Title

Fast and deep phosphoproteome analysis with the Orbitrap Astral mass spectrometer.

Permalink

Journal

Authors

Lancaster, Noah

Sinitcyn, Pavel

Forny, Patrick

et al.

Publication Date

DOI

Peer reviewed

Article

Fast and deep phosphoproteome analysis with the Orbitrap Astral mass spectrometer

Noah M. Lancaster [1,2,17], Pavel Sinitcyn [3,17], Patrick Forny [4,17], Trenton M. Peters-Clarke [1,2], Caroline Fecher [4], Andrew J. Smith [4], Evgenia Shishkova [2,5], Tabiwang N. Arrey[6], Anna Pashkova[6], Margaret Lea Robinson[2], Nicholas Arp [3,7], Jing Fan [3,7,8], Juli Hansen[9], Andrea Galmozzi [2,9,10], Lia R. Serrano[1,2], Julie Rojas[11], Audrey P. Gasch [11,12,13], Michael S. Westphall[2,5], Hamish Stewart[6], Christian Hock[6], Eugen Damoc[6], David J. Pagliarini [4,14,15], Vlad Zabrouskov[16] & Joshua J. Coon [1,2,3,5,11] ✉

Owing to its roles in cellular signal transduction, protein phosphorylation plays critical roles in myriad cell processes. That said, detecting and quantifying protein phosphorylation has remained a challenge. We describe the use of a novel mass spectrometer (Orbitrap Astral) coupled with data-independent acquisition (DIA) to achieve rapid and deep analysis of human and mouse phosphoproteomes. With this method, we map approximately 30,000 unique human phosphorylation sites within a half-hour of data collection. The technology is benchmarked to other state-of-the-art MS platforms using both synthetic peptide standards and with EGF-stimulated HeLa cells. We apply this approach to generate a phosphoproteome multi-tissue atlas of the mouse. Altogether, we detect 81,120 unique phosphorylation sites within 12 hours of measurement. With this unique dataset, we examine the sequence, structural, and kinase specificity context of protein phosphorylation. Finally, we highlight the discovery potential of this resource with multiple examples of phosphorylation events relevant to mitochondrial and brain biology.

Protein phosphorylation is an essential post-translational regulatory mechanism for myriad cellular functions including apoptosis, inflammation, metabolism, proliferation, protein trafficking, and many others[1]. Global detection of which proteins, and perhaps most importantly which residues, are subject to this dynamic modification has been a key technological gap for decades[2]. In the early 2000s, key advancements in technologies to enrich phosphorylated peptides—prior to mass spectrometry (MS) analysis—enabled a new era of large-

scale phosphorylation discovery experiments[3–13]. The evolution of that work, combined with gradual, but steady, improvements in both MS hardware and data analysis tools make it possible to map and quantify thousands of phosphorylation sites in a single experiment[14–27]. Still, current phosphoproteomic analyses are neither routine nor straightforward when compared to protein detection and quantification. Phosphoproteome analysis remains challenging due to four main requirements: need for site localization, dynamic range, reproducibility, and throughput. The quality of the tandem mass spectra required to determine precisely upon which residue the phosphoryl group resides is higher than that required to identify an unmodified peptide. Alternative dissociation methods, such as electron transfer dissociation (ETD), and improved mass resolving power and accuracy are both strategies that can aid in improving spectral quality and increasing the likelihood that the detected site can be localized with high confidence[15,21,28,29]. Additionally, sensitivity and dynamic range of the mass analyzer can also be important for detection of low level, but critical, product ions[11,23,27]. Next, the often sub-stoichiometric amounts of phosphorylated protein itself elevate the difficultly[2,13,17–19]. Beyond that, previous studies demonstrate that this dynamic range problem is further exacerbated in that ~10% of detected phosphopeptides account for ~80% of the observed signal[30]. This tremendous dynamic range necessitates both enrichment of phosphorylated peptides and chromatographic fractionation ahead of conventional capillary liquid chromatography tandem MS (nLC-MS/MS)[2,17,18].

Phosphorylation site quantification is often essential to elucidating biological insight[17]. The requirements outlined above also add challenges to achieving this goal. Quantification of a protein, for example, is done by summing the signals of multiple unique peptides all stemming from that single protein. Thus, if a single peptide is not reproducibly detected, the overall measurement can still be made. But unique phosphopeptides cannot be summed with other signals; they must be reproducibly and reliably detected from one sample to the next. Such demands make performing truly large-scale (i.e., >100 samples) comparisons of phosphoproteomes very difficult[18]. Finally, getting sufficient depth to detect targets of a particular kinase, for example, might require extensive fractionation, as discussed. However, scaling that experiment to multiple conditions or samples is often not possible from a throughput perspective, despite all the caveats noted. We conclude that global and quantitative phosphoproteomics technologies require improvement to enable routine and truly large-scale phosphoproteome measurements.

Recently, a new type of mass analyzer has been described–the Asymmetric Track Lossless analyzer (Astral™). The Astral analyzer can achieve high resolving powers (~80,000) and mass accuracy (5 ppm), single ion detection limit, and MS/MS scan speeds up to 200 Hz[31,32]. Here we describe the use of a quadrupole-Orbitrap™-Astral hybrid MS instrument for the analysis of phosphopeptides. Specifically, we examine the ability of this system to perform MS/MS scans of complex mixtures of phosphopeptides separated over durations ranging from 7 to 60 min. We examine the performance of these methods for various peptide mass loads and data acquisition settings as a function of detected and localized phosphorylation sites and overall reproducibility. Validation of site localization and quantification was accomplished using synthetic phosphopeptide standards spiked into a yeast phosphopeptide background. Next, we benchmarked the performance of our Orbitrap Astral method against an Orbitrap Ascend and a previously described timsTOF Pro method by replicating an EGF stimulation of HeLa cells[33]. Finally, we leverage this technology to collect an atlas of phosphorylation in the mouse–generating tens of thousands of localized phosphorylation sites from each of twelve unique tissues. With these data we present, to our knowledge, the deepest mouse phosphoproteome collected in a single study. Using this atlas combined with AlphaFold predicted protein structure[34], we confirm existing hypotheses that that most phosphorylation events are directed towards unstructured regions of proteins[35–37]. The incorporation of a previous kinome atlas allows investigation of tissue-specific kinase activity[38]. We additionally provide examples of how our resource can be mined for key phosphorylation events on biologically relevant proteins.

Results

Rapid phosphopeptide analysis with the Orbitrap Astral mass spectrometer

Owing to its fast MS/MS scan rate and high sensitivity, we hypothesized that the Orbitrap Astral™ mass spectrometer could resolve many of the aforementioned challenges in analyzing phosphoproteomes. Specifically, the Orbitrap Astral MS comprises a conventional quadrupole-Orbitrap coupled with a new mass analyzer (Astral, Fig. 1A). The combination of low ion losses and single ion detection drives the high sensitivity of the Astral analyzer, while very fast MS/MS scan rates allow it to cycle through large numbers of targets. In a typical Orbitrap Astral method, the Astral analyzer is set to generate 200 MS/MS spectra per second while the Orbitrap analyzer in parallel generates slower high resolution and high dynamic range MS data[31].

To test the utility of the Orbitrap Astral MS for the analysis of phosphopeptides, we purified tryptic phosphopeptides from HEK239T cells, loaded them onto a nanoflow capillary column with a pulled electrospray emitter[39], and eluted them into the Orbitrap Astral mass spectrometer. For this initial experiment, we utilized a data-dependent acquisition (DDA) method wherein MS scans were acquired in the Orbitrap analyzer while MS/MS scans were collected using the Astral analyzer. Figure 1B represents an example single-scan tandem mass spectrum that was identified following a traditional MaxQuant database search[25,40]. Here, a triply-protonated precursor having m/z value of 623.6314 and a sequence of RPsQNAISFFNVGHSK was selected, dissociated with beam-type collisional activation (HCD), and analyzed in the Astral analyzer– all within 5 ms. In total, from this single 30-min nLC-MS/MS DDA experiment, we collected 3201 MS and 174,944 MS/MS scans, from which we identified 12,327 phosphopeptides corresponding to 9,537 unique sites of phosphorylation. Note a previous study using the Orbitrap Tribrid platform with a DDA method reported a depth of ~9500 phosphosites following a 120-min method[20]. These MS/MS scans were collected within ~6.3 ms on average, enabling fast sampling of selected precursors, and sometimes exhausting the precursors available to sequence. Next, we investigated the Astral analyzer's capabilities by plotting the measured resolving power (Fig. 1C) and mass accuracy (Fig. 1D) for these product ions. The upper bound of observed resolving powers likely correspond to peaks with low numbers of ions, whereas the lower bound presumably arises from high intensity peaks exhibiting space charging[31]. Note the Astral resolving power persists with increasing m/z value. Finally, mass accuracy, as measured for these product ions, was within 5 ppm for 95% of the measurements.

With the exceptional speed of the Astral mass analyzer, combined with its ability to deliver high mass resolution and accuracy, we next wondered how this instrument would perform for DIA analysis of phosphopeptides[33,41–45]. Accordingly, we analyzed the same tryptic phosphopeptide mixture using a DIA method. The speed of the Astral analyzer allowed for use of DIA windows as narrow as 2 m/z across the range of 380–980 m/z while still maintaining a cycle time of 1.5 s. This scan method is visualized in Fig. 2A with the high-resolution Orbitrap MS scans (240 K resolving power) denoted in blue (~0.6 s cycle time) and the Astral MS/MS scans depicted in red.

A 250-ng injection of tryptic phosphopeptides was analyzed with a 30-min nLC separation and the DIA method described above (see Supplementary Fig. 1A for the base peak chromatogram). The resultant data was searched both with Spectronaut and a developmental build of CHIMERYS (see Methods). From the Spectronaut search, we localized 29,190 phosphorylation sites (Fig. 2B, see Supplementary Fig. 2A for
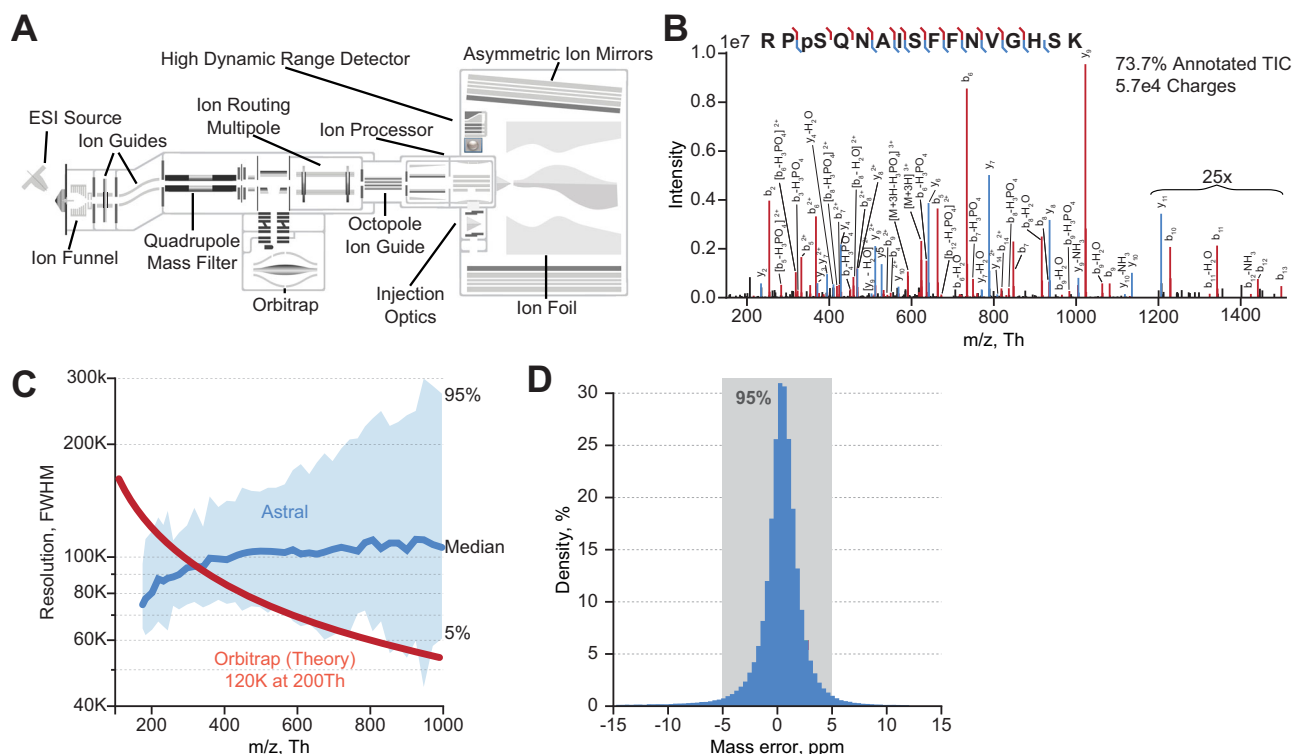
Fig. 1 | Overview of Orbitrap Astral MS and its key figures of merit. A Orbitrap Astral instrument schematic, highlighting the quadrupole, Orbitrap, and Astral analyzers. **B** Tandem mass spectrum of representative phosphopeptide collected using the Astral analyzer. **C** Distribution of Astral analyzer resolving power as a function of mass from phosphopeptide product ion spectra collected here. **D** Phosphopeptide product ion mass measurement error from Astral analyzer.

comparison of Spectronaut and CHIMERYS, and Supplementary Figs. 2B, C phosphopeptide multiplicity and localization probability distributions, respectively). We note that the use of a wider DIA window (4 *m/z*) produced 14% fewer localized phosphorylation sites. Next, we examined the impact of loading amount on performance using a dilution series ranging from 1 μg to 50 ng loaded on column (Fig. 2C). We note that similar results are obtained down to 250 ng loads. The sensitivity of the system is evident in that a 50-ng load produced ~60% of the detected sites observed with a 1 μg load. To determine whether the Astral scan speed would allow for increased throughput, we next evaluated a range of active gradient lengths (Fig. 2D, Supplementary Fig. 1A). A four-fold reduction in gradient length (30–7 min) results in just a ~20% reduction in localized phosphorylation sites. Additional method parameters are justified in the Supplementary Notes.

