al. Nat. Methods. 2009; 7:43–46. [PubMed: 19966807]

12. Stergachis AB, MacLean B, Lee K, Stamatoyannopoulos JA, MacCoss MJ. Nat. Methods. 2011; 8:1041–1043. [PubMed: 22056677]

13. Brownridge P, et al. Proteomics. 2011; 11:2957–2970. [PubMed: 21710569]

14. Patricelli MP, et al. Biochemistry. 2007; 46:350–358. [PubMed: 17209545]

15. Patricelli MP, et al. Chem. Biol. 2011; 18:699–710. [PubMed: 21700206]

16. Krokhin OV. Mol. & Cell. Proteomics. 2004; 3:908–919.

17. Ludwig C, Claassen M, Schmidt A, Aebersold R. Mol. & Cell. Proteomics. 2012; 11 M111.013987.

18. Käll L, Canterbury JD, Weston J, Noble WS, MacCoss MJ. Nat. Methods. 2007; 4:923–925. [PubMed: 17952086]

19. Vizcaíno JA, et al. Nat. Biotechnol. 2014; 32:223–226. [PubMed: 24727771]

20. MacLean B, et al. Anal. Chem. 2010; 82:10116–10124. [PubMed: 21090646]

21. MacLean B, et al. Bioinformatics. 2010; 26:966–968. [PubMed: 20147306]

22. Farrah T, et al. Proteomics. 2012; 12:1170–1175. [PubMed: 22318887]

23. Escher C, et al. Proteomics. 2012; 12:1111–1121. [PubMed: 22577012]

24. Cook, RD.; Weisberg, S. Chapman and Hall; New York: 1982.
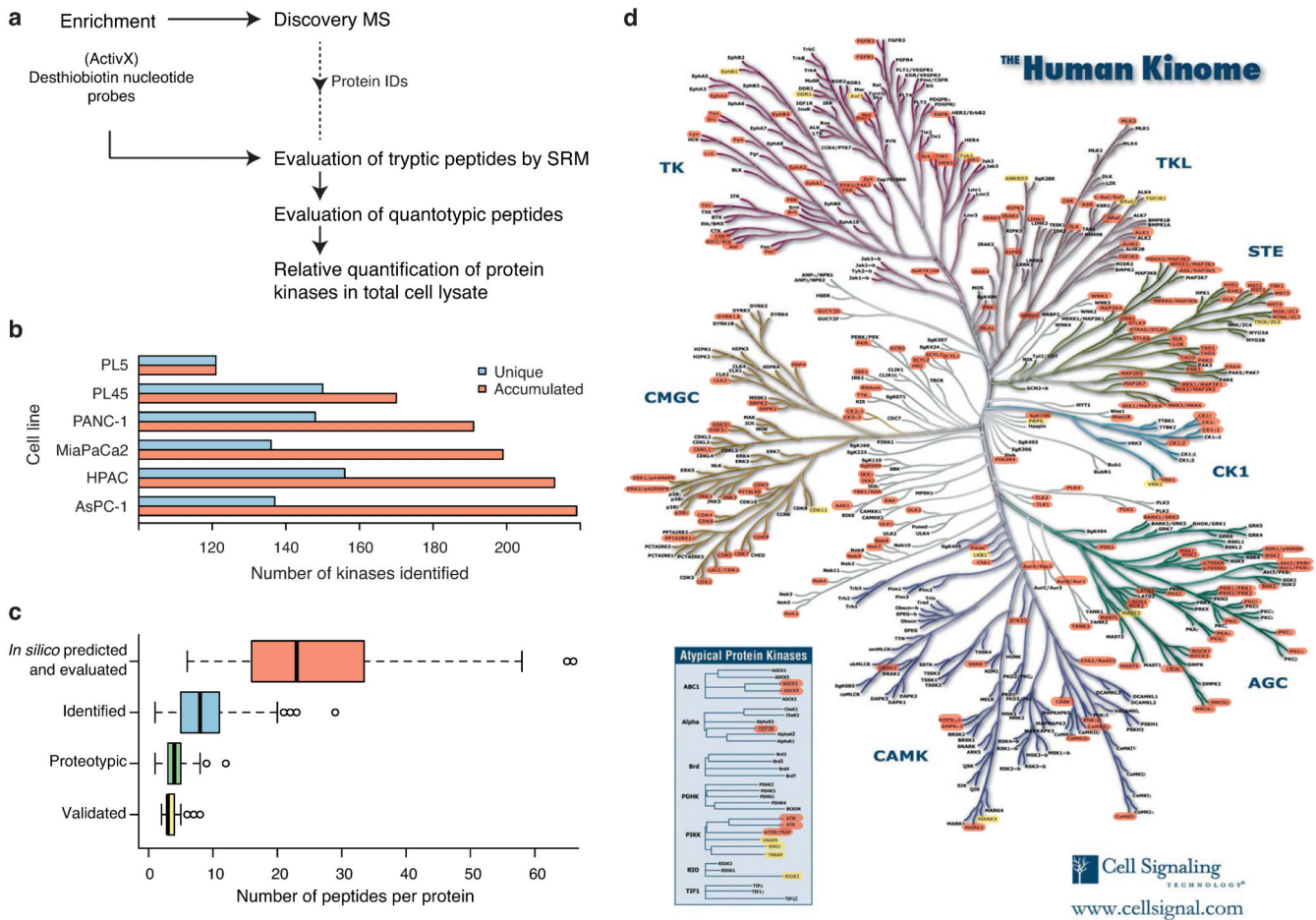
Figure 1. Development of a targeted proteomic assay for human protein kinases.
