

# Extensive protein pyrophosphorylation revealed in human cell lines

Received: 29 June 2023

Accepted: 27 March 2024

Published online: 25 April 2024

Check for updates

Jeremy A. M. Morgan <sup>1,5</sup>, Arpita Singh <sup>2,3,5</sup>, Leonie Kurz <sup>1,4</sup>,  
Michal Nadler-Holly<sup>1</sup>, Max Ruwolt <sup>1</sup>, Shubhra Ganguli <sup>2</sup>, Sheenam Sharma<sup>2,3</sup>,  
Martin Penkert<sup>1</sup>, Eberhard Krause<sup>1</sup>, Fan Liu<sup>1</sup>, Rashna Bhandari <sup>2</sup>   
& Dorothea Fiedler <sup>1,4</sup>

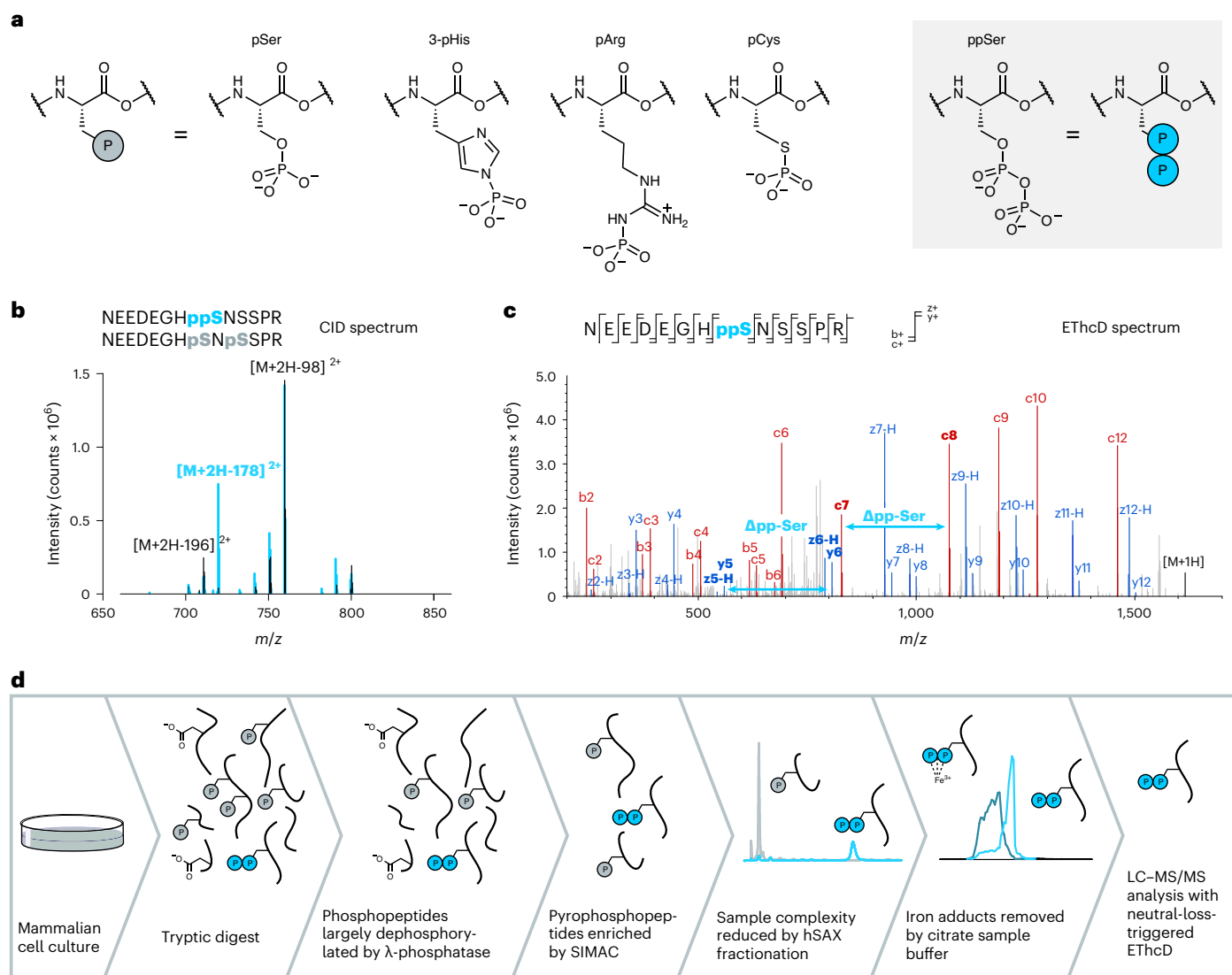
Reversible protein phosphorylation is a central signaling mechanism in eukaryotes. Although mass-spectrometry-based phosphoproteomics has become routine, identification of non-canonical phosphorylation has remained a challenge. Here we report a tailored workflow to detect and reliably assign protein pyrophosphorylation in two human cell lines, providing, to our knowledge, the first direct evidence of endogenous protein pyrophosphorylation. We manually validated 148 pyrophosphosites across 71 human proteins, the most heavily pyrophosphorylated of which were the nucleolar proteins NOLC1 and TCOF1. Detection was consistent with previous biochemical evidence relating the installation of the modification to inositol pyrophosphates (PP-InsPs). When the biosynthesis of PP-InsPs was perturbed, proteins expressed in this background exhibited no signs of pyrophosphorylation. Disruption of PP-InsP biosynthesis also significantly reduced rDNA transcription, potentially by lowering pyrophosphorylation on regulatory proteins NOLC1, TCOF1 and UBF1. Overall, protein pyrophosphorylation emerges as an archetype of non-canonical phosphorylation and should be considered in future phosphoproteomic analyses.

The specific phosphorylation of proteins is a fundamental mechanism of intracellular signal transduction across the domains of life<sup>1,2</sup>. In humans, kinases and phosphatases dedicated to the writing and erasing of protein phosphorylation make up almost 2.5% of the genome<sup>3–5</sup>. After the discovery of serine (Fig. 1a), threonine and tyrosine phosphorylation through biochemical approaches, mass spectrometry (MS)-based proteomics became the primary method to investigate the function and regulation of canonical Ser/Thr/Tyr phosphorylation<sup>6</sup>. As of 2024, more than 290,000 phosphorylation sites have been reported in the PhosphoSitePlus database, identified almost entirely by phosphoproteomic approaches<sup>7</sup>.

In contrast, phosphorylation of non-canonical amino acid residues in humans (histidine, arginine, cysteine, aspartate, glutamate and lysine; Fig. 1a) has been established with varying success. Adapting

traditional bottom-up phosphoproteomics workflows to allow for high-throughput detection of these modifications has proven challenging, as the low pH conditions used in sample preparation typically led to the hydrolysis of the acid-labile phosphoramidate, thiophosphate and acylphosphate moieties<sup>8,9</sup>. Rapid acidic enrichment employed at low temperatures has, however, proven successful for phosphohistidine and phosphoarginine detection in bacterial backgrounds<sup>10,11</sup>. There is also an additional burden on the spectral interpretation, as convincing assignment of a novel non-canonical site requires excluding a misassigned canonical site. To improve assignment accuracy, methods have incorporated characteristic neutral loss patterns<sup>12</sup> and immonium ion formation<sup>10</sup> into the detection and assignment process. In human backgrounds, identification of phosphohistidine (pHis) is arguably the most advanced, with 14 sites biochemically validated<sup>13</sup> and several

<sup>1</sup>Leibniz-Forschungsinstitut für Molekulare Pharmakologie (FMP), Berlin, Germany. <sup>2</sup>Laboratory of Cell Signalling, Centre for DNA Fingerprinting and Diagnostics, Hyderabad, India. <sup>3</sup>Graduate Studies, Regional Centre for Biotechnology, Faridabad, India. <sup>4</sup>Institute of Chemistry, Humboldt-Universität zu Berlin, Berlin, Germany. <sup>5</sup>These authors contributed equally: Jeremy A. M. Morgan, Arpita Singh. e-mail: [rashna@cdfd.org.in](mailto:rashna@cdfd.org.in); [fiedler@fmp-berlin.de](mailto:fiedler@fmp-berlin.de)



**Fig. 1 | Enrichment and detection of pyrophosphorylated peptides using mass spectrometry. a**, Examples of protein phosphorylation on different amino acid side chains. Canonical phosphorylation sites include phosphoserine (pSer), whereas non-canonical sites are exemplified by phosphohistidine (3-pHis), phosphoarginine (pArg), phosphocysteine (pCys) and pyrophosphoserine (ppSer). **b**, Fragmentation via CID of a pyrophosphopeptide results in

a characteristic neutral loss ( $-178$  Da) compared to the corresponding bisphosphopeptide. **c**, Fragmentation of a pyrophosphopeptide using electron transfer dissociation combined with EThcD shows excellent sequence coverage while leaving the modification intact. **d**, Development of a sample preparation workflow, tailored to the enrichment and subsequent detection of tryptic pyrophosphopeptides.

hundred sites identified across different phosphoproteomics studies. However, there are still considerable concerns regarding the validity of these high-throughput site assignments<sup>14</sup>.

