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

Jonathan D Worboys1, John Sinclair1, Yinyin Yuan2 & Claus Jørgensen1,3

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. We show that measuring ratios of proteotypic peptides across biological samples can be used to empirically identify peptides with good quantotypic properties. We applied this technique to identify quantotypic peptides for 21% of the human kinome.

Selected reaction monitoring (SRM) is an attractive method for accurate quantification of proteins in complex samples by mass spectrometry (MS)1,2. This approach is highly sensitive (low attomole levels), offers a broad dynamic range (five orders of magnitude) and has excellent analytical reproducibility3,4. Individual peptides that are both detectable by MS and unique to the protein of interest (proteotypic peptides5) 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 GPM6–9. When no prior information is available, proteotypic peptides are typically predicted and synthesized10,11. Although these approaches have been widely used, it has been recently shown that optimal proteotypic peptides could be defined only by empirically evaluating all in silico–predicted peptides across the entire protein-coding sequence12.

The underlying assumption for protein quantification in bottom-up proteomics is that the level of the measured peptides is stoichiometric to the level of the protein (quantotypic). Several factors may affect 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, owing to an incomplete understanding of the elements that affect peptide quantotypic behavior, there are currently only limited guidelines for predicting quantotypic peptides4,13. Importantly, because 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 (Fig. 1a). Initially, we sought to identify expressed protein kinases across a panel of six cell lines. Using ActivX nucleotide analogs (desthiobiotin-ATP and desthiobiotin-ADP), we enriched nucleotide-binding proteins and identified isolated proteins by data-dependent analysis on an Orbitrap Velos mass spectrometer14,15. This led to the accumulated identification of 219 protein kinases (Fig. 1b and Supplementary Fig. 1), covering 42% of the human kinome (518 kinases) (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 spectrometer12 (Online Methods). To increase sensitivity during this stage of assay development, we evaluated protein kinases from enriched samples for which the discovery analysis provided high sequence coverage. In total, we evaluated 35,954 transitions (y-series ions) targeting 5,806 peptides across 208 protein kinases (Fig. 1c and Online Methods). Owing to the high number of peptides that were evaluated, we predicted the retention time of each peptide using SSRCalc 3.0 (ref. 16), which facilitated the evaluation of close to 1,000 transitions in a single MS analysis. In total, this led to the identification of 4,375 transitions for 1,820 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 38% of the human kinome (Fig. 1d), of which 132 (25% of the human kinome) were covered by three or more proteotypic peptides. All proteotypic peptides, from the 132 kinases, were subsequently validated with synthetic counterparts: 453 of 466 peptides displayed a Pearson correlation >0.8 and ΔiRT <10 (<1.5 min), 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.

1Division of Cancer Biology, The Institute of Cancer Research, London, UK. 2Division of Molecular Pathology, The Institute of Cancer Research, London, UK. 3Present address: Systems Oncology Group, Cancer Research UK Manchester Institute, The University of Manchester, Manchester, UK. Correspondence should be addressed to C.J. (claus.jorgensen@cruk.manchester.ac.uk).

Received 7 October 2013; accepted 14 July 2014; published Online 24 August 2014; DOI:10.1038/NMETH.3072
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. Because quantotypic peptides are stoichiometric to the level of the protein, we reasoned that the relative level of 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.

In contrast, if a peptide is differentially modified across samples, the ratio to other proteotypic peptides changes. Because 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...
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 six cell lines, which we refer to as group 1 (Fig. 2b). To rigorously test the performance of the identified quantotypic peptides, we further evaluated their behavior across an independent panel of six additional cell lines that we refer to as 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 six cell lines in group 2 (Supplementary Fig. 3a). Notably, 72% of all quantotypic peptides from group 1 were validated in group 2, with an area under the curve from a receiver operating characteristic 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) affected the ability to identify quantotypic peptides. First, we fit a smoothing spline model to the data from each group (Fig. 2b and Supplementary Fig. 3a), which predicted that increasing the number of samples would 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 showed that inclusion of ten 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, or 21% of the human kinome (Fig. 2d). We hypothesize that the high fractions of identified quantotypic peptides are due to the use of endogenously expressed proteins for evaluating proteotypic properties; thus, poorly performing peptides will have been excluded in the initial evaluation.

As protein kinases are regulated by phosphorylation, we subsequently determined whether growth factor stimulation influenced the behavior of assigned quantotypic peptides. To evaluate the effect of early as well as late signaling events, we stimulated cells with insulin-like growth factor II or epidermal growth factor and subsequently determined whether growth factor stimulation influenced the behavior of assigned quantotypic peptides.
factor (IGF-II or EGF) for 5, 10 or 30 min, 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 (Cook’s distance D > 0.37; Online Methods) under these conditions (Supplementary Results). As such, the empirically identified quantotypic peptides were 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 an SRM experiment in biological and technical replicates. In total, we analyzed 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). As the most intense proteotypic peptides from individual proteins can be used as representative of the relative protein level, 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 evaluated proteotypic and quantotypic behavior of peptides for targeted analysis across the human kinome and showed that endogenously expressed proteins can serve as a practical source for SRM assay generation and optimization. This approach can be easily implemented across different classes of proteins and facilitates evaluation of the quantotypic behavior of peptides.