(**a**) Overview of the workflow combining enrichment and identification of nucleotide binding proteins with subsequent evaluation of proteotypic and quantotypic peptides. (**b**) Bar graph showing the unique and accumulated number of identified kinases across six cell lines. (**c**) Box plots displaying the number of evaluated peptides per protein including total number of *in silico* digested peptides evaluated by SRM, all peptides identified by SRM, assigned proteotypic peptides, and proteotypic peptides validated by synthetic counterparts. For each box, centre lines show the medians; box limits indicate the 25th and 75th percentiles, whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles and outliers are represented by circles. (**d**) Coverage of the human kinome. Kinases with successfully identified proteotypic peptides are shown in red. Kinases only detected by discovery MS are shown in yellow. The use of the kinome tree is reproduced courtesy of Cell Signaling Technology, Inc. (www.cellsignal.com).
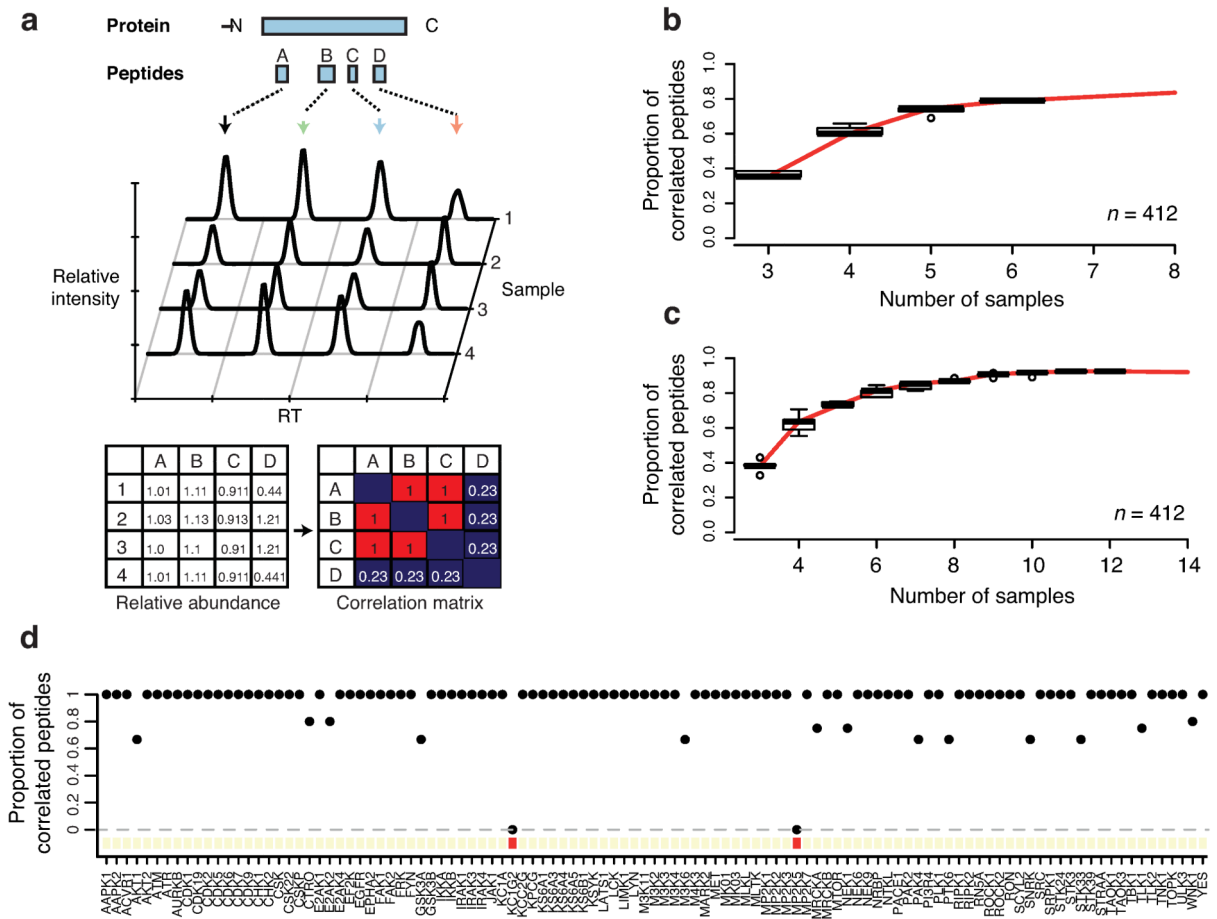
Figure 2. Systematic evaluation of quantotypic peptide properties.