Protein pyrophosphorylation, the phosphorylation of a phosphoserine (pSer) residue to yield pyrophosphoserine (ppSer; Fig. 1a), is an additional non-canonical phosphorylation that is often overlooked<sup>15–17</sup>. This non-enzymatic posttranslational modification is mediated by high-energy inositol pyrophosphate messengers (PP-InsPs), which can transfer their  $\beta$ -phosphoryl group to protein substrates in the presence of  $Mg^{2+}$  ions<sup>16</sup>. Pyrophosphorylation was established using radiolabeled 5-diphosphoinositol pentakisphosphate (SPP-InsP<sub>5</sub>), putatively the most abundant PP-InsP, demonstrating that a peptide or protein substrate can accept the radiolabel at only a pre-phosphorylated residue<sup>16</sup>. ppSer exhibits differences in stability compared to pSer, notably a resistance to hydrolysis by common protein phosphatases and  $\lambda$ -phosphatase.

The *Saccharomyces cerevisiae* proteins Nsr1, Srp40 and YGR130C were the first eukaryotic proteins shown to undergo

pyrophosphorylation in vitro<sup>15</sup>. Subsequently, a handful of mammalian targets of in vitro protein pyrophosphorylation were identified, including nucleolar and coiled-body phosphoprotein 1 (NOLC1), Treacher Collins syndrome protein 1 (TCOF1), adaptor protein complex AP-3 subunit beta-1 (AP3B1), cytoplasmic dynein 1 intermediate chain 2 (DCIL2) and the oncoprotein MYC<sup>16,18–20</sup>. Indirect evidence of endogenous pyrophosphorylation has relied on a ‘back-phosphorylation’ assay, where potential targets are expressed and purified from a PP-InsP-rich cell line. In such a cell line, endogenous pyrophosphorylation should be elevated, and the subsequent in vitro phosphoryl transfer to this target should be decreased, when compared to the protein obtained from a cell line with low PP-InsP levels<sup>21</sup>. Although essential to discovery, these tools are low throughput and cannot provide direct information on the sites of modification.

Detection of endogenous pyrophosphorylation by mass spectrometry (MS)-based proteomics would address these limitations. Compared to other modes of non-canonical phosphorylation, pyrophosphorylation is relatively acid stable<sup>15</sup>, suggesting that traditional

phosphoproteomics enrichment techniques should be compatible. However, the major technical challenge in enriching and detecting pyrophosphopeptides is their differentiation from peptides monophosphorylated at multiple positions. In particular, bisphosphorylated peptides (peptides containing two monophosphate groups) are isobaric with pyrophosphopeptides, making the observation of the molecular ion uninformative. An unambiguous assignment requires an inversion of the traditional probability-based assignment; the peptide must be assumed to be bisphosphorylated, unless pyrophosphorylation can be asserted.

Here we report the development of a dedicated pyrophosphoproteomics workflow for the detection and unambiguous assignment of endogenous pyrophosphorylation sites from mammalian cell lysates. Using neutral-loss-triggered electron transfer dissociation combined with higher-energy collision dissociation (ETHcD) liquid chromatography–tandem mass spectrometry (LC–MS/MS) analysis, 108 and 78 sites were identified from HEK293T and HCT116 cell lines, respectively. Protein pyrophosphorylation sites predominantly occurred in acidic serine-rich stretches, and most of the identified pyrophosphoproteins localized to the nucleus and nucleolus. Several proteins with newly identified pyrophosphorylation sites also accepted radiolabeled phosphate from [ $\beta^{32}\text{P}$ ]5PP-InsP<sub>3</sub> *in vitro*, supporting the non-enzymatic mechanism and the validity of both detection methods. In a functional readout for pyrophosphorylation of nucleolar proteins, we observed significantly impaired rDNA transcription in 5PP-InsP<sub>3</sub>-depleted cells. In sum, protein pyrophosphorylation can now be added unequivocally to the growing list of endogenous phosphorylation motifs in human cell lines.

## Results

### Establishment of a pyrophosphoproteomics workflow

Using synthetic peptide standards, we previously established specific mass spectrometric detection of pyrophosphopeptides based on a characteristic neutral loss of  $-178\text{ m/z}$ , corresponding to the loss of pyrophosphoric acid ( $\text{H}_4\text{P}_2\text{O}_7$ ) during collision-induced dissociation (CID) fragmentation (Fig. 1b)<sup>22</sup>. This neutral loss did not occur in Ser/Thr/Tyr monophosphorylated or bisphosphorylated peptides. Detection of this neutral loss from CID fragmentation of the proteome could be used as a trigger during shotgun proteomic analysis to identify candidate pyrophosphopeptide parent ions, which then undergo selective ETHcD fragmentation. The ETHcD spectra provided excellent sequence coverage while the modification stayed intact, enabling assignment of pyrophosphorylation sites. After implementing further improvements, including an optimized neutral loss filter, fine-tuned fragmentation parameters and the exclusion of low-charge precursor ions, this method was again applied to model pyrophosphopeptides (Fig. 1c and Extended Data Fig. 1a) and bisphosphopeptides (Extended Data Fig. 1b). Consistent neutral loss detection and reproducible site assignment indicated that this triggered MS approach was suitable for the identification of pyrophosphorylation sites in cell lysates.

In first attempts, samples were prepared using a standard high pH fraction and immobilized metal ion affinity chromatography (IMAC) enrichment approach for phosphoproteomics<sup>23</sup> and were subsequently analyzed using the triggered MS method. Unfortunately, although the spiked-in pyrophosphopeptide standards were reliably detected, only one endogenous site was identified. Therefore, a dedicated sample preparation workflow for enrichment of pyrophosphorylated tryptic peptides was developed (Fig. 1d). We tailored a standard phosphoproteomics workflow<sup>24</sup> toward pyrophosphopeptide selection utilizing a set of synthetic pyrophosphopeptides (and the corresponding phosphopeptides) of varying sequence characteristics for optimization (Extended Data Fig. 2). Proteomic material was generated from HCT116 or HEK293T cells using standard protocols for tryptic digestion. To reduce competition with phosphopeptides during subsequent

enrichment, the digested material was then treated with  $\lambda$ -phosphatase. As previously reported, this phosphatase hydrolyzes a large proportion of monophosphopeptides while leaving the pyrophosphoryl groups intact (Extended Data Fig. 2a)<sup>16,25</sup>. A standard sequential elution from immobilized metal ion affinity chromatography (SIMAC) enrichment, featuring an additional low pH washing step designed to elute acidic peptides, was then performed<sup>26</sup>. Pyrophosphopeptides were largely retained during SIMAC (Extended Data Fig. 2b), and the overall material mass was reduced more than 40-fold. This retention is likely supported by the lower predicted pK<sub>a</sub> value of the pyrophosphoryl moiety, in comparison to phosphoryl groups or acidic amino acid side chains.