To determine the impact of gradient length on reproducibility, we performed two additional analyses. First, we examined identification reproducibility across triplicate injections (Fig. 2E). The 7-,15-, and 30-min active gradient methods resulted in 56.7%, 61.8%, and 62.7% of phosphosites detected across all three injection replicates, respectively. Second, we assessed quantitative precision by determining percent relative standard deviation (RSD) of phosphorylation sites detected across all three replicates (Fig. 2F). The 15- and 30-min active gradients achieved median RSDs of 10.9 and 12.6%. Not unexpectedly, the 7-min active gradient method has considerably reduced precision (21.4% median RSD), most likely due to insufficient sampling across the narrower elution peaks (Supplementary Figs. 1B–E).

Since these data were collected with a novel mass analyzer and new instrument, we next sought to ensure that the phosphosites were identified and localized accurately. To test this, we first performed an entrapment search with a combined human and maize protein database[46]. In this experiment, which is commonly used to evaluate internal false discovery rate (FDR), maize false identifications allow for the calculation of an external FDR. The results of this analysis are

shown in Fig. 2G, where internal and external FDR is plotted for all precursors and phosphopeptide precursors. For lower q-values the internal and external FDRs disagree; however, these curves converge at approximately the 1% FDR threshold, suggesting that the identifications reported here are reliable. We conclude that, as the DIA software tools continue to evolve and are adopted for the new data type, this discordance will be diminished.

In the case of phosphoproteomics, peptide identification is not the final result—the site of phosphorylation is ideally localized to a specific residue with confidence. To evaluate the quality of phosphosite localization, we searched these data with phosphorylated proline as a variable modification where any detected proline phosphorylation is false[47]. S, T, and Y all have a significant amount of highly confident localized phosphosites whereas the P does not (Fig. 2H, Supplementary Fig. 2H). To further explore localization confidence, we analyzed a set of synthetic phosphopeptide standards that had been spiked into a complex mixture yeast tryptic phosphopeptides at various concentrations. Note this same 225 synthetic phosphopeptide mix has been used by others and thereby provides a means of both assessment and comparison[41,48]; specifically, the experiment allows for calculation of localization error rate based on ground truth knowledge. From these data we conclude that, depending on localization probability cutoff, the error rate ranges between 1 and 5 percent, a similar trend as the previous study (Fig. 2I, Supplementary Fig. 3A)[41]. Supplementary Fig. 3B demonstrates that we retain the majority of correct precursors even at more stringent localization probability cutoffs. Taken together, these results confirm that a library-free DIA search, using MS/MS spectra from an Astral analyzer, can reliably detect and localize phosphorylation sites.

The aforementioned phosphopeptide standards also provided an opportunity to evaluate quantitative linearity[42,48]. To accomplish this, phosphopeptide standards were diluted from 10,000 to 39 attomole (per standard, on-column) into a constant complex mixture of yeast
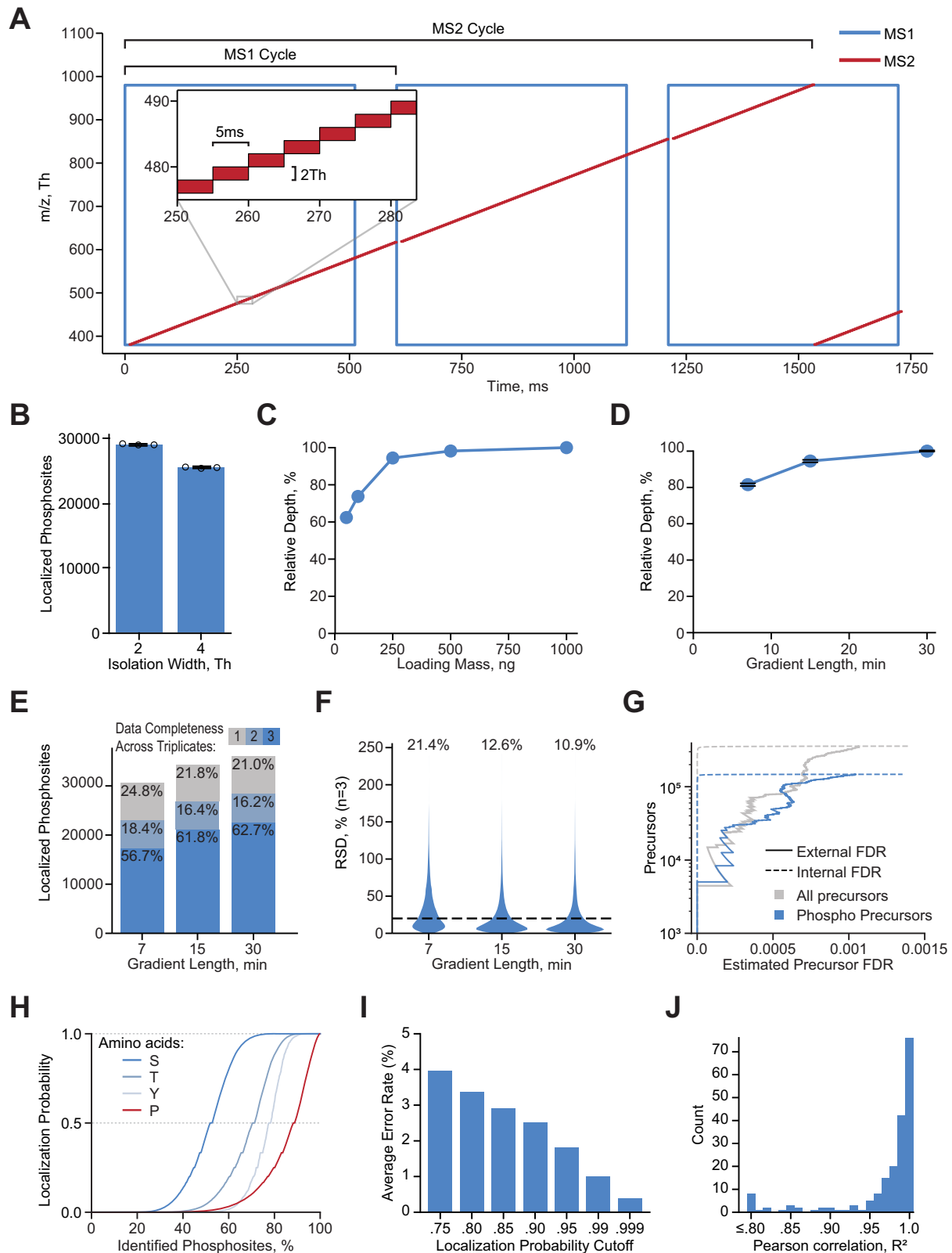
tryptic phosphopeptides. Using the 15-min DIA Orbitrap Astral method, we analyzed the various spike-in samples and then calculated the linear fit between the observed MS intensities and the load amount. Supplementary Fig. 3D displays three randomly chosen curves while Fig. 2J confirms that the majority of the phosphosite calibration curves have R^2 values greater than 0.95, indicating good quantitative linearity. Supplementary Fig. 3C confirms that even at loads as low as 39

attomole, half of the phosphorylation sites were detected across all three injections. Finally, we note that the phosphopeptide standard intensity distributions observed across the dilution series overlap with the yeast phosphopeptide intensity distribution (Supplementary Fig. 3E). Together, these results demonstrate that our method exhibits reliable detection and quantification over the concentration range typically observed in phosphoproteomic analysis.

Fig. 2 | Overview of DIA phosphoproteomics on the Orbitrap Astral MS.
A Illustration of DIA acquisition scheme. **B** Average number of localized phos-
phorylation sites identified using DIA method with various isolation widths for $n = 3$
injection replicates. Error bars indicate maximum and minimum observed values.
C Evaluation of phosphopeptide loading mass on performance. Data was collected
with a 30-min active gradient and 2 m/z DIA isolation width. **D** Effect of gradient
length on performance. Points represent the average across $n = 3$ injection repli-
cates. Error bars indicate maximum and minimum observed values. **E** Evaluation of
phosphosite identification reproducibility (for results from (**D**)). **F** Evaluation of
phosphosite quantitative precision. The relative standard deviation of phosphosite
quantities is shown for phosphosites detected across triplicate injections with the

median value displayed for each active gradient length. **G** Comparison of External
(maize/human entrapment experiment) and Internal (Spectronaut) FDR on pre-
cursor level. **H** Phosphoproline Decoy Search to test reliability of localization
algorithm. The cumulative distribution of localization probabilities is shown for
phosphorylation events at different amino acids. **I** Average site localization error
rate as a function of localization probability cutoff. Synthetic phosphopeptide
standards spiked into a yeast phosphopeptide background were used as a ground
truth for error rate determination. **J** The distribution of R^2 values for linear cali-
brations curves is shown for phosphopeptides standards detected in at least three
concentrations across a five-point dilution series into a constant yeast phospho-
peptide background. Source Data are provided as a Source Data file.

Comparison of Orbitrap Astral to other phosphoproteomics platforms

To compare the performance of the Orbitrap Astral to other phos-
phoproteomics platforms we replicated the HeLa EGF stimulation
experiment as previously reported by Skowronek et al. using the tim-
sTOF Pro hybrid mass spectrometer with dia-PASEF acquisition[33].
Briefly, we cultured HeLa cells in triplicate followed by 15 min of EGF
stimulation, cell lysis, protein extraction, trypsin digestion, and phos-
phopeptide enrichment. The resulting phosphopeptide samples were
then analyzed with a 15-min active gradient, closely replicating the
total acquisition time for the Skowronek et al. publication, with
detection using either the Orbitrap Astral or Orbitrap Ascend™ Tribrid
mass spectrometers. To enable a direct comparison, we searched our
raw files and those from the Skowronek et al. study with the same
protein database and Spectronaut library-free method. Figure 3A
shows the phosphoproteomic depth and completeness across tripli-
cates for all three datasets. The Orbitrap Astral method provided the
deepest phosphoproteomic analysis and yielded near 60% phospho-
site completeness across replicates from approximately three-fold
more localized sites as compared to either the timsTOF Pro or Orbitrap
Ascend instruments. Note the Orbitrap Ascend is not optimally oper-
ated in DIA mode given its slower scan speed; however, we chose to
use a DIA method for direct comparison. Interestingly, the Orbitrap
Ascend generated the highest level of phosphosite overlap of all three
instruments at 64%.

To examine the dynamic range of detected phosphorylation
sites across platforms we plotted the intensity distributions of
phosphorylation sites detected in all three replicates using only the
EGF-stimulated data. Figure 3B, C present the Orbitrap Astral inten-
sity distributions with overlapping sites indicated for either the
timsTOF Pro or Orbitrap Ascend datasets, respectively. Both plots
reveal that the unique phosphosites in the Orbitrap Astral dataset are
biased toward lower intensities. In contrast, unique phosphosites for
the timsTOF Pro dataset are distributed more evenly across the
intensity range (Fig. 3D). These observations suggest that our Orbi-
trap Astral exhibits improved sensitivity, resulting in a larger
dynamic range. Note the Orbitrap Astral and Ascend shows higher
overlap as compared to the timsTOF Pro (Fig. 3E); we suppose this
difference is likely due to variation in EGF treatment and sample
preparation.

To see how this performance variation translates to biological
discovery we performed a differential expression analysis between
control and EGF-stimulated samples for all three datasets (Fig. 3F).
Each datapoint in these plots presents a phosphorylation site with the
differentially expressed ones highlighted in blue. Phosphorylation
sites occurring directly on the EGFR are labeled, many of which are
detected across all platforms. To further compare we performed a
pathway enrichment analysis from each dataset (Fig. 3G)—all three
platforms had the EGF/EGFR signaling pathway as the top enriched
term. We conclude that despite differences in treatment, sample pre-
paration, and instrumentation, all technologies could arrive at the
same biological conclusions.

Deep phosphoproteome analysis of the mouse

Having demonstrated the speed and sensitivity of the Orbitrap Astral
MS for reliable identification of phosphopeptides, we next sought to
leverage this methodology to generate a comprehensive phospho-
protein atlas of the mouse. In 2010, Huttlin and co-workers described
the first mouse phosphoproteome atlas, reporting 35,965 phosphor-
ylation sites detected following 10 days of MS analysis and serving as a
valuable reference point for this study[49]. Fig. 4A presents our overall
experimental design wherein proteins were isolated, digested, and
enriched for phosphorylation from 12 distinct tissues. For each tissue,
the purified phosphopeptides were separated offline and con-
catenated into four fractions, each of which were analyzed using a 15-
min DIA nLC-MS/MS method on the Orbitrap Astral MS. In total, this
experiment required 12 h of active data acquisition time and resulted
in the detection of 81,120 unique phosphorylation sites (Supplemen-
tary Data 2). Note to provide complementary protein abundance we
separately analyzed non-enriched peptides from each tissue (Supple-
mentary Fig. 4, Supplementary Data 3).

Depending upon the source tissue, the number of identified
phosphorylation sites ranged from ~20,000 to 40,000 sites (Fig. 4B).
Of the 81,120 unique sites identified, 61,088 were localized to S; 16,276
localized to T; and 3756 to Y. It is noteworthy that ~4.6% of the total
sites identified stem from pY. Previous studies have placed this num-
ber closer to ~2% and often rely on pY-specific antibodies for enrich-
ment of this perhaps most functionally important phosphorylation
site[30,50]. We suppose that the increased sensitivity and depth afforded
by this new analyzer permits the detection of these low-expression
phosphorylation events (Supplementary Fig. 5C). Also shown in Fig. 4B
are the results, by tissue, from the previous mouse atlas. Note these
raw data were reprocessed using MaxQuant with the matching protein
database and filtered using similar quality metrics. On a per tissue
comparison, the method leveraged here provides a 5-fold boost in the
number of localized phosphorylation sites in ~1/24th the time.

