Plug-and-play analysis of the human phosphoproteome by targeted high-resolution mass spectrometry

Robert T Lawrence¹, Brian C Searle¹, Ariadna Llovet¹ & Judit Villén¹

Systematic approaches to studying cellular signaling require phosphoproteomic techniques that reproducibly measure the same phosphopeptides across multiple replicates, conditions, and time points. Here we present a method to mine information from large-scale, heterogeneous phosphoproteomics data sets to rapidly generate robust targeted mass spectrometry (MS) assays. We demonstrate the performance of our method by interrogating the IGF-1/AKT signaling pathway, showing that even rarely observed phosphorylation events can be consistently detected and precisely quantified.

Each human cell harbors a signaling landscape that likely spans hundreds of thousands of phosphorylated residues¹. Investigating how these residues are dynamically engaged to control cell behavior in the context of time, environment, cellular identity, and genetic variation requires systematic phosphoproteome analysis using high-throughput assays that are accurate, sensitive and reproducible. Measuring phosphorylation events in a targeted manner presents many hurdles. To date, it has been achievable only through tedious assay optimization and reliance on synthetic peptide standards²–⁴, which impede assay versatility and limit widespread adoption of the technique by researchers outside the proteomics community. Our goal was to develop the capability to easily generate 1-h 'plug-and-play' targeted phosphoprotein assays that are equivalent in sensitivity to prolonged deep fractionation experiments (>12-h analysis time) and have more reproducible sampling and quantification.

Much of the work on cellular signaling using mass spectrometry (MS)-based proteomics has focused on phosphorylation site discovery, generating vast catalogs of novel phosphorylation events and their regulation⁵–⁸. This workflow generally employs a data-dependent acquisition (DDA) strategy, in which the N most abundant features in each full MS scan are selected for MS/MS fragmentation and identification. However, one of the major problems that have plagued quantitative proteomics using DDA is stochastic sampling, which leads to extensive but sparse data sets that have many missing values across different experimental conditions⁹. Analysis of phosphopeptide-enriched samples is further complicated by their high dynamic range, limiting the sensitivity and reproducibility of DDA. Recently, more systematic and sensitive data acquisition strategies have emerged to meet these analytical challenges, including data-independent acquisition (DIA) and parallel reaction monitoring (PRM). In DIA, MS/MS scans are acquired across the full mass range each duty cycle¹⁰,¹¹. In PRM, MS/MS scans are targeted toward narrow, pre-specified mass and time windows corresponding to analytes of interest¹². Compared with selected-reaction monitoring (SRM), which has been the workhorse of targeted proteomics, PRM simplifies the targeted mass spectrometry workflow. To configure an assay, one needs to specify only the precursor mass-to-charge ratio (m/z) and the expected retention time, and no optimization is required a priori. Potential interferences can be identified and fragment ions quantified post hoc.

The promise of targeted quantitative phosphoproteomic analysis has been demonstrated in several recent studies³,⁴,¹³. However, selecting the best peptide sequence and charge state to monitor for phosphorylation sites still represents a major obstacle to targeted analysis. And, because protein phosphorylation is site specific, selection of MS-compatible peptide sequences is limited by the local sequence composition and by the protease enzyme used for digestion. Phosphorylation alters the local charge distribution, which interferes with routinely used enzymes like Lys-C and trypsin, further hampering peptide selection¹⁴,¹⁵. Thus, the preferred peptide cleavage and charge state are difficult to predict a priori.

Here, we instead relied upon a large-scale database of previously observed human phosphopeptide sequences. We assembled this database by searching nearly 1,000 liquid chromatography (LC)–MS/MS runs from label-free, trypsin-digested, phosphopeptide-enriched human samples. These samples were derived from a variety of human cell lines that were exposed to many different stimuli and processed using different phosphopeptide enrichment methods and single-shot as well as deep offline fractionation techniques. More than two-thirds of the data (727 runs) were collected in-house. Additionally, we searched 262 LC–MS/MS runs from three other groups (refs. 8, 16 and S.A. Gerber, unpublished data). Overall, we identified more than 7.5 million phosphopeptide spectral matches (PSM-level FDR < 1%) corresponding to 109,611 phosphorylation sites (90,103 localized P < 0.05) on 11,428 proteins (phosphosite-adjusted FDR < 5%), commensurate with the human phosphoproteome coverage provided in resources such as PhosphoSitePlus¹⁷ (Fig. 1a, Supplementary Fig. 1, and Supplementary Data).