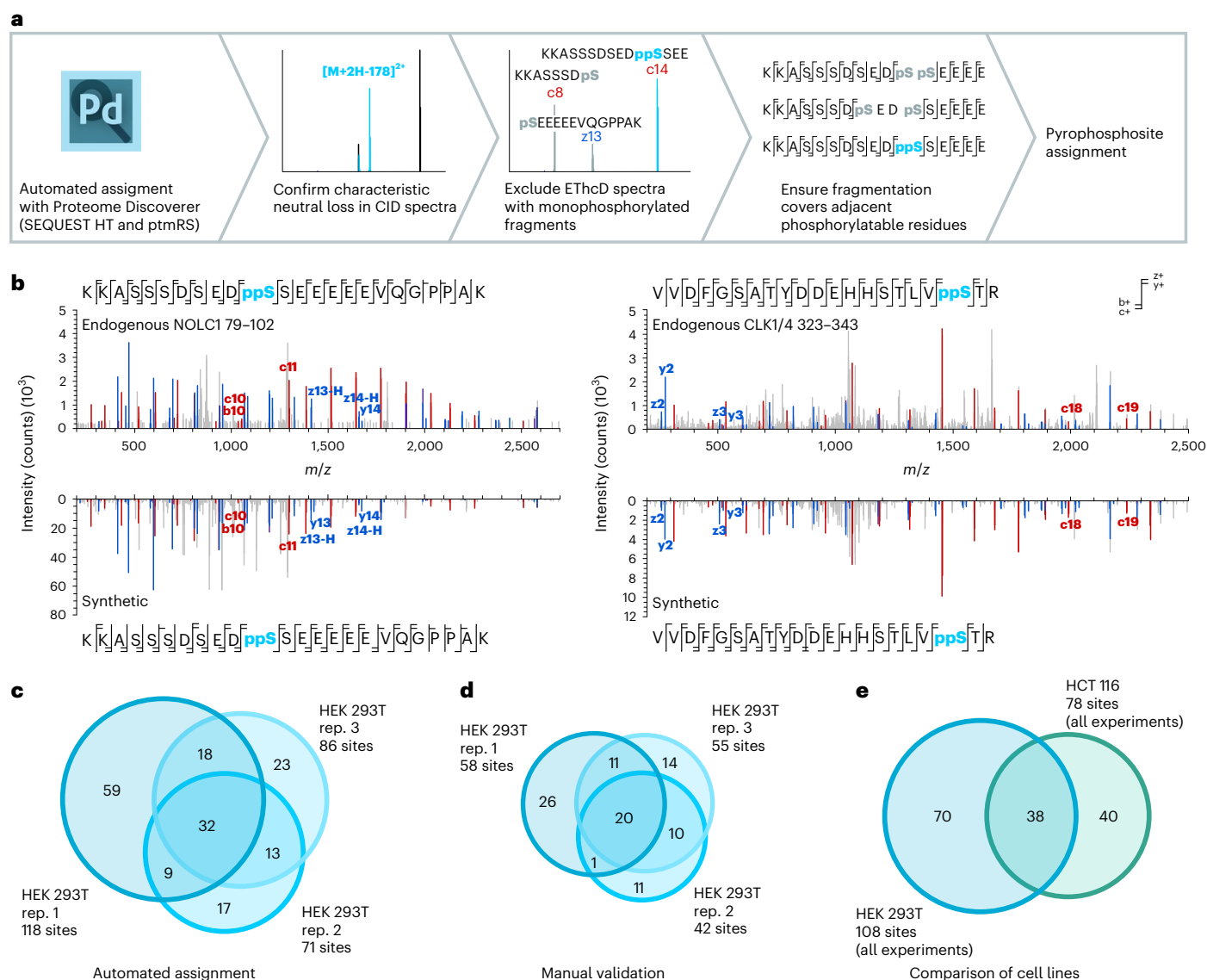
Despite the phosphatase treatment and SIMAC enrichment, peptides with polyacidic amino acid stretches and multiple monophosphorylation sites were still abundant, so an offline fractionation step was implemented to further reduce sample complexity (Supplementary Fig. 1). An ultra-performance liquid chromatography (UPLC) hydrophilic SAX (hSAX)<sup>27</sup> column using a quaternary ammonium stationary phase on a hydrophilic polymeric support was selected, owing to orthogonality with the low pH reverse-phase chromatography used in the LC–MS separation and the ability to separate analytes of differing negative charge and polarity<sup>6</sup>.

While establishing the workflow, we observed that pyrophosphorylated standard peptides were often detected in complex with Fe<sup>3+</sup> ions during LC–MS analysis. Crucially, these adducts formed in the liquid phase, as evidenced by distinct retention times and peak shapes. Similar behavior was previously reported for highly phosphorylated peptides and could be resolved through the addition of metal chelating agents<sup>28,29</sup>. Therefore, sodium citrate (50 mM) was added into the sample resuspension buffer, which led to a substantial decrease of Fe<sup>3+</sup> adduct formation (Extended Data Fig. 2c–f)<sup>28,29</sup>. Overall, the sample preparation workflow (Fig. 1d) now seemed adequate for the enrichment of pyrophosphopeptides from complex samples.

### Reliable annotation of endogenous pyrophosphorylation sites

We next subjected the widely used mammalian cell line HEK293T to the pyrophosphoproteomics workflow. Using CID neutral-loss-triggered ETHcD MS, many putative pyrophosphorylation sites were detected. To avoid incorrect assignment of multiply phosphorylated peptides (particularly bisphosphorylated peptides) as pyrophosphorylated, careful analysis of the data was required (Fig. 2a). Initial annotation was made on the basis of the SEQUEST HT engine with a fixed value peptide spectrum match (PSM) validator and ptmRS assignment as nodes in a Proteome Discoverer workflow. In the resulting dataset, some spectra could be directly assigned as pyrophosphopeptides based on unambiguous fragmentation, but many spectra were annotated both as pyrophosphorylated and as bisphosphorylated peptides with similar certainty. Increasing the threshold for the *P* value during automated assignment did not alleviate this problem. It became apparent that co-elution (and co-fragmentation) of bisphosphorylated peptides with pyrophosphopeptides could produce ambiguous mixed spectra: if two bisphosphopeptides with overlapping phosphorylation pairs are co-fragmented, all fragments required to annotate a pyrophosphorylation site to that central residue are present (Extended Data Fig. 3). This means that the interpretation of each spectrum must exclude the possibility of bisphosphopeptide mixtures before an assignment can be made. Therefore, a three-step protocol to assess automatically assigned pyrophosphorylation sites was established (Fig. 2a).

During MS analysis, each precursor ion is first fragmented by CID. In the synthetic pyrophosphopeptide standards, the  $-178\text{-m/z}$  ( $\text{H}_4\text{P}_2\text{O}_7$ ) neutral loss peak was consistently among the three most intense signals in the CID spectra, the other two being  $-98\text{ m/z}$  ( $\text{H}_3\text{PO}_4$ ) and  $-196\text{ m/z}$  ( $\text{H}_6\text{P}_2\text{O}_8$ ). The HEK293T proteomic data showed that parent ions



**Fig. 2 | Assignment and validation of endogenous pyrophosphorylation sites.**

**a**, Workflow for the assignment of pyrophosphorylation sites after automated assignment with Proteome Discoverer. **b**, Comparison of EThcD spectra obtained from complex samples (top) with synthetically prepared pyrophosphopeptides (bottom). Fragment ions critical for the assignment of pyrophosphorylation sites are indicated in red (b-ion and c-ion series) or blue (y-ion and z-ion series). See

detailed assignment in Supplementary Fig. 3. **c**, Venn diagram of the number of pyrophosphorylation sites detected by Proteome Discoverer in  $n = 3$  biological replicates prepared from HEK293T cell lysates. **d**, Venn diagram of the number of pyrophosphorylation sites detected after manual assignment using  $n = 3$  biological replicates from **c**. **e**, Overlap of pyrophosphorylation sites detected in HEK293T and HCT116 cells.

exhibiting weak  $-178$ - $m/z$  neutral loss peaks during CID fragmentation were frequently assigned as ambiguous based on EThcD fragmentation. This is consistent with bisphosphopeptide co-fragmentation, because bisphosphopeptides can produce only  $-98$ - $m/z$  and  $-196$ - $m/z$  neutral losses, making these signals proportionally higher. Therefore, in the first assessment step, parent ions exhibiting a  $-178$ - $m/z$  neutral loss with an intensity outside the three most intense signals were discarded, as they were likely bisphosphopeptide/pyrophosphopeptide mixtures (Supplementary Fig. 2).

In the second step, EThcD fragmentation spectra of candidate ions triggered by the detection of the neutral loss peak during CID were then manually examined using the Molecular Weight Calculator (<https://github.com/PNNL-Comp-Mass-Spec/Molecular-Weight-Calculator-VB6>)<sup>30</sup> for evidence of peptide fragments containing a single phosphorylated residue—this is possible only if the peptide is bisphosphorylated. Again, candidate ions exhibiting such fragments were discarded (Fig. 2a and Supplementary Fig. 2).

Finally, in the third step, the extent of fragmentation across the putative pyrophosphorylation site was assessed. Missing fragments, particularly those encompassing a canonically phosphorylatable residue, such as serine or threonine, can lead to the misassignment of a bisphosphorylated peptide as pyrophosphorylated, and, as such, spectra with key fragments missing were discarded (Supplementary Fig. 2). The remaining spectra correspond to genuine pyrophosphorylation sites.