Next, we plotted the distribution, by tissue, of phosphorylation
sites and whether they were detected in both studies (Supplementary
Fig. 5A). Despite the age and strain differences between mice, and the
dynamic nature of phosphorylation, we see good overlap in detection
of sites (median 55.3%). Tissue-specificity was also plotted (Fig. 4C and
Supplementary Fig. 5B) and compared to Huttlin et al. In both studies,
a large portion of phosphorylation sites appear to have single tissue
specificity; however, our results indicate that a sizeable number of sites
are detected across all twelve tissues. This result contrasts with the
Huttlin work and likely stems from the increased sensitivity and
reproducibility afforded by the new analyzer and DIA method. With
this result, we expect that many sites may indeed be present across all
tissues and that further improvements in sensitivity and reproduci-
bility are likely needed to detect them. This hypothesis is further
evinced by the results shown in Fig. 4D, E, where we examined the
detected peak intensities for each site as a function of how many tis-
sues in which it was observed. From these data, both studies show
concordance between number of tissues detected and overall signal
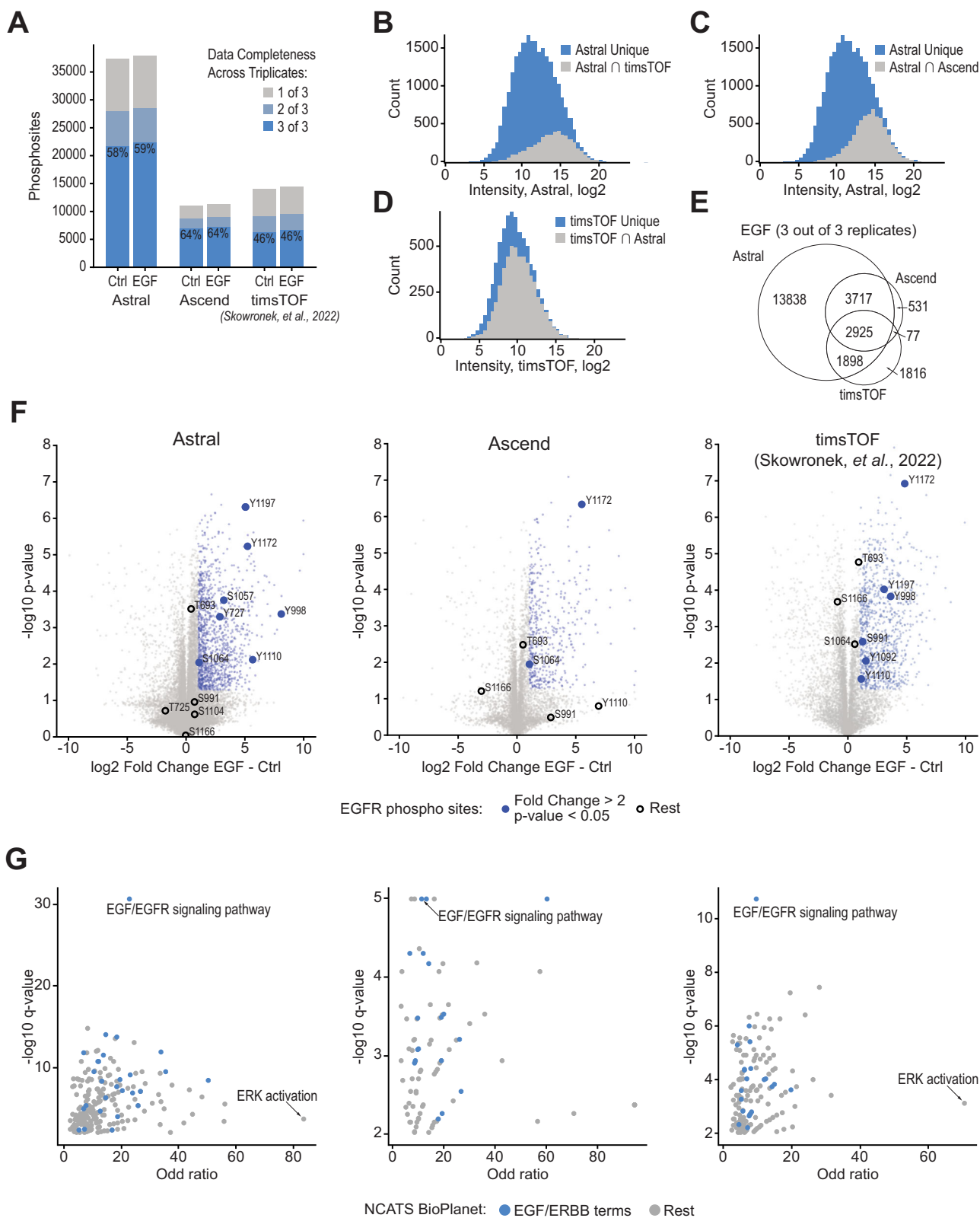strength (i.e., abundance).

Fig. 3 | Biological validation of phosphoproteomics platforms using EGF stimulation. A Reproducibility of detected phosphosites across biological triplicates for three phosphoproteomics platforms—Orbitrap Astral, Orbitrap Ascend, and timsTOF Pro[33]. Intensity distribution of phosphosites consistently detected (3 out of 3 EGF stimulated replicates) by Orbitrap Astral with overlap indicated for (**B**) timsTOF Pro and (**C**) Orbitrap Ascend. **D** Intensity distribution of phosphosites consistently detected (3 out of 3 EGF stimulated replicates) by timsTOF Pro with overlap indicated for Orbitrap Astral. **E** Venn diagram of phosphosites detected across all three mass spectrometry platforms. **F** Volcano plots of phosphosites between EGF stimulation and control across platforms with phosphosites meeting the differential expression criteria (fold-change > 2, $p < 0.05$ via two-sided t-test without multiple testing correction) indicated in blue and EGFR phosphosites labeled. **G** Pathway enrichment analysis using the NCATS BioPlanet terms was performed for the three platforms with EGF/ERBB-associated terms indicated in blue. Source Data are provided as a Source Data file.
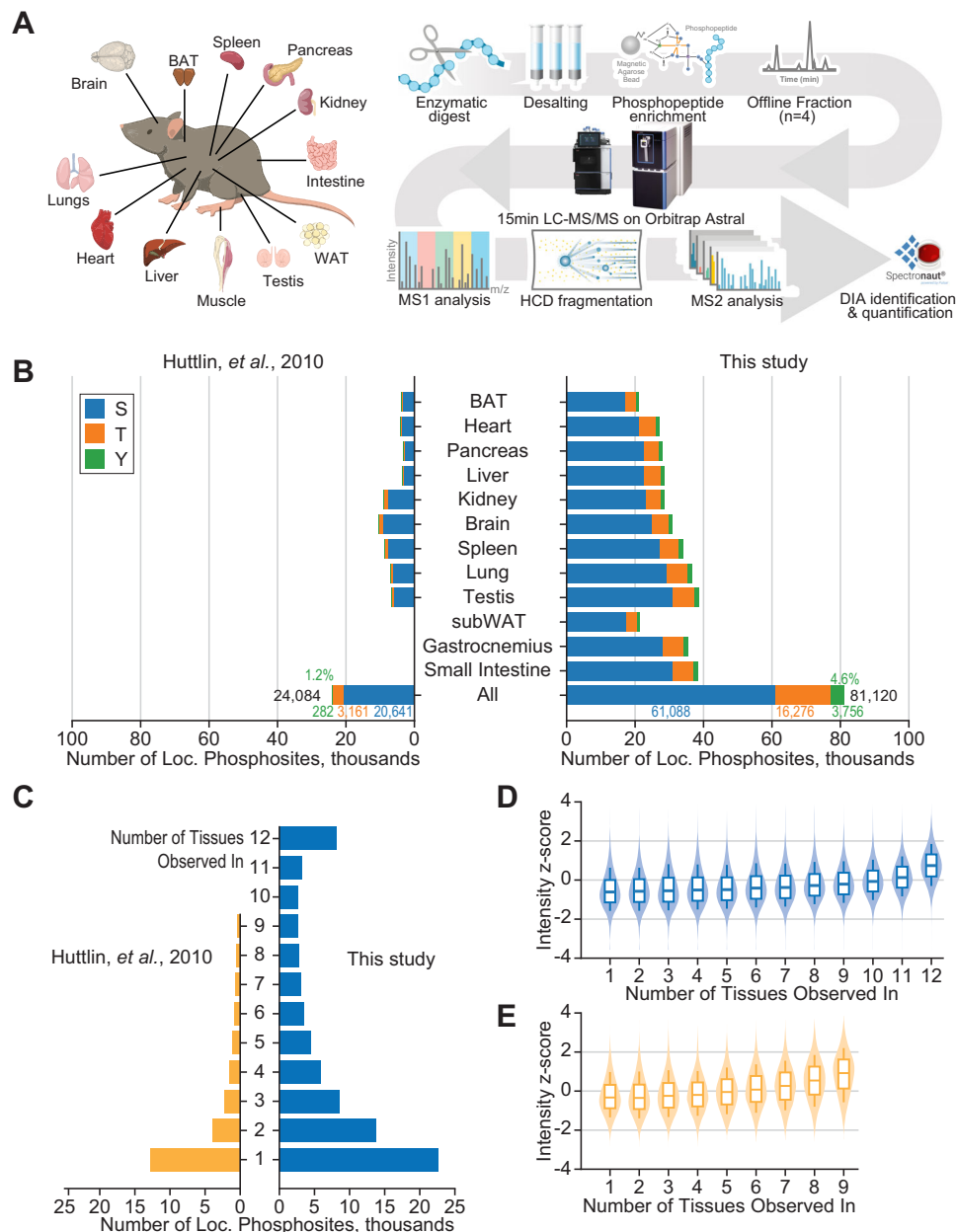
Fig. 4 | Mouse phosphorylation atlas workflow and results. A Mouse Phosphorylation Atlas Workflow. **B** Mouse Tissue Phosphoproteomic Analysis. Numbers of unique phosphorylation sites are shown for each tissue and the total unique sites with the fraction of S, T, and Y localizations indicated. The Huttlin et al. results were generated by researching the raw data in MaxQuant using the same protein database used in this study. **C** Tissue Specificity of Detected Phosphorylation sites. The y-axis indicates the number of tissues in which a phosphosite was detected.

D, E Intensity distributions for phosphorylation sites detected in a given number of tissues for this study and by Huttlin et al., respectively. For the distributions in (**D**) and (**E**), the 0.1, 0.25, 0.5, 0.75, and 0.9 percentiles are indicated with the lower error bar, lower box bound, center box line, upper box bound, and upper error bar, respectively. Each distribution was generated by down-sampling to 1000 data-points from the phosphosite categories in (**C**). Source data are provided as a Source Data file.

Furthermore, we compared our comprehensive phosphoprotein atlas of the mouse with the most recent multi-tissue study conducted by Giansanti and co-workers[51]. Similar to our sample preparation, the phosphopeptides from multiple tissues were offline fractionated and further concatenated into 4 samples. These samples were measured using a 90-min gradient DDA on the Q Exactive Orbitrap HF MS system. Despite a 6-fold shorter measurement time, the Orbitrap Astral detected 2–3 times more phosphosites per tissue, with improved agreement with our dataset compared to Huttlin et al. (median 64.2% versus 55.3%; Supplementary Fig. 6A). Interestingly, Giansanti's study shows a similar trend towards a higher number of sites that are present

across all tissues (Supplementary Fig. 6B). The portion of pY in this study is 1.1% (Supplementary Fig. 6C), which additionally highlights the capability of combining Astral sensitivity with DIA acquisition to enable the detection of 4.6% pY in our data.

Phosphorylation sites in the context of protein sequence and structure

Phosphorylation sites within proteins exhibit a specificity determined by both the sequence and structural motifs. In our investigation, we sought to unravel this specificity by clustering phosphorylation sites and their flanking regions. To achieve this, we
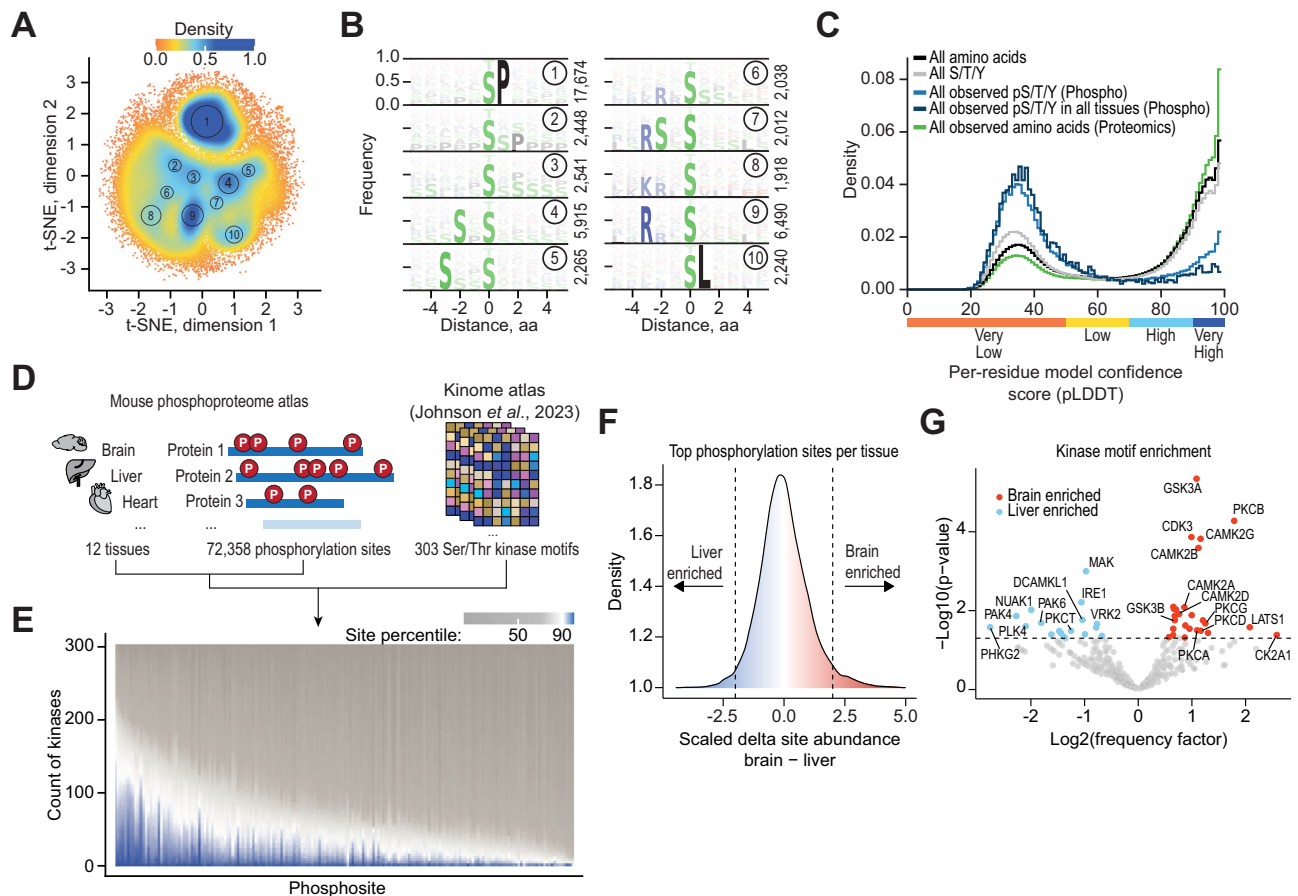
Fig. 5 | Sequence, structural and kinase specificity context of phosphosites.
A Two-dimensional representation of all phosphosites and their 5-amino acid flanking sequences, excluding the central amino acid from the comparison. Each cluster has been manually selected to emphasize the densest regions. **B** Sequence logo plot for all clusters depicted in (**A**). **C** Distribution of confidence scores for all amino acids, specifically S/T/Y, and for phosphorylated S/T/Y detected across all tissues. Source Data are provided as a Source Data file. **D** Our mouse phospho-proteome data derived from nine tissues was applied to the kinome atlas search tool. **E** All phosphorylation sites detected in our study are plotted on the x-axis, sorted by the number of kinases that scored higher than 90 for a specific site.