¹Department of Genome Sciences, University of Washington, Seattle, Washington, USA. Correspondence should be addressed to J.V. (jvillen@uw.edu)

RECEIVED 18 SEPTEMBER 2015; ACCEPTED 8 FEBRUARY 2016; PUBLISHED ONLINE 28 MARCH 2016; DOI:10.1038/NMETH.3811
We used the database to quantify several key parameters of data-dependent phosphoproteome analysis that we hypothesize can be addressed with targeted analysis. Without fractionation, DDA is limited in sensitivity. In line with our expectations, 64% of phosphopeptides we identified were observable only in experiments that used extensive fractionation before phosphopeptide enrichment (Fig. 1b). Phosphopeptide sampling stochasticity is detrimental in both single-shot and deep fractionation routines. Sample fractionation increases the depth of phosphoproteome coverage (on average, 5,834 unique phosphopeptides were identified per single-shot experiment versus 32,311 per fractionation experiment), but the overlap between samples was still lower than expected. For single-shot experiments, only 2,544 of 82,944 (3%) phosphopeptides identified were observable in at least 50% of experiments, compared with 16,677 of 210,107 (8%) for fractionation experiments (Fig. 1c).

Next, we examined phosphopeptides sequenced with deep coverage (identified a minimum of 100 times) to ascertain the distribution of charge and cleavage state specificity (preferred state/total observations). We found that most phosphosites were predominantly observed in only one cleavage state (Fig. 1d) and that charge state was moderately specific (Fig. 1e). The preferred phosphopeptide forms were both fully cleaved and of the expected charge state only 35% of the time. As expected, we observed a significant number of missed cleavages (Supplementary Fig. 2a,b), a small fraction (10%) of which were detected after diazotization and/or acetylation. Prediction of the preferred phosphopeptide forms was wrong 50% of the time (Supplementary Fig. 2c), with many of the ions predicted by heuristics falling outside the optimal mass range of the mass spectrometer (Supplementary Fig. 2d).

Lastly, we found that the preferred peptide sequence was the same between at least three of the four groups 87% of the time (Supplementary Fig. 2d), suggesting that the preferred phosphopeptide sequences provided in our database should be compatibile with most laboratory trypsin digestion and phosphopeptide enrichment protocols.

We hypothesized that leveraging a large-scale database of previously observed phosphopeptides would enable rapid assay deployment with higher success rates than traditional approaches. We evaluated the precision of the retention time scheduling by using DDA analysis of bovine serum albumin (BSA) or phosphoenriched tryptic digest to predict the retention times of independent phosphopeptides in a subsequent DDA run, and we established that when using a complex phosphopeptide mixture for assay calibration, 5-min retention time windows are sufficient to capture 95% of the targets (Fig. 2a and Supplementary Table 1). Next, we targeted pairs of phosphopeptides in which the sequence and charge state selected using heuristics differed from the database selection. Our data-driven selection approach outperformed heuristics (Fig. 2b, Supplementary Fig. 3 and Supplementary Table 2) by magnitudes similar to what was predicted by our previous analysis of sequence and charge specificity (Fig. 1d,e).