To validate the pyrophosphosite assignment, we synthesized pyrophosphopeptides based on two detected sequences: NOLC1 79–102 and CLK1/4 323–343 (Fig. 2b). Both peptide sequences contain phosphorylatable residues directly adjacent to the putative pyrophosphorylation sites. To prove that pyrophosphorylation is present, sequential ions consistent with the unphosphorylated peptide fragment immediately preceding the putative site (for example, ppNOLC1 79–88) and the sequential pyrophosphorylated peptide fragment containing the putative site (for example, ppNOLC1 79–89) must be detected simultaneously (in the absence of a singly phosphorylated

peptide fragment indicative of bisphosphorylation). For the NOLC1 sequence, the c10 and c11 ion couplet and the z13-H and z14-H ion couplet in the c/z ions series were observed in the fragmentation spectra of both the synthetic peptide and the endogenous peptide, confirming the presence of the putative pyrophosphorylation site. Similarly, the CLK1/4 sequence exhibited the z2/z3 and c18/c19 ion couplets in the c/z ions series and the y2/y3 ion couplet in the b/y ion series, in both synthetic and endogenous peptide fragmentation spectra, consistent with pyrophosphorylation at Ser341. No fragments indicative of monophosphorylation were detected in any spectra, and, crucially, the diagnostic ion couplets were absent in the fragmentation spectra of the corresponding bisphosphopeptide (Extended Data Fig. 1). Together, these data validated our assignment approach for the reliable and correct identification of endogenous pyrophosphorylation sites.

After applying the pyrophosphoproteomics workflow and assignment strategy to HEK293T cell lysates, three biological replicates were analyzed. A total of 171 unique pyrophosphorylation sites were detected by automated assignment, 93 of which were manually validated (an average of 51 per replicate; Fig. 2c,d and Supplementary Table 1). Forty pyrophosphoproteins were identified from manually validated sites. Although a core of 20 manually validated sites was seen in all replicates, a significant portion of the sites (51) were exclusively observed in a single replicate. This is likely due to variation in the sample background of each replicate, differentially co-eluting and masking these low abundant species. Three proteins exhibiting multiple pyrophosphorylation sites were heavily overrepresented in the triplicate. NOLC1, TCOF1 and serine/arginine repetitive matrix protein 1 (SRRM1) sites were found in all replicates and represented 43 of the 93 total sites assigned. Overall, compared to the HEK293T proteome, the identified pyrophosphoproteins were expressed at intermediate to high levels, illustrating the challenge of sufficiently enriching pyrophosphopeptides (Extended Data Fig. 4).

After establishing the detectability of pyrophosphorylation, we reanalyzed the replicate 1 HEK293T dataset using a decoy searching approach with the target decoy PSM validator node in Proteome Discoverer. Of the 58 manually assigned sites, 19 were correctly assigned, and a further 34 sites discarded during manual evaluation were identified (Supplementary Table 2). Many of these additional sites were plausible but did not fulfil the criteria for unambiguous assignment, highlighting the utility of an automated decoy search in a preliminary assessment of proteome pyrophosphorylation.

During method development, lysates from HEK293T and another human cell line, HCT116 colon cancer cells, were also frequently subjected to pyrophosphoproteomic analysis, and a total of 78 sites on 33 proteins were identified in HCT116 (Supplementary Table 3). Many of these pyrophosphorylation sites overlap with sites from HEK293T lysates (Fig. 2e), further corroborating the reliability of the assignment strategy.

### Pyrophosphorylation is commonly found on nucleolar proteins

Two nucleolar proteins, NOLC1 and TCOF1, were heavily pyrophosphorylated, containing 34 and 18 different pyrophosphorylation sites, respectively. These observations are consistent with previous biochemical studies, in which both NOLC1 (also called Nopp140) and TCOF1 were able to undergo in vitro radiolabeling by [ $\beta^{32}\text{P}$ ]5PP-InsP<sub>5</sub> (refs. 15,16). Both NOLC1 and TCOF1 are densely phosphorylated proteins, which presumably facilitates their pyrophosphorylation. Interestingly, the pyrophosphorylation sites on NOLC1 and TCOF1 exclusively localize to acidic regions, whereas phosphorylation sites are reported to be more evenly distributed across acidic and basic regions (Fig. 3a). The localization of pyrophosphorylation sites to acidic serine stretches was observed in previous studies on 5PP-InsP<sub>5</sub>-mediated pyrophosphorylation<sup>16,18–20,31,32</sup>. In all cases, the priming in vitro phosphorylation was catalyzed by acidophilic Ser/Thr kinases, especially casein kinase 2 (CK2).

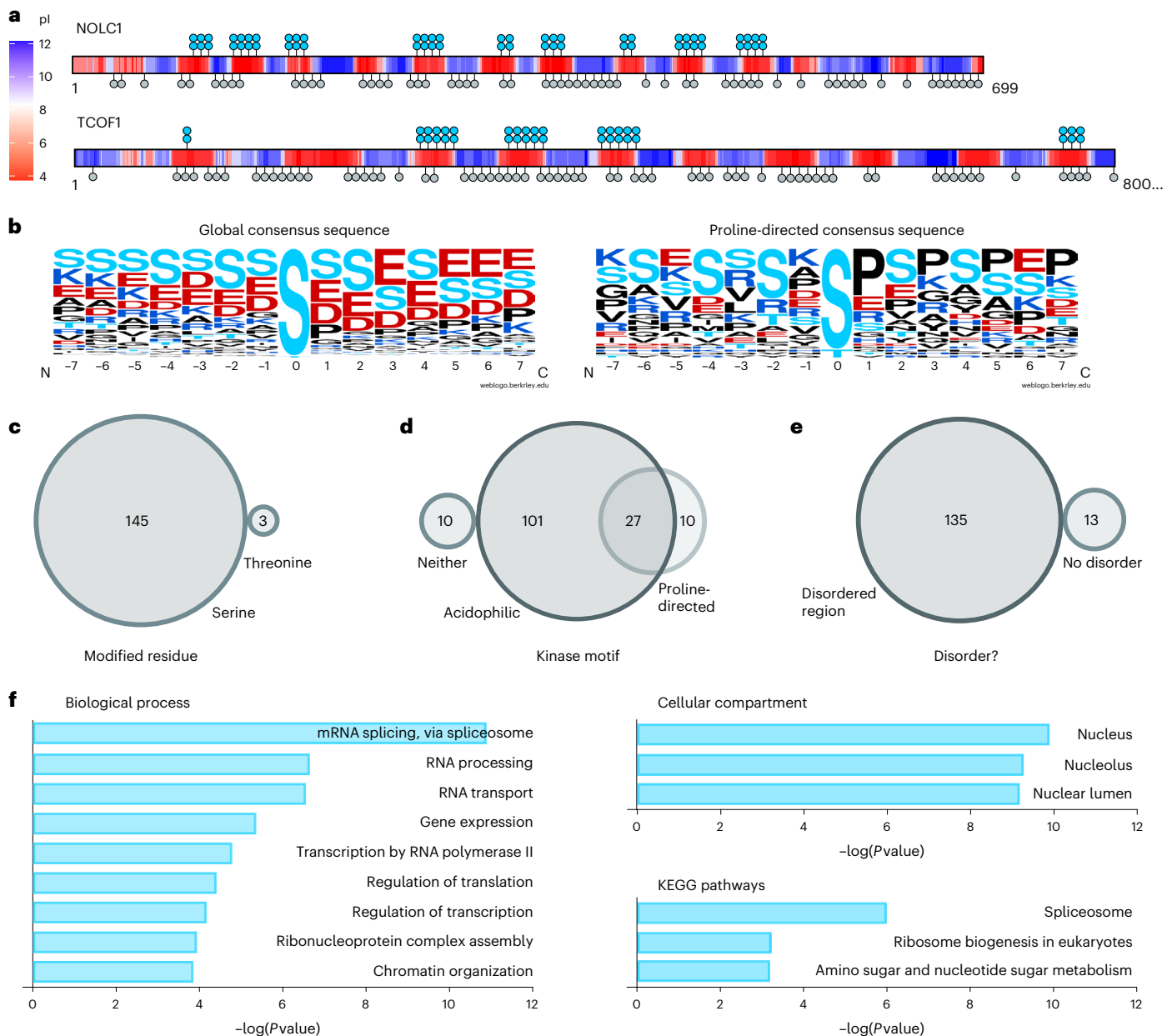
A global analysis of all endogenous pyrophosphorylation sites detected by MS in HEK293T and HCT116 lysates confirmed the central role for acidophilic Ser/Thr kinases—alignment of all sequences revealed a clear CK2 consensus sequence (Fig. 3b)<sup>33</sup>. However, some pyrophosphorylation sequences did not match the CK2 consensus sequence. When we removed all sites containing a Glu/Asp/Ser/Thr residue in position +3 from the alignment, a proline-directed kinase consensus sequence emerged (Fig. 3b)<sup>34</sup>, suggesting that this family of Ser/Thr kinases may also pre-phosphorylate residues prior to their pyrophosphorylation.