F Z-score transformed difference between abundances of shared phosphorylation sites in brain and liver tissue. Vertical dashed lines indicate thresholds for selection of phosphorylation sites that are used for kinase motif enrichment analysis. A chi-squared test was used to calculate p values after applying Haldane's correction.
G Based on the top sites per tissue, a motif enrichment analysis was performed and the resulting frequency of how often a kinase was predicted to act on a site was plotted on the x-axis, along with the p value on the y-axis. The scheme and types of analyses have been adapted from ref. 38. See Supplementary Data 4 for analysis results.

computed all pairwise comparisons of phosphosite sequences (site and five amino acids in both C and N-terminal direction) and projected this comparison onto a 2D plane using the t-SNE algorithm (Supplementary Fig. 7A). Notably, our analysis highlighted the importance of the phosphorylation amino acid as a major factor driving separation (Supplementary Fig. 7B). That is clusters 1-8 are Ser-based, while 9 is Tyr-based, and 10-12 are Thr-based. Intriguingly, upon repeating the analysis while excluding the phosphorylated amino acid from binary comparisons (Fig. 5A), we identified pre-dominant clusters organized by specific motifs (Fig. 5B), such as S/T-P (cluster 1), and RXX-S/T (cluster 9). Proline-directed protein kinases demonstrate a preference for phosphorylating serine or threonine residues immediately preceding a proline residue in proteins. Prominent examples of such kinases include extracellular signal-regulated kinases (ERKs) and cyclin-dependent kinases (CDKs). Leveraging annotations from the PhosphoSitePlus database[52], our analysis revealed that the majority of ERK/CDK annotated phosphorylation sites belong to the S/T-P cluster (Supplementary Fig. 7C). Note there is limited representation of Y in clusters 1 and 9, even though the phosphorylated residue was not considered in this analysis (Supplementary Fig. 7D). These data confirm known biology that Y is not a substrate for P and R-directed protein kinases[53].

To explore structural motifs, we mapped the 81,120 phosphorylation sites detected here onto protein structures predicted by AlphaFold[34]. The UniProt structural library (assembled using Alpha-Fold) contains structures with varying degrees of certainty as measured by a per residue confidence score (pLDDT) where 0 is very low and 100 is very high confidence. Note that lower confidence scores are indicative of either flexible or disordered protein regions[54]. Upon plotting the confidence score distribution for phosphorylated S, T, and Y sites, we observed a striking prevalence for these sites to be in low confidence regions (Fig. 5C), especially when compared to all S, T, and Y residues or even all amino acids. Phosphorylation sites that were detected in all tissues had an even slightly higher preference for these low-confidence regions. However, all amino acids observed in non-enriched proteomics experiments did not follow the same trend, emphasizing the specificity of this effect for phosphorylation sites. It is noteworthy that when plotted individually, S, T, and Y phosphorylation sites all exhibit the same trend (Supplementary Figs. 7E, F). These global results confirm previous work that suggests phosphorylation sites are enriched in intrinsically disordered regions[35,36]. In fact, some efforts have used intrinsically disordered regions to refine phosphorylation site prediction models[37]. Finally, although phosphorylation is more likely to occur in less structured regions, examination of the

surrounding environment shows the confidence scores are increasing with distance from the phosphorylated residue (Supplementary Fig. 7G). These data suggest further research on the role of structure and phosphorylation site is warranted.

Kinase predictions based on flanking regions of phosphorylation sites

Further leveraging our extensive dataset of phosphorylation sites across various tissues, we applied a recently published kinase prediction tool[38], which was developed by using synthetic peptide libraries to profile the substrate sequence specificity of 303 serine/threonine kinases (Fig. 5D). As expected, a systematic analysis of all phosphorylation sites in our dataset revealed a trend where many phosphorylation sites were assigned to a limited number of putative kinases; however, half of all phosphorylation sites were predicted to be targeted by 24 or more kinases (Fig. 5E). Furthermore, this approach enabled us to predict kinases for sites that showed tissue enrichment (Fig. 5F). By comparing enriched phosphorylation motifs between tissue pairs, we identified kinases with established activities in certain tissues (Supplementary Fig. 8, see Supplementary Data 4 for source data). For instance, comparing brain and liver enriched motifs revealed kinases such as GSK3A[55], CAMK subfamily 2 members[56–58], and PCKA/PCKB[59,60], which are known to modulate the brain phosphoproteome and have been implicated in neurological disease states (Fig. 5G). While these kinases have activities in other tissues, our dataset emphasized that the phosphorylation sites predicted to be clients of these kinases are strongly overrepresented in brain as compared to liver tissue. Additionally, we identified PHKG2 as the kinase with the strongest frequency factor for liver tissue, consistent with this enzyme's role as a liver isoform critical for glycogen breakdown[61]. Our analysis suggests additional kinases with probable tissue-specific activities, highlighting the potential of our resource to facilitate the discovery of such tissue-specific kinase functions.

Phosphorylation sites on mitochondrial proteins

We next sought to investigate how our comprehensive analysis of phosphorylation patterns in the mouse proteome can be further applied in the biochemical and biological context. We first compared our dataset to the mouse phosphorylation dataset from PhosphoSitePlus[52], revealing a capture of 38,454 previously unidentified phosphorylation sites, constituting 53% of the total identified sites within our dataset (Fig. 6A). Upon categorizing phosphorylation sites, we observed a significant representation of mitochondrial proteins, with 55% of known mitochondrial proteins harboring at least one phosphorylation site (Fig. 6B). This observation aligns with the growing appreciation for protein phosphorylation in regulating mitochondrial function[62]. An exemplary mitochondrial pathway whose protein phosphorylation sites showed the most dynamic pattern across all tissues was the tricarboxylic acid cycle (Supplementary Fig. 9A) with the majority of sites being not being detected until this study (Supplementary Fig. 9B), setting the stage for further exploration into the understudied regulation of this key metabolic pathway[63].

The occurrence of phosphorylation sites was variable across tissues in the entire dataset (Supplementary Fig. 9C) as well as in the mitochondrial subset of our data (Fig. 6C), indicating that tissue-specific phosphorylation patterns exist within the murine mitochondrial phosphoproteome. To investigate this in more detail, we calculated the intersections of all mitochondrial phosphorylation sites across tissues and found the liver to harbor the most unique phosphorylation sites (Fig. 6D). Moreover, these unique liver phosphorylation sites showed the largest proportion of mitochondrial matrix proteins, suggesting that phosphoregulation of proteins in this subcompartment is particularly important in this tissue. We identified carbamoyl phosphate synthetase I (CPS1) to harbor the most phosphorylation sites in this group (Fig. 6E).

Notably, 40 phosphorylation sites were detected on CPS1, which is the first and rate-limiting enzyme of the urea cycle. None of these phosphorylation sites have been functionally characterized, and 22 of these sites have not been reported previously (Fig. 6F), underscoring the significant discovery potential of our dataset. Deficiency of human CPS1 results in hyperammonemia ranging from neonatally lethal to environmentally induced adult-onset disease in affected individuals[64]. The phosphorylation site S913, located within the L2 integrating domain, overlaps with a known severe disease-causing missense variant in CPS1 (p.S913L)[65], indicating the potential regulatory significance of these sites in CPS1.

Our dataset also provides enhanced phosphotyrosine coverage without specific enrichment methods (Fig. 4B), revealing sites that may be functionally relevant. N-acetyl-L-glutamate (NAG) is an essential allosteric activator of CPS1. By binding to the C-terminal L4 domain of CPS1, NAG induces long-range conformational changes, impacting distant catalytic activities[66]. We identify Y1450 as a conserved phosphorylated residue in this domain, which is located proximal to the key R1453 residue that is required for NAG binding and is the location of two patient variants (c.4357C>T and c.4358G>A resulting in p.R1453W and p.R1453Q, respectively)[67] (Fig. 6G). Previous work suggested that a distinct post-translational modification of Y1450 (nitration) may impact CPS1 activity by obstructing NAG binding[68]. Structural modeling demonstrates shifts in this Y1450-R1453-containing helix of CPS1 upon NAG binding, highlighting the conformational change of the R1453 residue and its proximity to the Y1450 residue (Fig. 6H).

Most dynamic phosphoregulation in brain tissue

The brain is characterized by a diverse array of expressed kinases and phosphatases and its proteins have a higher tendency to be phosphorylated as compared to other tissues[49]. Our data substantiates this observation, as brain tissue showed the highest prevalence of unique phosphorylation sites (i.e., sites defined as exclusive to a singular tissue, Fig. 7A). Moreover, we investigated all pan-tissue expression of phosphorylation sites (sites detected across all twelve tissues) by a z-score analysis, enabling comparison of each phosphosite to the average organismal values, and found brain tissue to exhibit the strongest abundance changes of phosphorylation sites (Supplementary Fig. 10A). We then applied the same z-score-based principle on the entire dataset (including imputed values for undetected sites in some tissues) to impartially identify sites undergoing significant changes and observed a normal distribution in z-score patterns for all tissues, except the brain, which displayed a pronounced enrichment of phosphorylation sites with elevated z-scores (Supplementary Fig. 10B). By applying the same absolute z-score threshold of 2, we found 9541 phosphorylation site outliers in the brain (Fig. 7B), accounting for 23% of all outliers across all twelve tissues. Further prioritizing these brain sites based on z-score ranking within the mitochondrial phosphoproteome, we identified serine 298 of Optic atrophy protein 1 (OPA1) as a top hit (Fig. 7C).