We further benchmarked our method by selecting 101 phosphopeptides spanning a wide range of detectability and configuring a 1-h parallel reaction monitoring assay to detect those peptides in a phosphoenriched tryptic digest of MCF7 breast cancer cells treated with a cocktail of insulin-like growth factor-1 (IGF-1), epidermal growth factor, and pervanadate. We analyzed the same sample using PRM, DIA, and DDA strategies in technical quadruplicate. The PRM and DIA results were analyzed in a targeted 'peptide-centric' manner querying specifically for the target phosphopeptides as well as in a 'spectrum-centric' manner using a database search pipeline (Supplementary Table 3). Using the PRM method, we readily detected several species that were only sparingly detected in our phosphopeptide database, such as the peptide corresponding to the activation site of the tyrosine protein kinase SYK (Fig. 2c). Measured retention times correlated well with the retention times in the database ($R^2 > 0.99$), enabling efficient assay scheduling and interpretation (Supplementary Fig. 4a). Out of the 101 targets, PRM was superior to DIA and
DDA 1-h assays in terms of the number of peptides detected and sampling reproducibility (Fig. 2c, Supplementary Fig. 4b, and Supplementary Table 3). It has been suggested that peptide-centric targeted analysis might offer advantages over the traditional spectrum-centric approach since it more directly evaluates the evidence for a given peptide. In our analysis, peptide-centric analysis was more sensitive than database searching for both PRM and DIA, and the signal (if measurable) was consistently detectable across all four runs (Fig. 2c and Supplementary Fig. 4b).

Lastly, we designed and implemented a targeted assay to quantify phosphorylation sites on proteins within the IGF-1/AKT signaling pathway in MCF7 cells before and after stimulation with IGF-1 (Supplementary Table 4). Our assay enabled reproducible isoform-specific quantification of protein kinase AKT1/2 activation via phosphorylation at T308/T309 and S473/T474 (Fig. 2d). The specific isoforms of AKT are thought to have distinct roles in cellular signaling, but the respective kinase activation sites T308/T309 are not distinguishable using specific antibodies because of nearly identical local sequence composition. They are sparingly detectable by DDA even after deep fractionation but were reproducibly detected using PRM in the single-stage enrichment protocol used here.

In addition to its advantages of sensitivity and reproducibility, PRM has the capability to monitor isobaric peptide species, which are ubiquitous in phosphorylated proteins (Supplementary Fig. 5a). These positional isomers can often be resolved by retention time, allowing for more accurate quantification (Supplementary Fig. 5b).

Overall, we demonstrate the potential of label-free PRM assays for robust, high-throughput, targeted phosphoproteome analysis. In order to facilitate plug-and-play assay development, we created a web-based application that queries our database for optimal peptide selection and retention time scheduling (https://phosphopedia.gs.washington.edu). This application provides several tools for assay development, including precurated lists of phosphosites, information for sequence and charge state selection, an MS/MS spectra viewer, retention time calibration, automated variable window assignment for positional isomers, and dynamic schedule visualization and optimization. Using this tool, targeted phosphoproteomic assays are convenient to configure, sensitive enough to detect low-abundance analytes without sample fractionation, and more reproducible than DDA. The use of label-free targeted quantification in conjunction with data-driven peptide selection enables rapid deployment of assays to measure virtually any known phosphorylation event in human specimens. These qualities make the method suitable for interrogating the diverse dimensions of the cellular signaling landscape with high throughput and versatility.
METHODS
Methods and any associated references are available in the online version of the paper.

Accession codes. Raw MS data for the experiments performed in this study are available at MassIVE (MSV000079423) and ProteomeXchange (PXD003344).

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

ACKNOWLEDGMENTS
We would like to thank S.A. Gerber for providing data, the MacCoss lab for advice designing and analyzing DIA and PRM assays with Skyline, and the Villén lab for critical discussions. This work was supported by a Samuel and Althea Stroum Endowed Graduate Fellowship to R.T.L., an Interdisciplinary Training in Genome Sciences grant from NIH/NHGRI (T32 HG00035) to B.C.S., a Howard Temin Pathway to Independence Award from NIH/NCI (K99/R00CA140789) to J.V., and an Ellison Medical Foundation New Scholar Award (AG-NS-0953-12) to J.V.