Of the 148 pyrophosphorylated sites identified in our study, only three occur on threonine residues, and the remainder are on serine (Fig. 3c). We analyzed the pyrophosphosites using the Scansite 4.0 tool to predict motifs that are likely to undergo phosphorylation by specific protein kinases (<https://scansite4.mit.edu/#scanProtein>)<sup>34</sup>. Consistent with the consensus sequences above, most of the sites were predicted to be phosphorylated by acidophilic Ser/Thr kinases; a smaller number were potential substrates for proline-directed Ser/Thr kinases; and a few sites may be substrates for both families of kinases (Fig. 3d). Approximately 7% of the mapped pyrophosphorylated residues were not predicted as substrates for either acidophilic or proline-directed Ser/Thr kinases, suggesting that other families of protein kinases may also prime residues for pyrophosphorylation. An additional feature common to pyrophosphorylation sites is that they lie within intrinsically disordered regions (IDRs)<sup>17</sup>. A disorder prediction analysis using IUPred2A (<https://iupred2a.elte.hu/>; Supplementary Table 3) showed that 91% of the pyrophosphosites identified in our study lie within a continuous stretch of 20 or more residues that have a disorder score  $\geq 0.5$ , which is the cutoff for predicted disorder at that residue (Fig. 3e). This is in line with properties of phosphorylation sites in general, as IDRs are overrepresented in eukaryotic phosphoproteomic datasets<sup>35,36</sup>.

Gene Ontology analysis of pyrophosphorylation sites suggests a function for this modification in nuclear and nucleolar processes, as these two compartments were significantly overrepresented among pyrophosphorylated proteins (Fig. 3f and Supplementary Table 4). This localization preference is mirrored by the biological processes, in which RNA processing (specifically RNA splicing) emerged as a process putatively regulated by protein pyrophosphorylation of SRRM1, SRRM2, serine/arginine-rich splicing factor 2 (SRSF2), SRSF5, SRSF6, SRSF9, splicing factor 3B subunit 2 (SF3B2), WW domain-binding protein 11 (WBP11) and transformer-2 protein homolog beta (TRA2B). Another biological process potentially regulated by pyrophosphorylated proteins is chromatin organization, which is modulated by epigenetic regulators, including high mobility group protein A1 (HMGA1) and the histone deacetylase HDAC2, among others. Previous studies showed that inositol pyrophosphates can influence epigenetic modifications that regulate chromatin remodelling in yeast and mammals<sup>37</sup>. Proteins NOLC1, nucleolar protein 58 (NOP58), nucleophosmin (NPM1), H/ACA ribonucleoprotein complex subunit DKC1 (DKC1) and U3 small nucleolar ribonucleoprotein protein MPP10 (MPHOSPH10) are involved in ribosome biogenesis, another process overrepresented among pyrophosphoproteins (Fig. 3f). A functional connection between PP-InsPs and ribosome biogenesis was previously made in *S. cerevisiae* using genetics, and a mechanistic hypothesis for this regulation involved pyrophosphorylation of the RNA polymerase I subunits A190, A43 and A34.5 (ref. 32).

### Pyrophosphoproteins in the nucleolar fibrillar center

The nucleolus, which emerged as a major site for localization of pyrophosphorylated proteins, is a membrane-less organelle organized into three liquid–liquid phase-separated subcompartments: the fibrillar center (FC), the dense fibrillar component (DFC) and the granular component (GC) (Fig. 4a)<sup>38</sup>. rDNA repeats present in the FC are transcribed by RNA polymerase I at the FC/DFC boundary; the resulting pre-rRNAs are processed in the DFC and assembled into ribosomes in the GC<sup>38</sup>. Human IP6 kinase isoforms IP6K1 and IP6K2 are reported to be localized to the

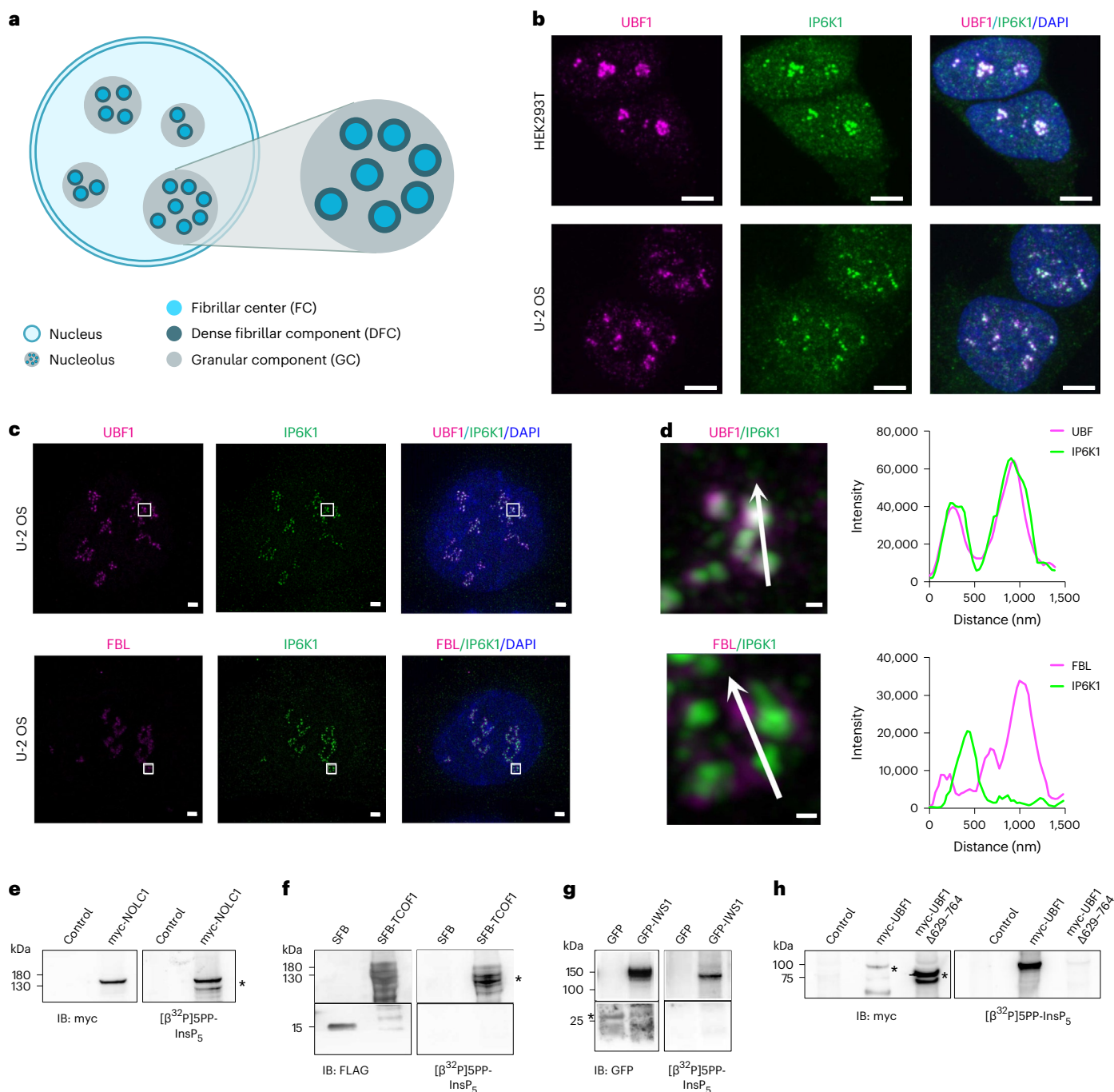


**Fig. 3 | Properties of pyrophosphorylation sites.** **a**, Illustration of the localization of pyrophosphorylation sites to acidic regions in NOLC1 and TCOF1. The local isoelectric point ( $pI$ ) was calculated using a custom R script. ppSer residues are shown in turquoise, and previously reported phosphorylation sites are indicated in gray (PhosphoSitePlus (ref. 7)). **b**, Consensus sequence of all pyrophosphorylation sites detected (left) and consensus sequence after

removal of CK2 consensus sites (right). Sequence logos were generated using WebLogo (ref. 65). **c**, Side chain of modification for all pyrophosphorylation sites. **d**, Kinase motifs surrounding the pyrophosphorylation sites. **e**, Analysis of order/disorder around pyrophosphorylation sites. **f**, Gene Ontology analysis of pyrophosphoproteins using Enrichr (Supplementary Table 4; <https://maayanlab.cloud/Enrichr/>).  $P$  values were obtained using Fisher's exact test.

nucleolar FC region (Human Protein Atlas, <https://www.proteinatlas.org/>)<sup>39</sup>. We used immunofluorescence to confirm the co-localization of IP6K1 with the FC marker protein upstream binding factor 1 (UBF1)<sup>40</sup> in HEK293T cells (Fig. 4b). The same pattern of localization was observed in the osteosarcoma cell line U-2 OS (Fig. 4b). Super-resolution microscopy revealed that IP6K1 is confined to the FC and does not co-localize with fibrillarin (FBL), which marks the DFC<sup>41</sup> (Fig. 4c,d). Local synthesis of 5PP-InsP<sub>3</sub> by IP6Ks in the FC would facilitate pyrophosphorylation of FC resident proteins. Indeed, the two most highly pyrophosphorylated proteins identified in our study, NOLC1 and TCOF1, are localized to the FC<sup>42</sup>. Four additional pyrophosphoproteins—NOP58; DKC1; SWI/SNF-related, matrix-associated, actin-dependent regulator of chromatin, subfamily A, member 4 (SMARCA4); and eukaryotic translation elongation factor