OPA1 facilitates multiple functions in mitochondria including membrane fusion, cristae biogenesis, mitochondrial DNA maintenance, and respiration[69]. Importantly, mutations in OPA1 are the most common cause of dominant optic atrophy (DOA). In our data, we detected multiple phosphorylation sites of OPA1 across the twelve tissues (Supplementary Fig. 10C), none of which are functionally characterized as documented in PhosphoSitePlus[52]. These sites are localized within the GTPase and PH domain of OPA1, which are also common areas for missense variants in patients with DOA[70]. Additionally, many of the detected phosphorylation sites have not been previously reported in mouse, among which is the S298 site within the first GTP-binding domain (G1/P-loop) of the GTPase (Fig. 7D). The consensus motif GxxxxGKS/T of G1/P-loop is highly conserved across species and other proteins of the dynamin superfamily (Fig. 7E). Further, the Q and S residue within this motif are important for the
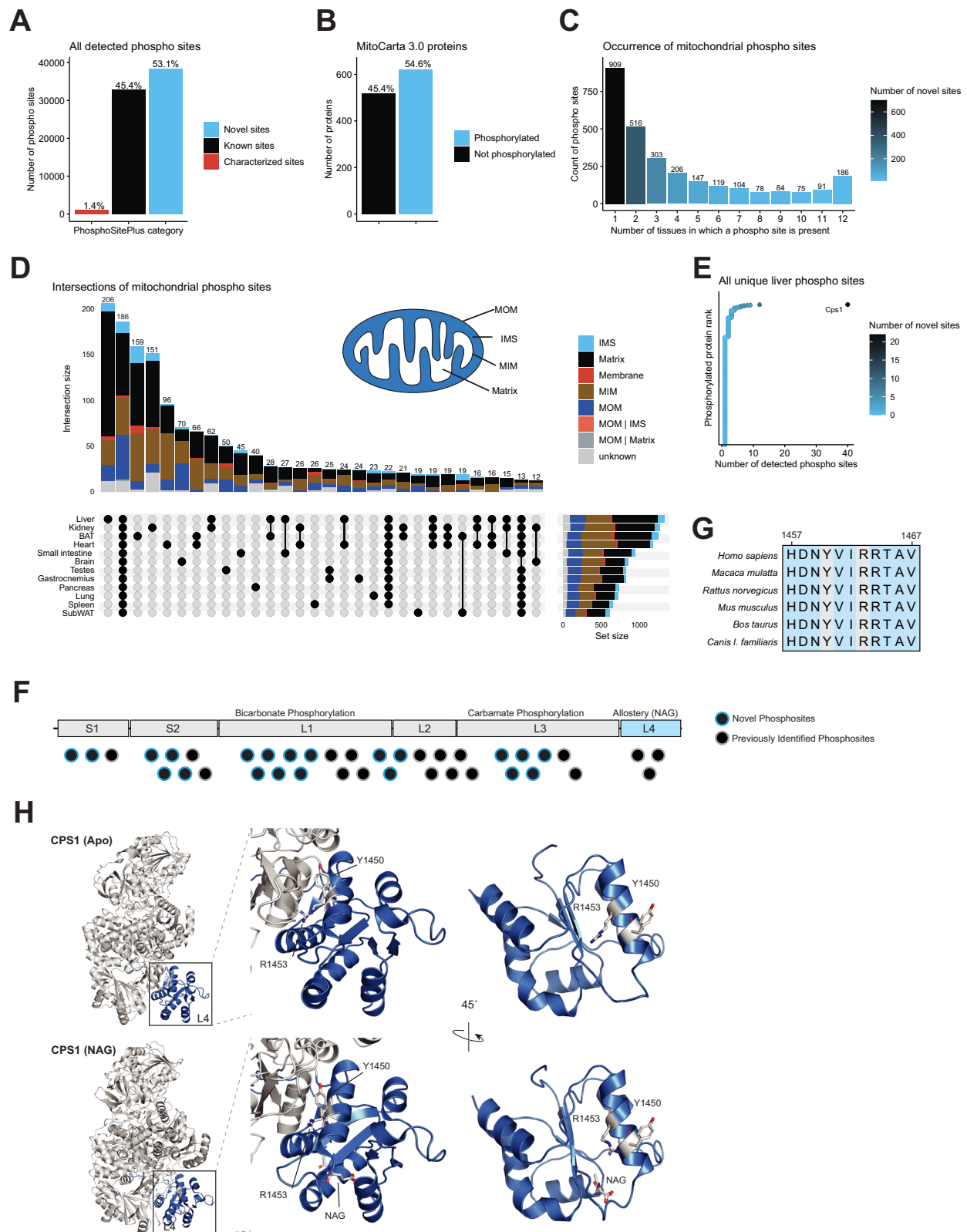
assembly stimulated GTPase activity as shown for dynamin[71]. Notably, S298 has been identified as a site of missense mutations in DOA patients (c.893G>A, p.S298N[72]; c.892A>G, p.S298G[73]), suggesting its importance for OPA1 function, too. Furthermore, a S298A mutation reduces GTPase activity and dimerization in vitro[74]; the S298N mutation abolishes respiratory growth in *S.cerevisiae* with mitochondrial DNA depletion and altered mitochondrial morphology[75]. Our

structural modeling of the OPA1's GTPase domain shows that S298 is close to the nucleotide-interacting G300 of the G1/P-loop and the nucleotide, suggesting that phosphorylation may affect nucleotide binding or hydrolysis (Fig. 7F). Taken together, current evidence clearly underscores the functional importance of S298 for OPA1 activity, but the effect and relevance of phosphorylation we have identified on S298 requires further study.

Fig. 6 | Mitochondrial phosphoproteomics reveals novel liver-specific phosphorylation site. A Bar plot of all detected phosphorylation sites in our study stratified into categories that are directly derived from the PhosphoSitePlus database (downloaded August 22, 2023). **B** Mouse MitoCarta 3.0[82] proteins with at least one detected phosphorylation site in our data versus the remainder. **C** Bar plot indicating the number of mitochondrial phosphorylation sites that occur in a specific number of tissues. **D** Intersections of phosphorylation sites on mitochondrial proteins per tissue subsets. Intersection sizes of 12 or more are shown. Submitochondrial localization is derived from MitoCarta 3.0. **E** Dot plot displaying all

proteins that harbor phosphorylation sites unique to liver tissue. The highest ranked protein harbors the most phosphorylation sites as indicated on the x-axis. Novel sites are according to the PhosphoSitePlus database. **F** Schematic representation of the CPS1 peptide chain with all detected phosphorylation sites indicated as novel or previously identified. **G** Sequence alignment of CPS1 orthologues using Clustal Omega[81]. Patient variant residue and phosphorylation site residue of interest are in gray. **H** Structural modeling based on structures (PDB: 5DOT (Apo), 5DOU (NAG)) from RCSB PDB[66]. Source Data are provided as a Source Data file.
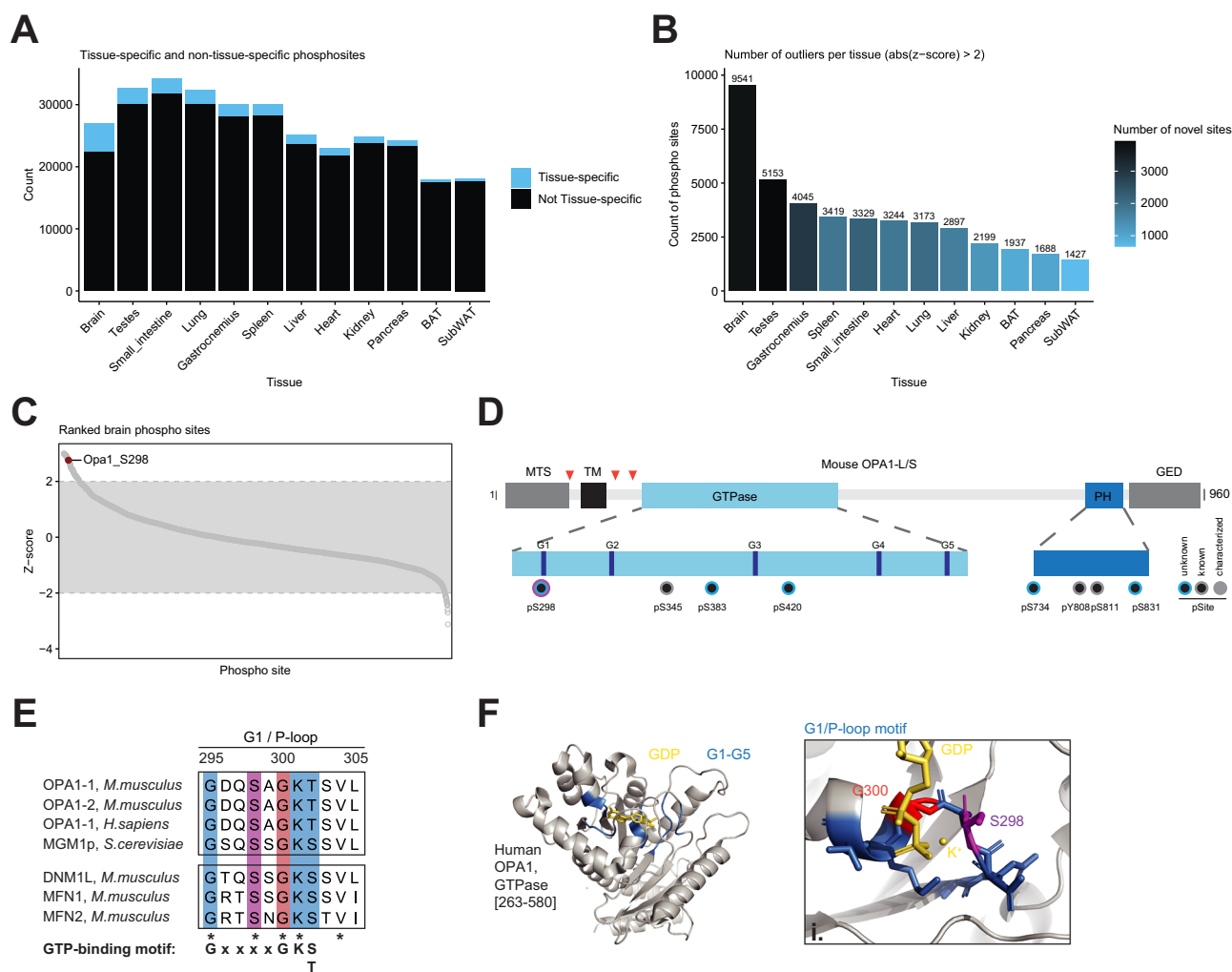

Fig. 7 | Unique phosphorylation sites in brain tissue. A Bar plot indicating all sites (black) that were detected per tissue, highlighting the sites that are unique in each tissue (blue). **B** Number of outliers per tissue based on z-score analysis with an absolute z-score cut-off of 2. **C** Ranked mitochondrial phosphorylation sites in brain based on z-score, highlighting the OPA1 top site. **D** Protein domain representation of mouse OPA1, including mitochondrial transition signal (MTS), transmembrane domain (TM), GTPase, PH, and GED domain. Red triangles represent sites of proteolytic cleavage generating the long and short proteo-forms of OPA1. Note, that all identified phosphosites are common to all OPA1 isoforms. Below, details of GTPase and PH domain with indicated phosphorylation sites (pS/Y) found in brain tissue

(See Supplementary Fig. 8C for presence of OPA1 phosphorylation sites in other tissues). **E** Sequence alignment of first GTP-binding site/P-loop among mouse OPA1 isoforms and homologs in human and yeast, as well as related dynamin family members, DNM1L, MFN1, and MFN2, in mouse. Conserved residues are marked with *, including serine 298 in mouse OPA1. **F** Structure modeling of human OPA1 GTPase domain [263-580] based on RCSB PDB structure 6JTG[74]. Left, Modeling of the entire domain with highlighted GTP-binding domains (G1-5) in blue and bound GDP (yellow). Right, Detail (i.) of the G1/P-loop domain containing the discussed S298 (magenta) and the nucleotide-binding G300 (red). K+ (yellow) stabilizes the vicinity of the nucleotide. Source Data are provided as a Source Data file.

Discussion

Here we demonstrated that the Orbitrap Astral mass spectrometer provides a platform for fast and deep phosphoproteome analysis. Key to empowering this application is the ability of the system's high MS/MS duty cycle to enable DIA data collection with narrow isolation windows (2 Th). This capability, combined with high mass accuracy,

mass resolution, and sensitive ion detection of the Astral analyzer allow confident phosphosite localization from low abundance peptides as evinced by the performance of the system for low overall phosphopeptide loads. Altogether this approach allowed for the collection of the deepest mouse phosphoproteomic atlas to date following only a half day of data collection. Nonetheless, the uniquely

large mouse phosphoproteome atlas defined here provides an opportunity to explore the sequence and structural context of kinase activity.

From the phosphoproteome analysis we discovered that, despite its bacterial origins, over fifty percent of mitochondrial proteins carry at least one site of phosphorylation site. An interesting and potentially important example is the CPS1 protein on which we detected 22 unreported sites. Notably, one of these sites, Y1450, is proximal to the NAG binding domain and near the location of known patient variants—suggesting potential clinical relevance.

As exemplified by the above example, this technology permitted detection of thousands of pY sites without the typical pY-specific antibody enrichments. Doubtless the sensitivity and depth afforded by the Orbitrap Astral instrument can allow direct access to these most critical, dynamic, and low-abundance phosphorylation events. Finally, we note that while all tissues exhibit a subset of unique phosphorylation events, the brain is distinguished and contains nearly 10,000 unique phosphorylation sites.

Protein phosphorylation analysis presents many challenges for mass spectrometric analysis; however, we demonstrate here that the Orbitrap Astral resolves many of these limitations and permits fast and deep phosphoproteome analysis. We suppose that these performance characteristics will be translated to the analysis of other post-translational modifications including acetylation, glycosylation, methylation, ubquitinylation, etc. Furthermore, we expect use of extensive fractionation, more tissues, and multiple proteases would allow detection of even more phosphorylation sites and increase tissue-to-tissue phosphorylation site reproducibility.

Methods
HEK293T cell preparation
HEK293T cells (ATCC CRL-3216) were cultured in Dulbecco's modified Eagle's medium (DMEM) (Gibco, 11995-065) with 1% penicillin/streptomycin (Thermo Fisher Scientific, 15-140-122) and 10% fetal bovine serum (FBS) (HyClone, 89133-098) at 37 °C and 5% CO_2. Cell line authentication was performed by the commercial distributor. Culture media was replaced every 24 h. Cells were expanded to appropriate cell number, detached from tissue culture plate with 0.05% trypsin-EDTA (Gibco, 25300062), washed once with phosphate buffered saline (PBS) (Gibco, 02-0119-0500), cell number determined, and 20×10^7 cells pelleted. Cell pellet was immediately stored at −80 °C until use. The cells were low passage number and tested negative for mycoplasma contamination. Frozen cell pellets were resuspended in 5.4 M guanidine hydrocholoride (from Sigma Life Science, 8 M, pH 8.5, G7294-100mL) in 100 mM Tris, pH 8 (Invitrogen, 1 M Tris pH 8.0, 0.2 μm filtered, AM9856) via vortexing, followed by heating in a sand bath for 5 min at 105 °C prior to brief (10–15 s) sonication with a probe sonicator. The sample was diluted with the guanidine buffer above to give a ~1.5 mg/mL estimated protein concentration via NanoDrop (Thermo Scientific) prior to beginning digestion.

EGF-stimulated HeLa cell preparation
HeLa cells (ATCC CCL-2) were cultured in Dulbecco's Modified Essential Medium (DMEM) (Gibco, 11995-065) with 10% FBS (HyClone, 89133-098) and 1% penicillin/streptomycin (Thermo Fisher Scientific, 15-140-122) in 37 °C/5% CO_2 incubator. Cell line authentication was performed by the commercial distributor. Cells were resuspended in PBS (Gibco, 02-0119-0500) with or without 100 ng/mL human Endothelial Growth Factor (Thermo, AF-100-15) for 15 min at room temperature. Cells were gently resuspended every 5 min during incubation. After incubation, cells were washed twice with PBS and stored in −80 °C until use. Protein extraction was performed as described above for the HEK293T samples. Protein concentrations were estimated via protein BCA (Pierce, 23235). Three biological replicates each were prepared for the control and EGF treatment groups.