AUTHOR CONTRIBUTIONS
R.T.L. and J.V. conceived the study. R.T.L., B.C.S., and J.V. designed the experiments. R.T.L. and B.C.S. performed the experiments and analyzed data. A.L. created the web resource. J.V. supervised the work. R.T.L., B.C.S., and J.V. wrote the paper.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.
ONLINE METHODS

Cell culture. MCF7 breast cancer cells were obtained from ATCC and tested biannually for the presence of mycoplasma. MCF7 cells were cultured at 37 °C in 5% CO₂ in Dulbecco’s modified Eagle’s medium (DMEM) supplemented with 4.5 g/L glucose, L-glutamine, and 10% FBS. To generate bulk phosphopeptides for method comparisons, cells were incubated in serum-free medium for 4 h before treatment with IGF-1 (100 ng/ml), EGF (100 ng/ml), and peroxanate (1 mM) for 15 min. For IGF-1 experiments, cells were incubated in serum-free medium for 4 h and stimulated with or without IGF-1 (100 ng/ml) for 15 min (n = 6). At the time of harvest, cells were rinsed 3 times quickly with ice-cold, phosphate-buffered saline and flash frozen on liquid nitrogen.

Sample preparation. Cell lysis was performed in 9 M urea, 50 mM Tris pH 8.2, 75 mM NaCl with protease inhibitors (Roche) and phosphatase inhibitors (50 mM beta-glycerophosphate, 50 mM sodium fluoride, 10 mM sodium pyrophosphate, 1 mM sodium orthovanadate). Cells were scraped off of plates directly into ice cold lysis buffer and subjected to 20 s of probe sonication, incubated on ice for 20 min to solubilize proteins and spun at 12,000g for 10 min, and protein content was assayed using the bicinchoninic acid method (Pierce). Proteins were reduced with 5 mM dithiothreitol for 30 min at 55 °C, alkylated with 10 mM iodoacetamide for 15 min at room temperature, and quenched with an additional 10 mM dithiothreitol. Protein extracts were diluted fivefold with 50 mM Tris pH 8.2 and digested overnight at 37 °C with sequencing grade trypsin (Promega) in a 1:200 enzyme/substrate ratio. Following digestion, the reactions were quenched with 10% TFA to pH ~2, desalted on C18 SepPak cartridges (Waters), and dried by vacuum centrifugation. For bulk phosphopeptide preparation, 5 mg of trypptic peptides were resuspended in immobilized metal affinity chromatography (IMAC) loading solution (80% MeCN, 0.1% TFA) and divided into 12 × 150 µl aliquots. To prepare IMAC slurry, Ni-Ni magnetic agarose (Qiagen) was stripped with 40 mM EDTA for 30 min, reloaded with 10 mM FeCl₃ for 30 min, washed 3 times and resuspended in IMAC loading solution. Phosphopeptide enrichment was performed using a KingFisher Flex robot (Thermo Scientific) programmed to incubate peptides with 150 µl 5% bead slurry for 30 min, wash 3 times with 150 µl 30% MeCN, 0.1% TFA, and elute with 60 µl 1:1 MeCN:1% NH₄OH. The eluates were acidified with 10% formic acid, pooled, and dried by vacuum centrifugation. For IGF-1 experiments, 350 µg trypptic peptides were enriched for each sample. To control variability in phosphopeptide enrichment and mass spectrometry, we used a spike-in standard of bovine serum albumin trypptic peptides previously subjected to in vitro phosphorylation by serum-stimulated HeLa cell lysate in the presence of 2.5 mM ATP for 60 min at 30 °C. The resulting peptides were purified by solid phase extraction on a C18 SepPak cartridge and spiked in before IMAC at a mass ratio of 1:50.