1 delta (EEF1D)—are also annotated to the FC<sup>42</sup>. UBF1, which co-localizes with IP6K1 in the FC (Fig. 4b–d), has a region of disorder at the C-terminus, which contains serine residues interspersed with glutamate and aspartate residues in a sequence that is reminiscent of the pyrophosphosites on TCOF1, NOLC1 and the nuclear protein IWS1 (Supplementary Fig. 4). UBF1 was, however, not identified in our pyrophosphoproteome MS screen, and so we resorted to radiolabeled [ $\beta$ -<sup>32</sup>P]5PP-InsP<sub>3</sub>, which has been used as a classical tool to test for pyrophosphorylation of candidate proteins<sup>21,43</sup>. We confirmed that NOLC1, TCOF1 and IWS1 overexpressed and isolated from HEK293T cells can undergo in vitro pyrophosphorylation with [ $\beta$ -<sup>32</sup>P]5PP-InsP<sub>3</sub> after pre-phosphorylation by CK2 (Fig. 4e–g). Similarly, we observed robust pyrophosphorylation of CK2 pre-phosphorylated UBF1 upon exposure to [ $\beta$ -<sup>32</sup>P]5PP-InsP<sub>3</sub>



#### Fig. 4 | Nucleolar FC proteins undergo SPP-InsP<sub>5</sub>-mediated

**pyrophosphorylation.** **a**, Architecture of the mammalian nucleolus and its subcompartments. **b**, Confocal micrographs representative of three independent experiments showing co-localization of IP6K1 (green) with nucleolar FC marker UBF1 (magenta) in HEK293T and U-2 OS cells. Nuclei were stained with DAPI (blue); scale bars, 5  $\mu$ m. **c**, SIM images representative of three independent experiments showing IP6K1 (green) co-localized with UBF1 or the nucleolar DFC marker FBL (magenta) in U-2 OS cells. Nuclei were stained with DAPI (blue); scale bars, 2  $\mu$ m. **d**, Magnification of the boxed region (c); scale bars, 0.2  $\mu$ m. Traces show fluorescence intensity profiles for IP6K1 (green)

and UBF1 or FBL (magenta), measured along the indicated white arrow. **e–h**, In vitro pyrophosphorylation of human proteins by [<sup>32</sup>P]SPP-InsP<sub>5</sub>. NOLC1, TCOF1, IWS1 and UBF1 (full-length and C-terminally deleted versions), with their indicated N-terminal tags, were expressed in HEK293T cells, isolated and prephosphorylated by CK2 before incubation with [<sup>32</sup>P]SPP-InsP<sub>5</sub>. Representative images show autoradiography to detect pyrophosphorylation (right) and immunoblotting with respective tag-specific antibodies (left) ( $n = 4$  (**e**),  $n = 5$  (**f**),  $n = 4$  (**g**) and  $n = 2$  (**h**)). Negative controls were cells transfected with plasmids pCMV-myc (**e**, **h**), pCDNA-SFB (**f**) or EGFP-C1 (**g**). The asterisks in **e–h** indicate specific bands. IB, immunoblot.

(Fig. 4h). We predicted pyrophosphorylation sites on UBF1 by identifying regions of disorder that contain sites for pre-phosphorylation by acidophilic Ser/Thr kinases (Supplementary Fig. 4). Deletion of the C-terminal disordered region of UBF1, which possesses a large number of potential pyrophosphosites, abrogated in vitro pyrophosphorylation

of the protein, confirming that UBF1 undergoes pyrophosphorylation at its C-terminus (Fig. 4h). Although UBF1 is a relatively abundant protein (Extended Data Fig. 4), the predicted pyrophosphosites in the C-terminus of UBF1 lie within either very short or very long tryptic fragments, both of which would evade detection by MS.

## PP-InsPs support pyrophosphorylation and rDNA transcription

Given the complementarity of the radiolabeling and MS methods, we used both approaches to affirm that 5PP-InsP<sub>5</sub> drives intracellular protein pyrophosphorylation. For this, we relied on *IP6K1*<sup>-/-</sup> HEK293T cells expressing either active or kinase-dead IP6K1, which display a three-fold difference in the levels of 5PP-InsP<sub>5</sub> (Extended Data Fig. 5a–c). TCOF1, NOLC1 or IWS1 expressed in these cells (Fig. 5a)<sup>44</sup> were purified, treated with λ-phosphatase and resolved by SDS-PAGE, and the excised gel pieces were digested with trypsin and measured by neutral-loss-triggered EThcD MS as described above. Equal loading between conditions was confirmed by analyzing absolute abundance of target proteins using the Precursor Ions Quantifier node in Proteome Discoverer (Extended Data Fig. 5d). Many automatic pyrophosphorylation site assignments in the samples were obtained from cells expressing active IP6K1, several of which could be confirmed to correspond to pyrophosphorylated peptides for all three substrates (Fig. 5a and Supplementary Table 5). By contrast, we could not confirm pyrophosphorylation on a single peptide sequence triggered by the characteristic loss of the pyrophosphoryl moiety from the cells expressing kinase-dead IP6K1. These results illustrate the strong dependence of endogenous pyrophosphorylation on cellular PP-InsP levels and point to 5PP-InsP<sub>5</sub> as the predominant phosphoryl donor. To interrogate the involvement of 1,5(PP)<sub>2</sub>-InsP<sub>4</sub> in protein pyrophosphorylation, we performed pyrophosphoproteomics on a lysate from *PP1P5K*<sup>-/-</sup> HCT116 cells, which cannot produce 1,5(PP)<sub>2</sub>-InsP<sub>4</sub> (ref. 45). Similar levels of pyrophosphorylation were observed in the *PP1P5K*<sup>-/-</sup> cells compared to the wild-type (WT) cells, suggesting that 1,5(PP)<sub>2</sub>-InsP<sub>4</sub> is not a major mediator of protein pyrophosphorylation in this cell line (Extended Data Fig. 6 and Supplementary Table 6). However, caution must be taken when interpreting these results, as *PP1P5K*<sup>-/-</sup> cell lines produce increased levels of 5PP-InsP<sub>5</sub> (refs. 46,47), which may compensate for a loss of 1,5(PP)<sub>2</sub>-InsP<sub>4</sub>.

As pyrophosphosites on UBF1 evade detection by MS, we resorted to the ‘back-pyrophosphorylation’ method to examine intracellular UBF1 pyrophosphorylation<sup>21</sup> (Fig. 5b). Overexpressed or endogenous UBF1 was isolated from *IP6K1*<sup>-/-</sup> HEK293T cells expressing either active or kinase-dead IP6K1 and incubated with radiolabeled [<sup>32</sup>P]5PP-InsP<sub>5</sub>. UBF1 isolated from cells expressing kinase-dead IP6K1 showed a two-fold higher pyrophosphorylation signal on the autoradiogram compared to UBF1 from cells with active IP6K1, reflecting higher levels of intracellular pyrophosphorylation on the latter form of UBF1 (Fig. 5c,d). The increase in cellular pyrophosphorylation on TCOF1, NOLC1, IWS1 and UBF1 in the presence of active IP6K1 is consistent with intracellular pyrophosphorylation mediated by 5PP-InsP<sub>5</sub>.

TCOF1, NOLC1 and UBF1 are known regulators of RNA polymerase I-mediated rDNA transcription<sup>38,39,48,49</sup>. Reduced intracellular 5PP-InsP<sub>5</sub>, which, in turn, will lower pyrophosphorylation levels of these proteins, is therefore likely to have an impact on rRNA synthesis. To examine this possibility, we used two different cellular models known to be depleted for 5PP-InsP<sub>5</sub>: (1) HCT116 cells lacking IP6K1 and IP6K2 (ref. 50) and (2) cells treated with the InsP<sub>6</sub> kinase inhibitors TNP or SC-919 (refs. 51,52). Quantitative RT-PCR analysis was used to assess the levels of the 45S pre-rRNA transcript, which is subsequently processed to yield mature rRNA for incorporation into ribosomes. We observed a more than two-fold decrease in the levels of 45S pre-rRNA in *IP6K1*<sup>-/-</sup> *IP6K2*<sup>-/-</sup> double knockout (DKO) cells compared to WT HCT116 cells (Fig. 5e). Treatment with IP6K inhibitors recapitulated these findings and similarly suppressed pre-rRNA synthesis in HCT116 and U-2 OS cells, compared to cells treated with the vehicle control (Fig. 5f,g). These observations point to an important role for 5PP-InsP<sub>5</sub> in the maintenance of rDNA transcription, likely via pyrophosphorylation of key nucleolar proteins.