Yeast cell preparation
Saccharomyces cerevisiae S288C-derivative strain, BY4741 (SRD GmbH, Y00000) was cultured in triplicate for ~5 generations into log phase in rich YPD medium at 30 °C. Authentication was performed by the commercial distributor. Cells were centrifuged at 2000 g, 3 min, rinsed in water, transferred to smaller tubes and centrifuged at 1600 g, 3 min, prior to snap-freezing in liquid nitrogen. Frozen cell pellets were resuspended in 8 M urea Sigma-Aldrich, U5378) in 100 mM Tris, pH 8 (Invitrogen, 1 M Tris pH 8.0, 0.2 μm filtered, AM9856) and vortexed with glass beads (425-600 μm, Sigma-Aldrich, G8772-500G) to lyse (2 min of total vortexing with 30 s vortexing followed by 30 s on ice). Protein concentrations were estimated via protein BCA (Pierce, 23235).

Mouse tissue preparation
All experiments were performed in accordance with the National Institute of Health Guide for the Care and Use of Laboratory Animals and were approved by the Animal Care and Use Committee at the University of Wisconsin-Madison. Mice were kept on a 12-h light–dark cycle at 23 °C and within a humidity range between 30% and 50%. Mice were fed standard chow diet (Teklad #2018). Six-week-old male C57BL6/J mice ($n = 3$) were euthanized by cervical dislocation and tissues were immediately collected, and flash frozen in liquid nitrogen. A total of twelve tissues (pancreas, small intestine, spleen, liver, kidney, testes, heart, lung, subcutaneous white adipose tissue (WAT), brown adipose tissue (BAT), gastrocnemius, and brain) were collected. All tissues were stored at −80 °C prior to cryo-pulverization. For each tissue, samples from three mice were pulverized together into a fine power under liquid nitrogen (i.e., one biological replicate per tissue). For each pulverized tissue, ~60 mg frozen wet weight was resuspended in 4 mL of 5.4 M guanidine hydrocholoride (Sigma Life Science, 8 M, pH 8.5, G7294-100 mL) in 100 mM Tris, pH 8 (Invitrogen, 1 M Tris pH 8.0, 0.2 μm filtered, AM9856) with Pierce™ Phosphatase Inhibitor Mini Tablets (A32957) with one tablet/10 mL. Tissue samples were vortexed and sonicated for 20 min in a bath sonicator (chilled to 4 °C) to homogenize. A probe sonicator was used briefly to sonicate samples on ice as needed. The protein concentration was estimated via protein BCA (Pierce, 23235) and samples were diluted with the guanidine hydrochloride buffer above to give a protein concentration of ~2 mg/mL prior to digestion.

Protein digestion
After extracting proteins from human cells, yeast cells, or mouse tissues, methanol (Optima LC/MS grade, Fisher Scientific) was added to 90% (v/v) to precipitate protein and samples were vortexed prior to centrifugation at 4000 g for 15 min. The supernatant was removed, and the pellet was resuspended in 8 M urea (Sigma-Aldrich, U5378), 100 mM Tris (Invitrogen, 1 M Tris pH 8.0, 0.2 μm filtered, AM9856), 10 mM TCEP (Sigma-Aldrich, C4706-2G), 40 mM 2-chloroacetamide (Sigma-Aldrich, ≥98%, C0267-100G) pH 8 at ~1.5 mg protein/mL. Lysyl Endopeptidase (LysC, 100369-826, VWR) was added at a ratio of 1:50 enzyme:protein and gently rocked at ambient temperature for 4 h, followed by the dilution of the solution to 2 M urea with 100 mM Tris, pH 8 (Invitrogen, 1 M Tris pH 8.0, 0.2 μm filtered, AM9856). Promega Sequencing Grade Modified Trypsin (V5113) was added at a ratio of 1:50 enzyme:protein and incubated overnight. Following overnight digestion, the solution was acidified to <pH 2 with 10% trifluoroacetic acid (Sigma-Aldrich, HPLC grade, >99.9%) to quench the digestion. The sample was then centrifuged at 4000 g for 10 min to remove particulate matter prior to desalting with a Strata-X 33 μm polymeric reversed phase SPE cartridge. Peptides were dried via a SpeedVac (Thermo Scientific) and stored at −80 °C until phosphopeptide enrichment or, in the case of the mouse proteomics experiments, until fractionation.

Phosphopeptide enrichment

Phosphopeptides were enriched from digested peptides using MagReSyn Ti-IMAC HP beads (ReSyn Biosciences, MR-THP005). A volume of 100 μL beads were used per 1 mg of peptides. Input peptide masses of 2–3 mg were utilized for human and yeast enrichments, and input peptide masses ranging from -0.3 mg to 1.2 mg were utilized for the different mouse tissues. Beads were washed three times with 1 mL 80% acetonitrile (Optima LC-MS grade, Fisher Scientific)/6% trifluoracetic acid (Sigma-Aldrich, HPLC grade, >99.9%) prior to resuspending the sample in 1 mL 80% acetonitrile (Optima LC-MS grade, Fisher Scientific)/6% trifluoracetic acid (Sigma-Aldrich, HPLC grade, >99.9%) and vortexing the sample with the beads for 1 h. After the 1-h of vortexing, the beads were washed three times with 1 mL 80% acetonitrile (Optima LC-MS grade, Fisher Scientific) /6% trifluoracetic acid (Sigma-Aldrich, HPLC grade, >99.9%), once with 1 mL 80% acetonitrile, once with 1 mL 80% acetonitrile (Optima LC-MS grade, Fisher Scientific)/0.5 M glycolic acid (Sigma-Aldrich, 99%, 124737-500 G), and three times with 1 mL 80% acetonitrile (Optima LC-MS grade, Fisher Scientific). The phosphopeptides were eluted from the beads with the addition of 300 μL 50% acetonitrile (Optima LC-MS grade, Fisher Scientific)/1% ammonium hydroxide (28% in H_2O, ≥99.99% trace metals basis, Sigma-Aldrich), followed by a second elution with another 300 μL 50% acetonitrile/1% ammonium hydroxide. The samples were acidified via addition of 15 μL 10% trifluoroacetic acid (Sigma-Aldrich, HPLC grade, >99.9%). The samples were then dried down in a SpeedVac (Thermo Scientific) prior to being resuspended in 0.2% trifluoroacetic acid (Sigma-Aldrich, HPLC grade, >99.9%) and desalted as described above. Desalted phosphopeptide samples were dried and resuspended in 0.1% formic acid (Fisher Scientific, LC-MS grade). The phosphopeptide concentration was estimated via NanoDrop (Thermo Scientific). The HEK293T phosphopeptides were pooled to generate a sample for method evaluation. Each mouse phosphopeptide sample was fractioned as described below.

High-pH peptide fractionation

High-pH fractionation of peptides was performed on an Agilent 1260 Infinity BioInert LC with an automated fraction collector. A 20-min method was performed on a Waters XBridge, Peptide BEH C18, 3.5 μm, 130 Å, 4.6 mm × 150 mm column with a flow rate of 0.8 mL/min. Mobile phase A and B were 10 mM ammonium formate (Sigma-Aldrich, >99/0%, LC-MS grade, 70221-100GF), pH 10 and 20% 10 mM ammonium formate (pH 10)/80% methanol (Optima LC/MS grade, Fisher Scientific), respectively. The gradient went from 0 to 35%B from 0 to 2 min, 35 to 75%B from 2-8 min, 75% to 100%B from 8 to 13 min, followed by washing at 100%B from 13 to 15 min and equilibration at 0%B from 15 to 20 min. UV absorbance at 210 and 280 nm was recorded. For HEK293T spectral library generation and mouse proteomics samples, 16 fractions were collected from 5 to 18 min and concatenated into the final fractions by combining fraction 1 and 9, fraction 2 and 10, etc., resulting in 8 final fractions. For mouse phosphoproteomics samples, 8 fractions were collected from 5 to 18 min and concatenated into the final fractions by combining fraction 1 and 5, fraction 2 and 6, etc. Samples were dried down in a SpeedVac (Thermo Scientific) prior to being resuspended in 0.1% formic acid (Fisher Scientific, LC-MS grade) for LC-MS analysis.

Synthetic phosphopeptide dilution series preparation

Five sets of phosphopeptide standards were acquired: SpikeMix PTM-Kit 52 (JPT, SPT-PTM-POOL-Yphospho-1), SpikeMix PTM-Kit 54 (JPT, SPT-PTM-POOL-STphospho-1), MS PhosphoMix 1 (Sigma, MSP1L-1VL), MS PhosphoMix 2 (Sigma, MSP2L-1VL), and MS PhosphoMix 3 (Sigma, MSP3L-1VL). The standards were reconstituted in 0.2% formic acid/20% acetonitrile/80% water via vortexing. The standards were pooled into an equimolar mixture of the 225 total phosphopeptide standards. The pooled equimolar mixture was then diluted and mixed with the yeast phosphopeptide sample to construct a dilution series comprised of five points of four-fold dilutions starting at 10,000 amol, resulting in total quantities loaded onto the column of 10000, 2500, 625, 156.25, and 39.0625 amol per phosphopeptide standard along with a constant yeast phosphopeptide load of 250 ng. A summary of all the phosphopeptide standards used in the analysis is provided in Supplementary Data 1.

LC-MS operation for phosphoproteomics with Orbitrap Astral analysis

Nanoflow capillary columns (75 μm I.D., 360 μm O.D.) with pulled nanoESI emitters were packed to 40 cm at high pressures with C18 1.7 μm diameter, 130 Å pore size BEH C18 particles (Waters) as previously described[39]. A slurry of C18 particles dissolved in chloroform was loaded into a custom-built packing setup with a high pressure pneumatic pump (Haskel) and ultrahigh-pressure capillary fittings (HiP) and packed into the column while slowly ramping up to 30,000 psi, holding at 30,000 psi for 4 h, and then allowed to depressurize slowly. Samples were analyzed with a Vanquish Neo UHPLC (Thermo Scientific) coupled to an Orbitrap Astral mass spectrometer (Thermo Scientific) using a NanoSpray Flex source (Thermo Scientific). A source voltage of 2000 V was used for all experiments. Mobile phase A and B were 0.1% formic in water (Fisher Scientific, Optima LC-MS grade) and 0.1% formic acid/80% acetonitrile (Fisher Scientific, Optima LC-MS grade), respectively. The column was heated to 50 °C with the Column Oven PRSO-V2 (Sonation Lab Solutions) and the flow rate was set to 400 nL/min at the start of the method to decrease delay time and turned to 300 nL/min at the start of the active gradient. Initial conditions of 2%B were ramped to 14% from 0 to 5 min. The active gradient was generally set to 14% to 54%B with curve type 6 beginning at 5.2 min, with the exact %B settings adjusted for each active gradient length to evenly distribute peptide signal across the gradient. The column was washed for 5 min at 100%B and 400 nL/min at the end of the gradient, followed by fast equilibration on the Vanquish Neo LC with an upper-pressure limit of 1100 bar.

For initial DDA experiments on the Orbitrap Astral MS, MS^1 spectra were collected in the Orbitrap every 0.6 s at a resolving power of 240,000 at m/z 200 over m/z 350–1350 with a normalized AGC target of 300% (3e6 charges) and a maximum injection time of 10 ms. The MIPS filter was applied with Peptide mode and "Relax Restrictions when too few Precursors are Found" set to True. Precursors were filtered to charges states 2-6. A Dynamic Exclusion filter was applied with 10 s duration and 10ppm low and high mass tolerance and exclude isotopes set to True. An intensity filter was applied with a minimum precursor intensity of 5000 required for selection. MS^2 scans were collected in the Astral mass analyzer with an isolation window of 0.7 m/z, normalized collision energy of 27, a scan range of 150–2000 m/z, an AGC target of 100% (1e4 charges), and a maximum injection time of 10 ms.

For DIA experiments on the Orbitrap Astral MS, MS^1 spectra were collected in the Orbitrap every 0.6 s at a resolving power of 240,000 at m/z 200 over m/z 380–980. The MS^1 normalized AGC target was set to 300% (3e6 charges) with a maximum injection time of 10 ms. DIA MS^2 scans were acquired in the Astral analyzer over a 380–980 m/z range with a normalized AGC target of 500% (5e4 charges) and a maximum injection time of 3.5 ms and an HCD collision energy setting of 27% and a default charge state of +2. Window placement optimization was turned on. Isolation widths of 2 Th and active gradient lengths of 30 min were used with HEK293T fraction analysis for HEK293T spectral library generation. Isolation bin widths of 2 Th and 4 Th with 1 Th overlap were compared for HEK293T phosphoproteomics method evaluation. Isolation widths of 2 Th were used for mouse phosphoproteomics experiments. Comparisons of different DIA m/z ranges and AGC targets are shown in Supplementary Fig. 2D–G.

Eight HEK293T phosphopeptide fractions were injected once for library generation. For

HEK293T shotgun experiments, one injection replicate per loading mass was performed for the loading mass experiment. Three injection replicates per method were performed for the gradient length and isolation width experiments. For the phosphopeptide standard analysis, each concentration point was analyzed with three injection replicates. For the EGF treatment vs control HeLa phosphopeptide and proteomics experiments, each of the three biological replicates per condition was injected once. For the HeLa phosphopeptide method evaluation experiments, two injection duplicates were performed per method. For the mouse phosphopeptide analysis, each tissue sample was fractioned into 4 fractions and injected once.

LC-MS operation for benchmarking on the Orbitrap Ascend

Nanoflow capillary columns (75 μm I.D., 360 μm O.D.) with pulled nanoESI emitters were packed to 40 cm at high pressures with C18 1.7 μm diameter, 130 Å pore size BEH C18 particles (Waters) as previously described[39]. See above for column packing description. Samples were analyzed with a Vanquish Neo UHPLC (Thermo Scientific) coupled to an Orbitrap Ascend mass spectrometer (Thermo Scientific) using a NanoSpray Flex source (Thermo Scientific) incorporating a homebuilt column heating compartment. The column was heated to 50 °C. The flow rate was set to 400 nL/min at the start of the method to decrease delay time and turned to 300 nL/min at the start of the active gradient. Initial conditions of 2%B were ramped to 14% from 0 to 5 min. The active gradient was generally set to 14–54%B with curve type 6 beginning at 5.2 min, with the exact %B settings adjusted for each active gradient length to evenly distribute peptide signal across the gradient. The column was washed for 5 min at 100%B and 400 nL/min at the end of the gradient, followed by fast equilibration on the Vanquish Neo LC with an upper pressure limit of 1100 bar.