Mass spectrometry. Phosphopeptide-enriched samples were resuspended in 4% formic acid, 3% MeCN and subjected to liquid chromatography on an EASY-nLC 1000 system equipped with a 100 µm inner diameter × 25 cm column packed in-house with Reprosil C18 1.9 µm particles (Dr. Maisch GmbH) and column oven set to 50 °C. All separations were performed using a gradient 9% to 32% MeCN in 0.15% formic acid over 44 min (60 min total method length) at a flow rate of 500 nl/min. The HPLC was coupled directly with a Q-Exactive mass spectrometer. The DDA method consisted of a full MS scan (70k resolution, 3e6 automatic gain control (AGC) target, 240 ms maximum injection time, 400 to 1,200 m/z, centroid mode) followed by up to 20 data-dependent MS/MS acquisitions on the top 20 most intense precursor ions (35k resolution, 5e5 AGC target, 120 ms maximum injection time, 2 m/z isolation window, 27% normalized collision energy, centroid mode). The DIA method consisted of a full MS scan configured as above followed by 33 data-independent MS/MS acquisitions configured using an inclusion list with 25 m/z overlapping windows (12.5 m/z with deconvolution) covering the 400 to 1,200 m/z mass range (35k resolution, 5e5 AGC target, 120 ms maximum injection time, 25 m/z isolation window, 27% normalized collision energy, centroid mode). The PRM method consisted of a full MS scan configured as above followed by up to 20 targeted MS/MS scans as defined by a time-scheduled inclusion list (35k resolution, 5e5 AGC target, 120 ms maximum injection time, 2 m/z isolation window, 27% normalized collision energy, centroid mode). To prevent systematic bias, the order of acquisition for ‘control’ and “IGF-1” samples was randomized. Benchmarking experiments for retention time, sequence and charge selection were performed on a nanoACQUITY liquid chromatography system coupled to a Q-Exactive Plus mass spectrometer with the following modifications to the above parameters: flow rate was set to 400 nl/min and for the PRM assays 25 unscheduled targeted MS/MS scans using 50 ms maximum injection time and 17.5k resolution were collected after each full MS scan. For DDA and DIA, AGC targets were optimized for speed with a goal of reaching the target before reaching maximum injection time. For PRM, the AGC target was selected for enhanced sensitivity and dynamic range with a goal of reaching the maximum injection time before reaching the target. PRM assay scheduling was performed within Skyline (version 3.1.0.7382). To calibrate the schedule, an initial pilot run was conducted with 10-min-wide acquisition windows, and aligned to the normalized retention database provided with this manuscript (HumanPhosphoproteomeRT,.irtdb) to build a retention time predictor. Subsequently, a scheduled isolation list with refined 6 min windows was exported from Skyline as a .csv file and imported directly into the instrument PRM method as an inclusion list. Any group of peptides from the database may be used as retention time calibrators, including any of the phosphopeptides, the Peptide Retention Time Calibration (PRTC) mixture (Pierce), or tryptic peptides from BSA. An equivalent retention time scheduling tool with other capabilities such as variable windows for positional isomeric phosphopeptides is available with the web portal that accompanies this manuscript.

DDA data processing and analysis. Raw DDA data files were converted to mzXML and searched using Comet (version 2015.01) against the human SwissProt database including reviewed isoforms (April 2015; 42,121 entries) allowing for variable oxidation of methionine, protein N-terminal acetylation, and phosphorylation of serine, threonine, and tyrosine residues. Carbamidomethylation of cysteines was set as a fixed modification. Tryptsin (KR|P) digestion allowing for up to two missed cleavages was selected. Precursor mass tolerance was set to 50 ppm, and fragment ion tolerance to 0.02 Daltons. Search results were filtered using Percolator to reach a 1% false discovery rate at
the PSM level. Phosphosite assignment was performed using an in-house implementation of Ascore\textsuperscript{23}, and sites with Ascore \(\geq 13\) were considered localized (\(P = 0.05\)). To construct the large-scale phosphopeptide database we imposed additional filters to prevent accumulation of false hits associated with data aggregation. First, phosphopeptides in the database with multiple non-localized instances spanning the same sequence were only considered to correspond to the minimum number of phosphosites that explain the data. Second, we carried forward the best posterior error probability for each phosphopeptide spectral match, phosphoisoform, and phosphosite in order to compute an adjusted FDR at each level. A phosphoisoform represents multiple peptide sequences containing the same combination of phosphorylation sites. Multiple peptides with different degrees of phosphorylation or cleavage may represent the same phosphosite. Without imposing additional filtering beyond peptide spectral matches, 196,744 phosphosites were identified, but the phosphosite-level false discovery rate after data aggregation was 29.2\% and the adjusted posterior error probability for phosphosites identified by only a single MS/MS scan was 44\% (Supplementary Fig. 1b). Accordingly, we suspect that phosphorylation site databases that aggregate large volumes of spectral data without imposing additional filters are also likely to aggregate false discoveries. Lastly, spectral libraries were constructed from aggregate phosphopeptide search results and assembled into a normalized retention time database (HumanPhosphoproteomeRT\_irtdb) using Skyline.