## Discussion

In this work, we developed a tailored pyrophosphoproteomics workflow to detect and assign protein pyrophosphorylation in two human

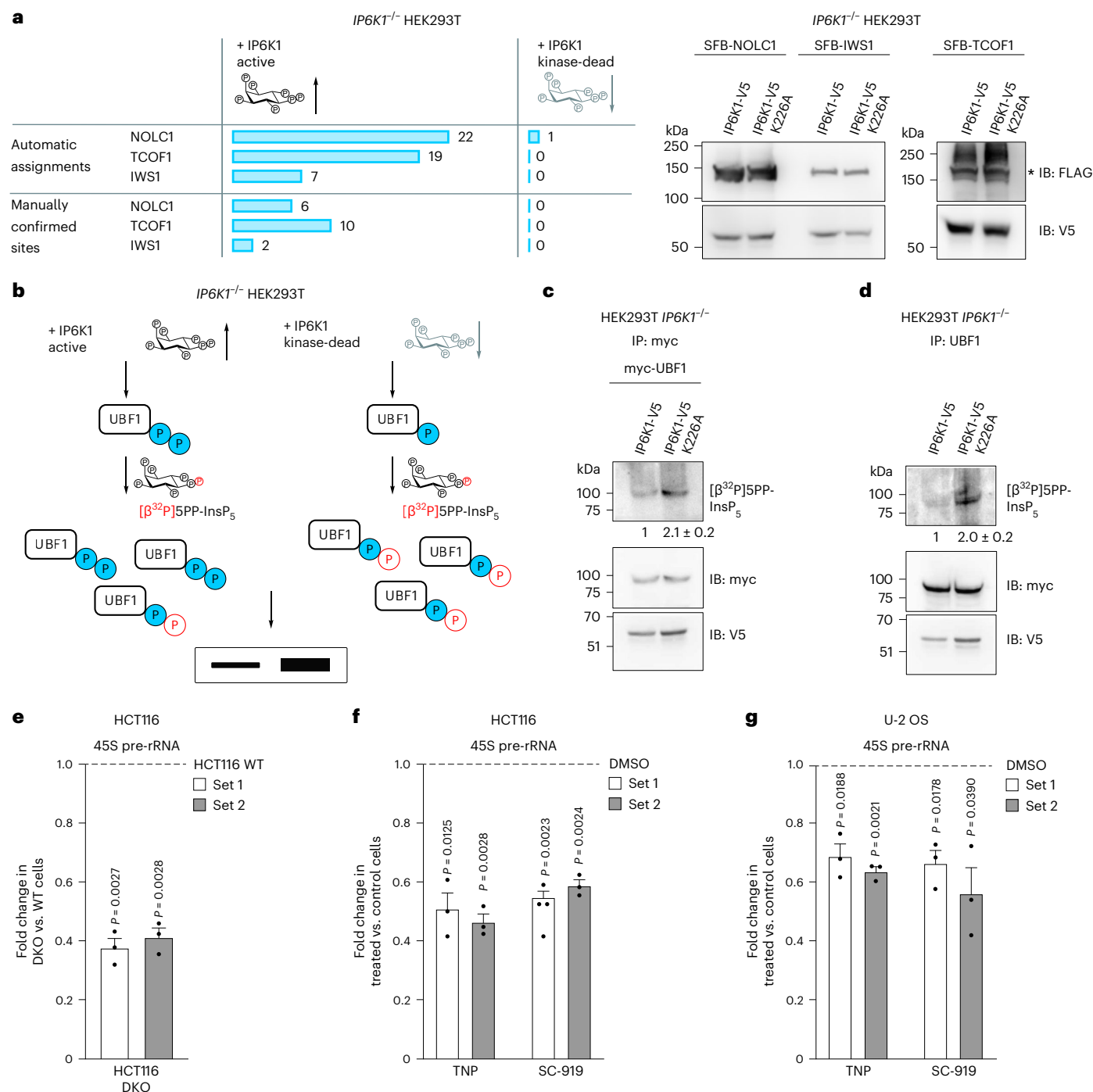
cell lines (148 manually validated sites across 71 proteins), providing, to our knowledge, the first direct evidence of endogenous protein pyrophosphorylation. Over time, efforts to characterize non-canonical phosphorylation with phosphoproteomic methods have advanced, as sample handling, enrichment and MS techniques have been adapted to the properties of the modification. For example, pHis proteomics has evolved by leveraging antibody enrichment and, later, strong anion exchange, culminating in a dedicated pHis database, HisPhosSite, containing more than 270 putative human sites<sup>53</sup>. Despite these advances, site localization and, therefore, distinction from canonical phosphorylation has remained a major technical challenge. As site localization probabilities are tightened in assignment workflows, unambiguous assignments are markedly reduced<sup>8</sup>. The pHis ammonium ion was recently used as additional evidence of correct pHis assignment, but only 0.5% of putative pHis-containing peptides in a human dataset exhibited this ion, suggesting significant overassignment by current automated methods<sup>14</sup>.

In the case of pyrophosphorylation, with our tailored sample preparation workflow and manual assignment strategy, we could achieve unambiguous pyrophosphosite assignment. The proteomic datasets indicated that pyrophosphorylation often occurred in acidic stretches known to be multiply phosphorylated, meaning that mixtures of isobaric bisphosphopeptide isomers and pyrophosphopeptides may be co-fragmented during EThcD, increasing the risk of false assignment. To exclude this possibility, both the CID and EThcD fragmentation patterns were assessed, and spectra containing monophosphorylated fragments were excluded from the analysis. Although this manual workflow is more laborious than automated methods, the assigned pyrophosphorylation sites are unambiguous and can be directly used to inform further biological investigation. We expect that it is possible to integrate the data analysis into software-based assignment, avoiding manual assessment of sites in the future. Even with the current automated assignment alone, an indicative number of pyrophosphorylation sites are identified, sufficient for preliminary investigations.

The dependence of pyrophosphorylation on PP-InsPs was demonstrated based on the lack of characteristic neutral loss triggers and assigned pyrophosphosites originating from NOLC1, TCOF1 and IWS1 proteins expressed in a PP-InsP-depleted (IP6K1-kinase dead) cell line. All three proteins were also able to accept the radiolabeled phosphoryl group from [<sup>32</sup>P]5PP-InsP<sub>5</sub>, demonstrating that the pyrophosphorylation of these proteins is consistent with the non-enzymatic model of protein pyrophosphorylation.

In the future, a quantitative MS approach likely involving tandem mass tag (TMT) labeling or stable isotope labeling by amino acids in cell culture (SILAC) will be needed to quantify levels of protein pyrophosphorylation in different cellular backgrounds or treatment conditions. This approach would also be required to investigate site occupancy; as peptides are deliberately dephosphorylated during sample preparation, comparing amounts of pyrophosphopeptides and their phosphorylated or unmodified precursors in a single analysis is not possible. Further optimization to reduce the quantity of material required per analysis and to improve the level of overlap between biological replicates would make the method more accessible. Both issues likely relate to inefficiencies in the λ-phosphatase treatment and enrichment steps, causing non-pyrophosphorylated material to remain after sample preparation and stochastically mask pyrophosphopeptide ions during MS analysis. Implementation of online Fe-IMAC chromatography<sup>54</sup> might allow for improved selectivity, reducing material requirements and increasing reproducibility. Optimizing the λ-phosphatase reaction would further reduce the selectivity requirement of the enrichment step. Additionally, the lack of Lys/Arg residues for tryptic cleavage in acidic polyserine stretches may prevent detection of certain pyrophosphopeptides with our approach. Pyrophosphorylation of residues adjacent to the cleavage site may also prevent protein digestion.