DIA experiments on the Orbitrap Ascend utilized similar method settings to those previously described by Bekker-Jensen et al.[41], but with the same m/z range as utilized on the Orbitrap Astral to enable direct comparisons. Briefly, MS[1] spectra were collected in the Orbitrap at a resolving power of 60,000 at m/z 200 over m/z 380–980 with a AGC target of 250% and maximum injection time of 123 ms. DIA MS[2] scans were collected in the Orbitrap with 15,000 at 200 m/z resolving power, a scan range of 150–2000 m/z, an AGC target of 200%, a maximum injection time of 27 ms, and a collision energy of 30%. A m/z range from 380–980 m/z was iterated through with 14 Th isolation windows with 1 Th overlap.

For the phosphopeptide standard analysis, each concentration point was analyzed with three injection replicates. For the EGF treatment vs control HeLa phosphopeptide and proteomics experiments, each of the three biological replicates per condition was injected once.

LC-MS operation for proteomics with Orbitrap Eclipse analysis

The chromatography setup described above for the Orbitrap Ascend analysis was utilized for the Orbitrap Eclipse experiments. A source voltage of 2000 V was used for all experiments. Mobile phase A and B were 0.2% formic acid (Optima LC-MS grade, Fisher Scientific) in water (Optima LC-MS grade, Fisher Scientific) and 0.2% formic acid/80% acetonitrile (Optima LC-MS grade, Fisher Scientific), respectively. The column was heated to 50 °C and a flow rate of 300 nL/min was used. Initial conditions of 0%B were ramped to 6%B from 0 to 1 min. The active gradient was set to 6% to 52%B with curve type 6 from 1 to 73 min. The column was washed for 5 min at 100%B, followed by fast equilibration on the Vanquish Neo LC with an upper pressure limit of 1150 bar.

For DDA experiments on the Orbitrap Eclipse MS, MS[1] spectra were collected in the Orbitrap analyzer every 0.6 s at a resolving power of 240,000 at m/z 200 over m/z 300–1350 with a normalized AGC target of 250% and a maximum injection time of 50 ms. The MIPS filter was applied with Peptide mode and "Relax Restrictions when too few Precursors are Found" set to True. Precursors were filtered to charges states 2-5. A Dynamic Exclusion filter was applied with 10 s duration and 25ppm low and high mass tolerance and exclude isotopes set to True. MS[2] scans were collected in the ion trap with an isolation window of 0.5 m/z, normalized collision energy of 25, a scan range of 150–1350 m/z, an AGC target of 300%, and a maximum injection time of 14 ms.

Each mouse tissue was fractioned into 8 fractions, each of which was injected once for DDA proteomics.

DDA data processing

Proteomics DDA data was processed in MaxQuant 2.4.9.0 with default parameters[40]. Phosphoproteomics DDA data was processed with the same version of MaxQuant with default parameters and Phospho (STY) enabled as a variable modification. The "Phospho (STY)Site.txt" was used for analysis with a filter for localization probabilities greater than or equal to 0.75.

DIA data processing

Phosphoproteomics DIA data was processed in Spectronaut version 17.6.230428.55965 or 18.6.231227. A HEK293T spectral library was generated in Spectronaut (searching against the human proteome database (Swiss-Prot and TrEMBL) downloaded from UniProt on 15-Jan-2023) from the eight HEK293T phosphopeptide fractions analyzed with a 30-min active gradient 2 Th DIA method and the library was used to search the HEK293T phosphoproteomics method evaluation experiments shown in Fig. 2. Note that the entrapment search in Fig. 2G and the phosphoproline variable modification search in Fig. 2H were performed as a library-free search. Mouse phosphoproteomics data was searched with the directDIA mode. Default Spectronaut search parameters were used with a variable Phospho (STY) modification included. The PTMSiteReport file was used for analysis of phosphorylation sites. To count the number of unique phosphorylation sites detected within an experiment, the PTMSiteReport was filtered for rows with "Phospho (STY)" in the "PTM.ModificationTitle" column and values greater than or equal to 0.75 in the "PTM.Site-Probability" column. The data table was then sorted according to "PTM.CollapseKey" and filtered for unique values in the "PTM.Group". At this stage, there can be rows with the same values of "PTM.CollapseKey" that are not filtered by "PTM.Group" due to PTM grouping differing across multiple files, so the dataset was filtered for unique values of "PTM.CollapseKey" to determine the total number of unique phosphorylation sites measured in a single file. Mouse phosphoproteomics data was searched against the mouse proteome downloaded from UniProt from Swiss-Prot and TrEMBL (downloaded on 25-Aug-2023). For the mitochondrial biology investigation, a separate search was performed with just the Swiss-Prot database to facilitate the analysis.

HEK293T phosphoproteomics DIA data was also processed in Proteome Discoverer 3.1.0.638 with CHIMERYS. Numbers of localized phosphosites were reported from the Modification Sites table after filtering for rows with "Phospho" in the Modification Name column. These searches were performed in a developmental version of Proteome Discoverer 3.1 to show that the phosphoproteomic depth achieved with our methods is not restricted to Spectronaut searches. However, this developmental version does not yet incorporate quantitative values for phosphorylation sites detected with DIA methods. Furthermore, from internal conversations, we know the results of this development build should be treated as preliminary and that the algorithms being utilized are still in active development. Consequently, we chose to perform our analyses using Spectronaut, which provides identification, localization, and quantification for DIA phosphoproteomics.

Localization error rate calculation

To calculate the localization error rate, a reference sheet of all the possible precursors that could be detected from the phosphopeptide standards was constructed with the phosphorylation sites specified. Precursors with sequences that could also arise from the yeast proteome (based on an *in-silico* digest) were removed from consideration. Then, the error rate was calculated as $\text{ErrorRate}(\%) = \frac{\text{False Sites}}{\text{False Sites} + \text{True Sites}} \times 100$, where "False Sites" is the number of phosphopeptide standard precursors detected with phosphorylation states not indicated in the reference sheet, and "True Sites" is the number of phosphopeptide standard precursors detected with phosphorylation states indicated in the reference sheet. This calculation was performed for each phosphopeptide standard raw data file, and the average error rates are reported as a function of the localization probability cutoff.

Phosphopeptide standard quantification calculations

The phosphopeptide standard reference sheet described above was modified so that each row corresponds to a phosphorylation site, so that the "PTM.Quantity" value in the PTM Site Report could be utilized for quantification assessment. As described above, phosphopeptides with sequences that could arise from the yeast proteome were removed from consideration. Linear regression and R^2 calculations were performed using the "pearsonr" and "linregress" functions within the "scipy.stats" module for phosphosites detected across at least three concentration points[76].

EGF-stimulated HeLa differential expression and pathway analysis

Differential expression analysis for the EGF treatment experiment was performed on all phosphosites that were detected across triplicates in either the EGF-treatment or control group. Missing value imputation was performed in Perseus[77] assuming a normal distribution with "width = 0.3" and "down shift = 1.8". Enriched phosphosites were defined as those with an absolute fold change of more than 2 and a p value less than 0.05 using a two-sided t-test without a multiple testing correction.

Pathway enrichment analysis was performed using Enrichr[78]. The enriched subset contains genes corresponding to phosphosites that are enriched as defined above, and the background was performed for the rest of genes with at least one detected phosphosite.

Sequence-based phosphosite clustering

Phosphosite sequences, along with their 5-amino acid flanking regions, were compared using the Blosum62 substitution matrix. The top 4 scores, representing a typical number of amino acids in kinase recognition motifs, were selected. These scores were subsequently averaged and transformed to a range between 0 (indicating identical sequences) and 1 (representing maximally different sequences). Subsequently, a similarity matrix between all sequence pairs was constructed and utilized to generate a 2D embedding space through t-SNE, implemented using the scikit-learn library[79]. Default parameters were used, except for metric = "precomputed", init = "random", n_iter=500, n_iter_without_progress=150, and random_state=42. Additionally, Perseus was employed for visualization purposes[77].

Kinase motif enrichment

Initially, each phosphorylation site and its flanking region in our dataset was used to search the kinase library website (https://kinase-library.phosphosite.org/site, accessed on February 22nd, 2024) for predictions of all 303 serine/threonine kinases for each site. For kinase motif enrichment analysis, only the top 15 kinases with regards to percentile score for each site were retained. To nominate tissue-enriched phosphorylation sites, we proceeded as follows: for each tissue-tissue comparison the log-transformed abundances of all shared phosphorylation sites were subtracted from each other. For selection of top phosphorylation sites per tissue two different approaches were applied: a z-score threshold of 2 was used, or the top and bottom 200 sites were selected. For the final analysis, the top and bottom 200 ranked sites were used. To calculate the frequency factor for each kinase the proportion of how often a kinase was predicted within the top phosphorylation sites versus the total of top phosphorylation sites was computed, divided by the same ratio for the unchanged sites, i.e., not ranked among the top/bottom 200. A chi-squared test was calculated after applying Haldane's correction to the contingency table of the two proportions and the p value was extracted.

Human variant information

Annotation of human variants in CPS1 and OPA1 were identified using The Human Gene Mutation Database at the Institute of Medical Genetics in Cardiff (HGMD) using the public site entries (retrieval date 2023-11-18[80]).

Multiple sequence alignment

Clustal Omega[81] hosted at the EMBL-EBI webserver was used for protein sequence alignment using the following UniProt identifiers for CPS1: P08686, F7EZ, P07756, Q8C196, F1ML89, A0A8C0NKH0; for OPA1: P58281, P58281-2, O60313, P32266; and other dynamin family members: Q811U4, Q80U63, Q8K1M6 (retrieved from UniProt 2023-11-18).

Structural modeling prediction

The PyMOL Molecular Graphics System, Version 2.5.7 Schrödinger, LLC. was used for predicting structural models for CPS1 (RCSB PDB: 5DOT (Apo), 5DOU (NAG)[66]) and OPA1 (RCSB PDB structure 6JTG[74]). For modeling purposes, the human crystal structures (highly homologous to mouse) were used.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All mass spectrometry data and data search outputs generated in this study have been deposited in the MassIVE database under accession code MSV000093613 [https://massive.ucsd.edu/ProteoSAFe/dataset.jsp?task=bda6396ce50547bdaac0deccdb110bd5]. A map of provided data files is included in Supplementary Data 5. Results from downstream data analyses are provided as Supplementary Data files. Source data for figures have been uploaded at the same MassIVE repository as above due to file size limitations. Source data are provided with this paper.

Code availability

All code used for visualization is available via the link: https://github.com/coongroup/PhosphoAstral.

References

1. Ardito, F., Giuliani, M., Perrone, D., Troiano, G. & Lo Muzio, L. The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy (Review).*Int. J. Mol. Med.* **40**, 271–280 (2017).
2. Solari, F. A., Dell'Aica, M., Sickmann, A. & Zahedi, R. P. Why phosphoproteomics is still a challenge. *Mol. Biosyst.* **11**, 1487–1493 (2015).
3. Neville, D. C. A. et al. Evidence for phosphorylation of serine 753 in CFTR using a novel metal-ion affinity resin and matrix-assisted laser desorption mass spectrometry. *Protein Sci.* **6**, 2436–2445 (1997).
4. Anguenot, R., Yelle, S. & Nguyen-Quoc, B. Purification of tomato sucrose synthase phosphorylated isoforms by Fe(III)-immobilized metal affinity chromatography. *Arch. Biochem. Biophys.* **365**, 163–169 (1999).

5. Posewitz, M. C. & Tempst, P. Immobilized gallium(III) affinity chromatography of phosphopeptides. *Anal. Chem.* **71**, 2883–2892 (1999).

6. Stensballe, A., Andersen, S. & Jensen, O. N. Characterization of phosphoproteins from electrophoretic gels by nanoscale Fe(III) affinity chromatography with off-line mass spectrometry analysis. *Proteomics* **1**, 207–222 (2001).

7. Ficarro, S. B. et al. Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae. *Nat. Biotechnol.* **20**, 301–305 (2002).

8. Pinkse, M. W. H., Uitto, P. M., Hilhorst, M. J., Ooms, B. & Heck, A. J. R. Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns. *Anal. Chem.* **76**, 3935–3943 (2004).

9. Sano, A. & Nakamura, H. Chemo-affinity of titania for the column-switching HPLC analysis of phosphopeptides. *Anal. Sci.* **20**, 565–566 (2004).

10. Aryal, U. K. & Ross, A. R. S. Enrichment and analysis of phosphopeptides under different experimental conditions using titanium dioxide affinity chromatography and mass spectrometry. *Rapid Commun. Mass Spectrom.* **24**, 219–231 (2010).

11. Olsen, J. V. et al. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* **127**, 635–648 (2006).

12. Zhang, G. & Neubert, T. A. Use of detergents to increase selectivity of immunoprecipitation of tyrosine phosphorylated peptides prior to identification by MALDI quadrupole-TOF MS. *Proteomics* **6**, 571–578 (2006).

13. Beltran, L. & Cutillas, P. R. Advances in phosphopeptide enrichment techniques for phosphoproteomics. *Amino Acids* **43**, 1009–1024 (2012).

14. Schroeder, M. J., Shabanowitz, J., Schwartz, J. C., Hunt, D. F. & Coon, J. J. A neutral loss activation method for improved phosphopeptide sequence analysis by quadrupole ion trap mass spectrometry. *Anal. Chem.* **76**, 3590–3598 (2004).

15. Syka, J. E. P., Coon, J. J., Schroeder, M. J., Shabanowitz, J. & Hunt, D. F. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl Acad. Sci. USA* **101**, 9528–9533 (2004).