**METHODS**

Methods and any associated references are available in the online version of the paper.

**Accession codes.** ProteomeXchange: MS proteomics data have been deposited with the data set identifier PXD00531. PeptideAtlas SRM Experiment Library (PASSEL): targeted proteomics data have been deposited with the data set identifier PASS00531.

*Note: Any Supplementary Information and Source Data files are available in the online version of the paper.*

**ACKNOWLEDGMENTS**

This work is supported by a Biotechnology and Biological Sciences Research Council (BB/SRC)/Pfizer CASE Studentship (BB/IS32392/1; J.W.) and a Cancer Research UK Career Establishment Award (C37293/A12905; C.J.). We thank colleagues in the Cell Communication Team (The Institute of Cancer Research) for valuable input and useful discussions and the PRIDE team (The European Bioinformatics Institute, UK) for help with data submission.

**AUTHOR CONTRIBUTIONS**

J.D.W. and C.J. conceived the idea and designed experiments. J.D.W. performed all SRM assays. J.D.W. and J.C. analyzed samples for discovery analysis. Y.Y. and J.D.W. developed and applied computational analysis techniques. J.D.W., C.J. and Y.Y. analyzed data. J.D.W. and C.J. wrote the manuscript. C.J. oversaw the project.

**COMPETING FINANCIAL INTERESTS**

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Picotti, P., Bodenmiller, B. & Aebersold, R. Nat. Methods 10, 24–27 (2013).
2. Gillette, M.A. & Carr, S.A. Nat. Methods 10, 28–34 (2013).
3. Domon, B. & Aebersold, R. Nat. Biotechnol. 28, 710–721 (2010).
4. Lange, V., Picotti, P., Domon, B. & Aebersold, R. Mol. Syst. Biol. 4, 222 (2008).
5. Kuster, B., Schirle, M., Mallick, P. & Aebersold, R. Nat. Rev. Mol. Cell Biol. 6, 577–583 (2005).
6. Vizcaíno, J.A. et al. Nucleic Acids Res. 41, D1063–D1069 (2013).
7. Craig, R., Cortens, J.P. & Beavis, R.C. J. Proteome Res. 3, 1234–1242 (2004).
8. Deutsch, E.W., Lam, H. & Aebersold, R. EMBO Rep. 9, 429–434 (2008).
9. Picotti, P. et al. Nat. Methods 5, 913–914 (2008).
10. Mallick, P. et al. Nat. Biotechnol. 25, 125–131 (2007).
11. Picotti, P. et al. Nat. Methods 7, 43–46 (2010).
12. Stengachis, A.B., MacLean, B., Lee, K., Stamatoyanopoulos, J.A. & MacCoss, M.J. Nat. Methods 8, 1041–1043 (2011).
13. Brownridge, P. et al. Proteomics 11, 2957–2970 (2011).
14. Patricelli, M.P. et al. Biochemistry 46, 350–358 (2007).
15. Patricelli, M.P. et al. Chem. Biol. 18, 699–710 (2011).
16. Krokhin, O.V. Mol. Cell. Proteomics 3, 908–919 (2004).
17. Ludwig, C., Claassen, M., Schmidt, A. & Aebersold, R. Mol. Cell. Proteomics 11, M111.013987 (2012).
ONLINE METHODS

Cell culture. All cell lines used were obtained from ATCC and included 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 Technologies), 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 checked every second month.

Cell lysis. Cells were lysed using PLC buffer (50 mM Tris–HCl, pH 7.5, 150 mM NaCl, 1.5 mM MgCl2, 1 mM EDTA, 10 mM NaPPi, 10% glycerol and 1% Triton X-100, 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 were cleared by centrifugation at 4 °C for 15 min at 16,000g. 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 dithiothreitol (DTT)) and separated on 10% polyacrylamide gels. Gels were transferred to a nitrocellulose membrane (LI-COR) by a wet transfer system (Bio-Rad) and were blocked in 1× Roti-Block (Carl Roth). Primary antibodies used were against GAPDH (1:2,000, Santa Cruz, FL-335, sc-25778) and P-Tyr-1000 (1:2,000, Cell Signaling Technology (CST), #8954). Secondary antibodies used were anti-mouse IgG DyLight 680 (1:10,000, CST, #5470) and anti-rabbit IgG DyLight 800 (1:10,000, CST, #5151). Blots were visualized fluorescently using a LI-COR Odyssey imaging system.