**DIA and PRM data processing and analysis.** For spectrum-centric analysis of DIA and PRM mass spectrometry results, we used DIA-Umpire\textsuperscript{24} version 1.4 with default parameters to assemble pseudo-MS/MS spectra for the database search pipeline described above. Peptide-centric analysis was performed using Skyline. Signal extraction was performed on +2, +3, +4 precursors and +1, +2 b and y fragment ions. Full MS resolving power was set to 70,000 and MS/MS resolving power set to 17,500. After importing an initial run, extracted ion chromatograms were aligned to the retention time library to generate a predictor and all results were reimported using retention time filtering to within 5 min of predicted RT. Peptide identifications were further refined by manual interpretation using several criteria including product ion mass accuracy, correlation of precursor and fragment ion peak shapes, and signal-to-noise ratios. Specifically, we required at least three highly-resolved fragment ions without interference to consider a peptide identified. To consider a peptide localized, we required at least 1 site-diagnostic ion. For IGF-1 experiments, integrated peak areas were measured for each peptide in Skyline and exported for analysis. Values were normalized to the average peak areas of 3 spiked-in phosphorylated bovine serum albumin peptides and log\textsubscript{2} transformed. Statistical significance was assessed using a two-sample unpaired \(t\)-test.

**Benchmarking experiments.** Retention times from bovine serum albumin digest (110 peptides) or phosphopeptide-enriched MCF7 digest (phospho-mixture) were used to predict the retention times of a subsequent analysis of the same MCF7 phospho-mixture. For the phospho-mixture prediction, half of the identifications (2,357 phosphopeptides) were randomly selected as ‘training’ data and used to predict the retention times of an independent set of phosphopeptides in the subsequent run. Unscheduled PRM experiments were used to evaluate disagreements in data-driven versus heuristic peptide sequence and charge state selection. For sequence selection, the heuristic was the fully cleaved phosphopeptide (i.e., cleavage after all lysines and arginines except when residue at +1 is proline), and 100 phosphosites were analyzed (200 targets). For charge selection, the heuristic was positively charged amino acids +1, and 50 phosphopeptides were analyzed (100 targets).

**Statistical analysis.** Statistics regarding peptide and phosphorylation site identification are discussed in the above section entitled “Data processing and analysis.” Sample sizes necessary for non-parametric comparisons are not easily predictable since the expected shape of the distribution is unknown. For the comparison of peptide sequence and charge state selection we used \(n = 100\) and \(n = 50\), respectively, which we predicted would be sufficient to detect a roughly tenfold difference in the median peak area intensity. Peptide sequence selection implicitly also requires charge state selection for each sequence, hence the sample size was increased for that experiment to account for additional variability. A Wilcoxon signed rank test (paired nonparametric) was used because the intensity of different peptides representing the same phosphorylation site in the same sample are related but the intensity of peptides arising from different phosphorylation sites in the sample are unrelated and not normally distributed. Similarly, sample size necessary for high-throughput measurements is also difficult to predict, since different analytes in the assay each have different expected effect sizes and precision. For the quantitative analysis of IGF-1 stimulation versus control, we assumed that replicate measures of a typical phosphopeptide would follow a normal distribution after log transformation with coefficient of variation of approximately 20\%. Under these assumptions we used a sample size of \(n = 6\), which we predicted would be sufficient to detect a twofold change in most targets. Unpaired \(t\)-tests were used to assess significant differences between control and IGF-1 treated samples. Replicates were from independent treatments of the same source of MCF7 cells.

10. MacLean, B. et al. *Bioinformatics* **26**, 966–968 (2010).
21. Eng, J.K., Jahan, T.A. & Hoopmann, M.R. *Proteomics* **13**, 22–24 (2013).
22. Käll, L., Canterbury, J.D., Weston, J., Noble, W.S. & MacCoss, M.J. *Nat. Methods* **4**, 923–925 (2007).
23. Beausoleil, S.A., Villén, J., Gerber, S.A., Rush, J. & Gygi, S.P. *Nat. Biotechnol.* **24**, 1285–1292 (2006).
24. Tsou, C.C. et al. *Nat. Methods* **12**, 258–264 (2015).