(**a**) Diagram representing the approach underlying quantotypic evaluation. Multiple proteotypic peptides from the same protein are quantified across several samples. The relative ratios of each of these peptides are calculated and statistically evaluated by multiple linear regression analysis. Peptides with highly correlated relative abundance have good quantotypic behaviour. (**b,c**) Box plots representing the proportion of quantotypic peptides determined across a panel of 6 (**b**) and 12 (**c**) cell lines included in the resampling analysis. Correlation analysis with resampling was conducted for differing numbers of total cell lines. The fitted red line is a prediction from a cubic smoothing spline estimating the number of quantotypic peptides identified if additional cell lines were included. Box plots are defined as in Fig. 1. (**d**) Dot plot showing the proportion of quantotypic peptides per protein analysed, across all 12 cell lines. Kinases where no quantotypic peptides were identified are marked in red.
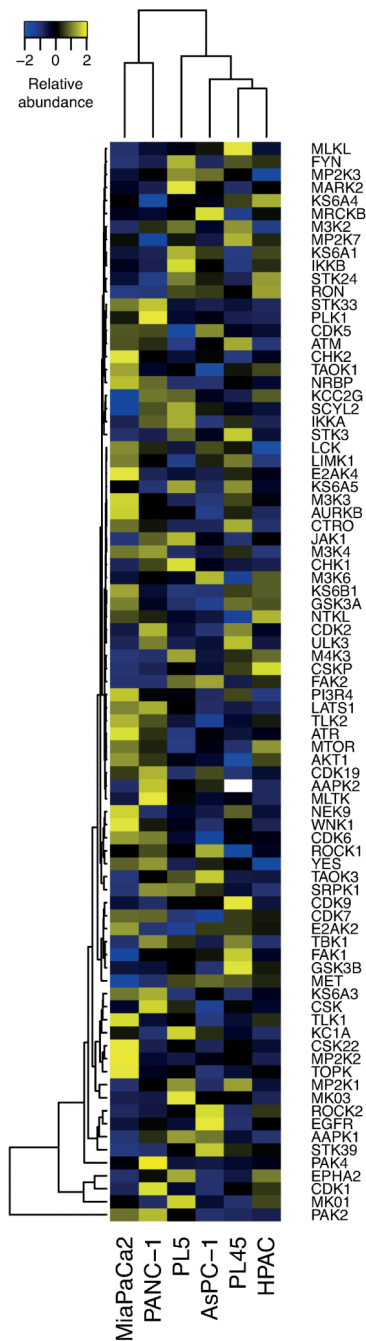
Figure 3. Relative quantification of kinases directly from total cell lysate using quantotypic peptides.
The relative abundance of each kinase is displayed in a heatmap, depicting the 83 protein kinases across six cell lines. The colour key represents the relative abundance (by row z-score). Kinases and cell lines are labelled and dendrograms from hierarchical clustering are plotted.