Kinase enrichment with ATP/ADP probes. Cell lysates were enriched for kinases using ActivX dethiobiotin-ATP and dethiobiotin-ADP probes (Thermo Scientific), essentially according to the manufacturer’s instructions. Briefly, cell lysates were desalted using Zeba spin desalting columns (7K MWCO, 5 ml, Thermo Scientific) to remove endogenous ATP. Lysates were eluted with reaction buffer (20 mM HEPES, pH 7.4, 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 labeling with the ActivX probes, 1 mg of cell lysate was adjusted to 2 mM MgCl2 and incubated with 20 µM of ActivX probe in a final volume of 500 µl for 30 min at room temperature. Afterward, 500 µl urea lysis buffer (8 M urea, 5 mM Tris–HCl, pH 7.4, 150 mM NaCl, 1 mM EDTA, 1% NP-40, 5% glycerol) were added to the lysate to stop the reaction. Samples were then incubated with 25 µl high-capacity streptavidin agarose resin (Thermo Scientific) for 1 h at room temperature. Beads were collected by centrifugation at 800g for 30 s, 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 precast Any kD mini-Protein gels (Bio-Rad). Proteins were visualized 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 ten 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 min and dried under vacuum in a Savant SC250 express SpeedVac concentrator (Thermo Scientific) for 10 min. The dried gel bands were reduced in 10 mM DTT, 5 mM ammonium bicarbonate (Ambic), pH 8, for 45 min at 50 °C, which was followed by alkylation in 50 mM iodoacetamide (IAA) in 5 mM Ambic for 1 h at room temperature in the dark. Gel pieces were subsequently washed three times with 50% MeCN and dried under vacuum for 10 min. Proteins were digested with 100 ng sequence-grade trypsin (Promega) in 5 mM Ambic for 18 h 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 analyzed by liquid chromatography–mass spectrometry (LC-MS).

Mass spectrometry. Discovery-based analysis was conducted on an LTQ Orbitrap Velos mass spectrometer (Thermo Scientific) coupled to a NanoLC-Ultra 2D with a HPLC-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 (MeCN, 0.1% FA) against solvent A (H2O, 0.1% FA) 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 × 106 ions. The 20 most abundant ions were selected for fragmentation and dynamically excluded for 8 s. 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 (March 2013, 89,601 entries). Searches were performed with a precursor mass tolerance set to 10 p.p.m., fragment mass tolerance set to 0.8 Da and a maximum number of missed cleavages set to 2. Static modifications were limited to carboxymethylation 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 Percolator18). Proteins were assigned by a minimum of one unique peptide. The mass spectrometry proteomics data have been deposited to the ProteomeXchange consortium19 via the PRIDE partner repository with the data set identifier PXD001026.

Targeted analysis was conducted on a TSQ Vantage triple quadrupole mass spectrometer (Thermo Scientific) coupled to a NanoLC-Ultra 1D with a HPLC-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 of 0.7 Th. Q2 was operated at 1.5 mtrr with predicted

doi:10.1038/nmeth.3072
collision energies for each peptide. 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 analyzed using Skyline. We used the extracted ion chromatograms for the two 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) with the data set identifier PASS00531.

**In silico digests 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. 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 an m/z ratio between 300 and 1,500. Retention times for each peptide were predicted using SSRCalc 3.0 (ref. 16). Peptides were considered detected when all fragment ions coeluted, with at least three data points across the peak, and a signal-to-noise ratio of at least 7 (ref. 3).

We were unable to identify unique peptides with sufficiently high number of transitions and a high signal-to-noise ratio for 12 proteins. Therefore, to identify peptides affected by growth factor levels and this analysis was conducted across only a single cell line, peptide quantotypic behavior 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 tenth 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.

Next, a cubic smoothing spline model was fit to the median of data. The degrees of freedom for the smoothing spline was chosen to minimize the penalized cross-validation error of the model and avoid overfitting (see Supplementary Software). Finally, a smoothing spline with the optimal degree of freedom was fit to the data and subsequently predicted values for n = 7–10, in the case of six cell lines, and n = 13–16, in the case of 12 cell lines.

To reflect differences in numbers of peptides analyzed 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 penalized when more peptides do not correlate. Where no peptides correlate, zeros are assigned. peptide ranking was assigned on the basis of the multiple intraprotein 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).

Peptide CV and intensity was compared between quantotypic and nonquantotypic peptides (Supplementary Fig. 9).

**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 cutoff. In addition, relative elution time was assigned for all peptides using a standard peptide mix (Thermo Scientific, #88320, iRT values) for calibration. Overall, the measured iRTs 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 preparation.

Samples were then enriched with ActivX, and all quantotypic peptides were analyzed. As the multiple correlation analysis for identification of quantotypic peptides relies on differences in protein levels and this analysis was conducted across only a single cell line, peptide quantotypic behavior 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 tenth 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.

**Growth factor stimulation.** MiaPaCa2 cells were grown to 50% confluence and either left untreated or stimulated with 100 ng/ml EGF or IGF-II for 5, 10 or 30 min (Supplementary Fig. 5). Samples were then enriched with ActivX, and all quantotypic peptides were analyzed. As the multiple correlation analysis for identification of quantotypic peptides relies on differences in protein levels and this analysis was conducted across only a single cell line, peptide quantotypic behavior 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 tenth 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.

**Statistical analysis of quantotypic peptides.** Correlation analysis was performed with two-sided Pearson’s product moment correlation on total, log-normalized, 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 six cell lines. For the extrapolation analysis, we first calculated the fraction of quantotypic peptides on the basis of the correlation when the number of cell lines n = 3−6 (Fig. 2b) and n = 3−12 (Fig. 2c), each with five resampling without replacement.