16. Olsen, J. V. & Mann, M. Status of large-scale analysis of post-translational modifications by mass spectrometry*. *Mol. Cell. Proteom.* **12**, 3444–3452 (2013).

17. Nilsson, C. L. Advances in quantitative phosphoproteomics. *Anal. Chem.* **84**, 735–746 (2012).

18. Riley, N. M. & Coon, J. J. Phosphoproteomics in the age of rapid and deep proteome profiling. *Anal. Chem.* **88**, 74–94 (2016).

19. Grimsrud, P. A., Swaney, D. L., Wenger, C. D., Beauchene, N. A. & Coon, J. J. Phosphoproteomics for the masses. *ACS Chem. Biol.* **5**, 105–119 (2010).

20. Muehlbauer, L. K., Hebert, A. S., Westphall, M. S., Shishkova, E. & Coon, J. J. Global phosphoproteome analysis using high-field asymmetric waveform ion mobility spectrometry on a hybrid orbitrap mass spectrometer. *Anal. Chem.* **92**, 15959–15967 (2020).

21. Chi, A. et al. Analysis of phosphorylation sites on proteins from saccharomyces cerevisiae by electron transfer dissociation (ETD) mass spectrometry. *Proc. Natl Acad. Sci. USA* **104**, 2193–2198 (2007).

22. Swaney, D. L., Wenger, C. D., Thomson, J. A. & Coon, J. J. Human embryonic stem cell phosphoproteome revealed by electron transfer dissociation tandem mass spectrometry. *Proc. Natl Acad. Sci. USA* **106**, 995–1000 (2009).

23. Beausoleil, S. A., Villén, J., Gerber, S. A., Rush, J. & Gygi, S. P. A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat. Biotechnol.* **24**, 1285–1292 (2006).

24. Taus, T. et al. Universal and confident phosphorylation site localization using phosphoRS. *J. Proteome Res.* **10**, 5354–5362 (2011).

25. Cox, J. et al. Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **10**, 1794–1805 (2011).

26. Bailey, C. M. et al. SLoMo: automated site localization of modifications from ETD/ECD mass spectra. *J. Proteome Res.* **8**, 1965–1971 (2009).

27. Savitski, M. M. et al. Confident phosphorylation site localization using the mascot delta score. *Mol. Cell. Proteom.* **10**, S1–S12 (2011).

28. Molina, H., Horn, D. M., Tang, N., Mathivanan, S. & Pandey, A. Global proteomic profiling of phosphopeptides using electron transfer dissociation tandem mass spectrometry. *Proc. Natl Acad. Sci. USA* **104**, 2199–2204 (2007).

29. Ferries, S. et al. Evaluation of parameters for confident phosphorylation site localization using an orbitrap fusion tribrid mass spectrometer. *J. Proteome Res.* **16**, 3448–3459 (2017).

30. Sharma, K. et al. Ultradeep human phosphoproteome reveals a distinct regulatory nature of Tyr and Ser/Thr-based signaling. *Cell Rep.* **8**, 1583–1594 (2014).

31. Stewart, H. I. et al. Parallelized acquisition of orbitrap and astral analyzers enables high-throughput quantitative analysis. *Anal. Chem.* **95**, 15656–15664 (2023).

32. Stewart, H. et al. A conjoined rectilinear collision cell and pulsed extraction ion trap with auxiliary DC electrodes. *J. Am. Soc. Mass Spectrom.* https://doi.org/10.1021/jasms.3c00311 (2023).

33. Skowronek, P. et al. Rapid and in-depth coverage of the (phospho-)proteome with deep libraries and optimal window design for dia-PASEF. *Mol. Cell. Proteom.* **21**, 100279 (2022).

34. Jumper, J. et al. Highly accurate protein structure prediction with alphafold. *Nature* **596**, 583–589 (2021).

35. Bludau, I. et al. The structural context of posttranslational modifications at a proteome-wide scale. *PLoS Biol.* **20**, e3001636 (2022).

36. Tyanova, S., Cox, J., Olsen, J., Mann, M. & Frishman, D. Phosphorylation variation during the cell cycle scales with structural propensities of proteins. *PLoS Comput. Biol.* **9**, e1002842 (2013).

37. Iakoucheva, L. M. et al. The Importance of Intrinsic disorder for protein phosphorylation. *Nucleic Acids Res.* **32**, 1037–1049 (2004).

38. Johnson, J. L. et al. An atlas of substrate specificities for the human serine/threonine kinome. *Nature* **613**, 759–766 (2023).

39. Shishkova, E., Hebert, A. S., Westphall, M. S. & Coon, J. J. Ultra-high pressure (>30,000 Psi) packing of capillary columns enhancing depth of shotgun proteomic analyses. *Anal. Chem.* **90**, 11503–11508 (2018).

40. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).

41. Bekker-Jensen, D. B. et al. Rapid and site-specific deep phosphoproteome profiling by data-independent acquisition without the need for spectral libraries. *Nat. Commun.* **11**, 787 (2020).

42. Kitata, R. B. et al. A data-independent acquisition-based global phosphoproteomics system enables deep profiling. *Nat. Commun.* **12**, 2539 (2021).

43. Srinivasan, A., Sing, J. C., Gingras, A.-C. & Röst, H. L. Improving phosphoproteomics profiling using data-independent mass spectrometry. *J. Proteome Res.* **21**, 1789–1799 (2022).

44. Oliinyk, D. & Meier, F. Ion mobility-resolved phosphoproteomics with Dia-PASEF and short gradients. *Proteomics* **23**, 2200032 (2023).

45. Lou, R. et al. Benchmarking commonly used software suites and analysis workflows for DIA proteomics and phosphoproteomics. *Nat. Commun.* **14**, 94 (2023).

46. Bruderer, R. et al. Optimization of experimental parameters in data-independent mass spectrometry significantly increases depth and

reproducibility of results*. *Mol. Cell. Proteom.* **16**, 2296–2309 (2017).

47. Leijten, N. M., Heck, A. J. R. & Lemeer, S. Histidine phosphorylation in human cells; a needle or phantom in the haystack? *Nat. Methods* **19**, 827–828 (2022).

48. Vaca Jacome, A. S. et al. Avant-garde: an automated data-driven dia data curation tool. *Nat. Methods* **17**, 1237–1244 (2020).

49. Huttlin, E. L. et al. A tissue-specific atlas of mouse protein phosphorylation and expression. *Cell* **143**, 1174–1189 (2010).

50. Stopfer, L. E. et al. High-density, targeted monitoring of tyrosine phosphorylation reveals activated signaling networks in human tumors. *Cancer Res.* **81**, 2495–2509 (2021).

51. Giansanti, P. et al. Mass spectrometry-based draft of the mouse proteome. *Nat. Methods* **19**, 803–811 (2022).

52. Hornbeck, P. V. et al. PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res.* **43**, D512–520 (2015).

53. Pinna, L. A. & Ruzzene, M. How do protein kinases recognize their substrates? *Biochim Biophys. Acta* **1314**, 191–225 (1996).

54. Alderson, T. R., Pritišanac, I., Kolarić, Đ., Moses, A. M. & Forman-Kay, J. D. Systematic identification of conditionally folded intrinsically disordered regions by AlphaFold2. *Proc. Natl Acad. Sci. USA* **120**, e2304302120 (2023).

55. Chico, L. K., Van Eldik, L. J. & Watterson, D. M. Targeting protein kinases in central nervous system disorders. *Nat. Rev. Drug Discov.* **8**, 892–909 (2009).

56. Küry, S. et al. De novo mutations in protein kinase genes CAMK2A and CAMK2B cause intellectual disability. *Am. J. Hum. Genet.* **101**, 768–788 (2017).

57. Kool, M. J. et al. CAMK2-dependent signaling in neurons is essential for survival. *J. Neurosci.* **39**, 5424–5439 (2019).

58. Rigter, P. M. F. et al. Role of CAMK2D in neurodevelopment and associated conditions. *Am. J. Hum. Genet.* **111**, 364–382 (2024).

59. Lordén, G. et al. Enhanced activity of alzheimer disease-associated variant of protein kinase Cα drives cognitive decline in a mouse model. *Nat. Commun.* **13**, 7200 (2022).

60. Callender, J. A. & Newton, A. C. Conventional protein kinase C in the brain: 40 years later. *Neuronal Signal.* **1**, NS20160005 (2017).

61. Maichele, A. J., Burwinkel, B., Maire, I., Søvik, O. & Kilimann, M. W. Mutations in the testis/liver isoform of the phosphorylase kinase gamma subunit (PHKG2) cause autosomal liver glycogenosis in the Gsd rat and in humans. *Nat. Genet.* **14**, 337–340 (1996).

62. Niemi, N. M. & Pagliarini, D. J. The extensive and functionally uncharacterized mitochondrial phosphoproteome. *J. Biol. Chem.* **297**, 100880 (2021).

63. Arnold, P. K. & Finley, L. W. S. Regulation and function of the mammalian tricarboxylic acid cycle. *J. Biol. Chem.* **299**, 102838 (2023).

64. Häberle, J. et al. Suggested guidelines for the diagnosis and management of urea cycle disorders: first revision. *J. Inherit. Metab. Dis.* **42**, 1192–1230 (2019).

65. Häberle, J. Urea Cycle Disorders Consortium; Lichtarge, O.; Rubio, V.; Wong, L.-J.; Summar, M. et al. Molecular defects in human carbamoy phosphate synthetase I: mutational spectrum, diagnostic and protein structure considerations. *Hum. Mutat.* **32**, 579–589 (2011).

66. de Cima, S. et al. Structure of human carbamoyl phosphate synthetase: deciphering the on/off switch of human ureagenesis. *Sci. Rep.* **5**, 16950 (2015).

67. Pekkala, S. et al. Understanding carbamoyl-phosphate synthetase I (CPS1) deficiency by using expression studies and structure-based analysis. *Hum. Mutat.* **31**, 801–808 (2010).

68. Takakusa, H. et al. Protein tyrosine nitration of mitochondrial carbamoyl phosphate synthetase 1 and its functional consequences. *Biochem Biophys. Res. Commun.* **420**, 54–60 (2012).

69. Del Dotto, V., Fogazza, M., Carelli, V., Rugolo, M. & Zanna, C. Eight human OPA1 isoforms, long and short: what are they for? *Biochim Biophys. Acta Bioenerg.* **1859**, 263–269 (2018).

70. Ferré, M. et al. Improved locus-specific database for OPA1 mutations allows inclusion of advanced clinical data. *Hum. Mutat.* **36**, 20–25 (2015).

71. Chappie, J. S., Acharya, S., Leonard, M., Schmid, S. L. & Dyda, F. G. Domain dimerization controls dynamin's assembly-stimulated GTPase activity. *Nature* **465**, 435–440 (2010).

72. Santarelli, R. et al. OPA1-related auditory neuropathy: site of lesion and outcome of cochlear implantation. *Brain* **138**, 563–576 (2015).

73. Loo, J. L. et al. Multiethnic involvement in autosomal-dominant optic atrophy in Singapore. *Eye* **31**, 475–480 (2017).

74. Yu, C. et al. Structural insights into G domain dimerization and pathogenic mutation of OPA1. *J. Cell Biol.* **219**, e201907098 (2020).

75. Del Dotto, V. et al. Deciphering OPA1 mutations pathogenicity by combined analysis of human, mouse and yeast cell models. *Biochim Biophys. Acta Mol. Basis Dis.* **1864**, 3496–3514 (2018).

76. Virtanen, P. et al. SciPy 1.0 contributors. SciPy 1.0: fundamental algorithms for scientific computing in python. *Nat. Methods* **17**, 261–272 (2020).

77. Tyanova, S. et al. The perseus computational platform for comprehensive analysis of (Prote)omics data. *Nat. Methods* **13**, 731–740 (2016).

78. Chen, E. Y. et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinform.* **14**, 128 (2013).

79. Pedregosa, F. et al. Scikit-learn: machine learning in python. *JMLR* **12**, 2825–2830 (2011).

80. Stenson, P. D. et al. The human gene mutation database (HGMD®): optimizing its use in a clinical diagnostic or research setting. *Hum. Genet.* **139**, 1197–1207 (2020).

81. Madeira, F. et al. Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res.* **50**, W276–W279 (2022).

82. Rath, S. et al. MitoCarta3.0: an updated mitochondrial proteome now with sub-organelle localization and pathway annotations. *Nucleic Acids Res.* **49**, D1541–D1547 (2021).

Acknowledgements

Author contributions

N.A. and J.F. performed cell culture and collected HEK293T and HeLa cells. J.H. and A.G. collected mouse tissues. J.R. and A.P.G. performed yeast culture and collected yeast cells. T.N.A., A.P., H.S., C.H., E.D., and V.Z. developed the new instrument used in this study. N.M.L., T.M.P.C., T.N.A., A.P., M.L.R., L.R.S., and E.D. performed mass spectrometry sample preparation and/or data acquisition. N.M.L., P.S., T.M.P.C., and L.R.S. performed mass spectrometry data processing. N.M.L., P.S., P.F.,

C.F., and A.J.S. performed data analysis and assessment of the biological relevance of the study. N.M.L., P.S., P.F., C.F., A.J.S., E.S., T.N.A., L.R.S., M.S.W., E.D., V.S., and J.J.C. contributed to the figure content and design. N.M.L., P.S., P.F., C.F., A.J.S., E.S., H.S., D.J.P., and J.C.C. wrote the manuscript. N.M.L., P.S., P.F., T.M.P.C., C.F., A.J.S., E.S., T.N.A., A.P., H.S., D.J.P., V.Z., and J.J.C. edited the manuscript. M.S.W., D.J.P., V.Z., and J.J.C. provided project supervision.

Competing interests

J.J.C. is a consultant for Thermo Fisher Scientific and on the scientific advisory board for Seer and 908 Devices. T.N.A., A.P., H.S., C.H., E.D., and V.S. are employees of Thermo Fisher Scientific. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at
https://doi.org/10.1038/s41467-024-51274-0.

Correspondence and requests for materials should be addressed to Joshua J. Coon.

Peer review information *Nature Communications* thanks Nurhan Ozlu, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

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

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.