**Fig. 5 | 5PP-InsP<sub>5</sub> drives cellular pyrophosphorylation and promotes rRNA synthesis.** **a**, Comparative MS analysis of pyrophosphosites detected in SFB-tagged NOLC1, TCOF1 or IWS1 co-expressed with V5-tagged active or kinase-dead IP6K1 (left). Number of pyrophosphopeptides assigned by automatic or manual analysis. An additional analysis, including the number of ETcD triggers, can be found in Supplementary Fig. 5. Immunoblot (IB) shows that equal amounts of protein were subjected to MS (right). Asterisk indicates the specific band ( $n = 1$ ). **b**, Back-pyrophosphorylation method to detect intracellular UBF1 pyrophosphorylation. **c,d**, Overexpressed myc-tagged UBF1 (**c**) or endogenous UBF1 (**d**) was immunoprecipitated from *IP6K1*<sup>-/-</sup> HEK293T cells expressing either active or kinase-dead V5-tagged IP6K1 and incubated with [<sup>32</sup>P]5PP-InsP<sub>5</sub>. Images

show autoradiography to detect pyrophosphorylation (top) and immunoblotting with the indicated antibodies (bottom). IB with V5 antibody detects input levels of IP6K1. Numbers show mean fold change  $\pm$  s.e.m. in the extent of UBF1 back-pyrophosphorylation in cells expressing kinase-dead compared to active IP6K1 ( $n = 3$  (**c**) and  $n = 4$  (**d**) independent experiments). **e-g**, RT-qPCR analysis to measure 45S pre-rRNA transcript levels using two different primer sets. Values indicate the fold change in transcript levels in *IP6K1*<sup>-/-</sup> *IP6K2*<sup>-/-</sup> DKO cells compared to HCT116 WT cells (**e**) and TNP-treated or SC-919-treated HCT116 (**f**) or U-2 OS (**g**) cells compared to cells treated with the vehicle control (DMSO). Data are mean  $\pm$  s.e.m. ( $n = 3$  independent experiments). *P* values were determined using a two-tailed one-sample *t*-test. IP, immunoprecipitation. IB, immunoblotting.

In the future, these possibilities could be addressed by integrating additional proteases, such as GluC, into the digestion step alongside trypsin<sup>55</sup>.

NOLC1 and TCOF1 were the two most heavily pyrophosphorylated proteins identified in this study. Along with UBF1, here characterized as a novel in vitro pyrophosphorylation substrate, NOLC1 and TCOF1 are

known to upregulate rRNA synthesis via RNA polymerase I binding<sup>56–59</sup>. CK2-dependent phosphorylation at the C-terminal region of UBF1 promotes its interaction with SL1 to form a stable RNA polymerase I pre-initiation complex<sup>59,60</sup>. As most of the predicted UBF1 pyrophosphorylation sites lie within its C-terminus (Supplementary Fig. 4), pyrophosphorylation on UBF1 after CK2 priming may conceivably regulate rRNA synthesis. Depletion of 5PP-InsP<sub>3</sub> leads to decreased rDNA transcription (Fig. 5e–g), consistent with regulation via polymerase I. We previously showed that, in budding yeast *S. cerevisiae*, the absence of 5PP-InsP<sub>3</sub> leads to severely reduced rRNA synthesis, correlating with pyrophosphorylation of RNA polymerase I subunits<sup>32</sup>. A recent study showed that elevation of intracellular 5PP-InsP<sub>3</sub> in *PIP5K*<sup>−/−</sup> HCT116 cells does not alter the steady state levels of rRNA, but this work did not examine pre-rRNA transcript levels that reflect rDNA transcription by polymerase I (ref. 61). 5PP-InsP<sub>3</sub> is thought to be a ‘metabolic messenger’ as its synthesis by IP6Ks is uniquely sensitive to ATP availability<sup>62,63</sup>. If the interaction between NOLC1/TCOF1/UBF1 and RNA polymerase I were indeed dependent on 5PP-InsP<sub>3</sub>-mediated pyrophosphorylation, this would represent a straightforward cellular energy sensing mechanism to control rDNA transcription.

The localization of pyrophosphoproteins to the nucleolus was striking and raises the question of whether pyrophosphorylation plays a general regulatory role in this biomolecular condensate. It was recently demonstrated that phosphorylation of specific sites influenced partitioning of NPML1 (pSer125) and heterogeneous nuclear ribonucleoprotein A1 (HNRNPA1) (pSer6) to the nucleolus<sup>64</sup>. Interestingly, both proteins were found to be pyrophosphorylated at these specific positions in our MS data. It is, therefore, valuable to develop tools that can accurately represent or mimic pyrophosphorylation at the protein level. Such tools would enable researchers to investigate the influence of pyrophosphorylation on protein partitioning into membrane-less condensates.

Overall, protein pyrophosphorylation has emerged as an abundant, non-canonical phosphorylation. The ability to detect this modification within complex samples using MS and to definitively assign the modification sites now opens the door to investigate the functional role of protein pyrophosphorylation at the biochemical and cellular level. Although the installation of pyrophosphorylation appears to be non-enzymatic in biochemical assays, the question remains as to whether the presence of cell lysates or other co-factors can accelerate protein pyrophosphorylation. How the features of the PP-InsP phosphoryl donor influence the degree and the specificity of pyrophosphorylation has not been addressed to date. After installation, pyrophosphorylation could be detected by specific reader domains, and characterization of the pyrophosphoproteome will facilitate the identification of such readers. Finally, the removal of pyrophosphorylation sites will need to be explored. Are there dedicated protein pyrophosphatases that convert pyrophosphoserine back to phosphoserine or serine? Answers to these questions will provide fundamental insight into the regulation of protein pyrophosphorylation and its interplay with signaling pathways controlled by canonical protein phosphorylation.

## Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41589-024-01613-5>.

## References

- Cohen, P. The origins of protein phosphorylation. *Nat. Cell Biol.* **4**, E127–E130 (2002).
- Humphrey, S. J., James, D. E. & Mann, M. Protein phosphorylation: a major switch mechanism for metabolic regulation. *Trends Endocrinol. Metab.* **26**, 676–687 (2015).
- Hunter, T. A journey from phosphotyrosine to phosphohistidine and beyond. *Mol. Cell* **82**, 2190–2200 (2022).
- Manning, G., Whyte, D. B., Martinez, R., Hunter, T. & Sudarsanam, S. The protein kinase complement of the human genome. *Science* **298**, 1912–1934 (2002).
- Alonso, A. et al. Protein tyrosine phosphatases in the human genome. *Cell* **117**, 699–711 (2004).
- Beausoleil, S. A. et al. Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc. Natl Acad. Sci. USA* **101**, 12130–12135 (2004).
- Hornbeck, P. V. et al. PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res.* **43**, D512–D520 (2015).
- Hardman, G. et al. Strong anion exchange-mediated phosphoproteomics reveals extensive human non-canonical phosphorylation. *EMBO J.* **38**, 1–22 (2019).
- Bertran-Vicente, J. et al. Site-specifically phosphorylated lysine peptides. *J. Am. Chem. Soc.* **136**, 13622–13628 (2014).
- Potel, C. M., Lin, M. H., Heck, A. J. R. & Lemeer, S. Widespread bacterial protein histidine phosphorylation revealed by mass spectrometry-based proteomics. *Nat. Methods* **15**, 187–190 (2018).
- Prust, N., Van Breugel, P. C. & Lemeer, S. Widespread arginine phosphorylation in *Staphylococcus aureus*. *Mol. Cell. Proteom.* **21**, 100232 (2022).
- Oslund, R. C. et al. A phosphohistidine proteomics strategy based on elucidation of a unique gas-phase phosphopeptide fragmentation mechanism. *J. Am. Chem. Soc.* **136**, 12899–12911 (2014).
- Fuhs, S. R. & Hunter, T. pHosphorylation: the emergence of histidine phosphorylation as a reversible regulatory modification. *Curr. Opin. Cell Biol.* **45**, 8–16 (2017).
- 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).
- Saiardi, A., Bhandari, R., Resnick, A. C., Snowman, A. M. & Snyder, S. H. Phosphorylation of proteins by inositol pyrophosphates. *Science* **306**, 2101–2105 (2004).
- Bhandari, R. et al. Protein pyrophosphorylation by inositol pyrophosphates is a posttranslational event. *Proc. Natl Acad. Sci. USA* **104**, 15305–15310 (2007).
- Ganguli, S. et al. A high energy phosphate jump—from pyrophospho-inositol to pyrophospho-serine. *Adv. Biol. Regul.* **75**, 100662 (2020).
- Azevedo, C., Burton, A., Ruiz-Mateos, E., Marsh, M. & Saiardi, A. Inositol pyrophosphate mediated pyrophosphorylation of AP3B1 regulates HIV-1 Gag release. *Proc. Natl Acad. Sci. USA* **106**, 21161–21166 (2009).
- Chandhuri, M. et al. Inositol hexakisphosphate kinase 1 (IP6K1) activity is required for cytoplasmic dynein-driven transport. *Biochem. J.* **473**, 3031–3047 (2016).
- Lolla, P., Shah, A., Unnikannan, C. P., Oddi, V. & Bhandari, R. Inositol pyrophosphates promote MYC polyubiquitination by FBW7 to regulate cell survival. *Biochem. J.* **478**, 1647–1661 (2021).
- Chandhuri, M. & Bhandari, R. Back-pyrophosphorylation assay to detect in vivo InsP<sub>7</sub>-dependent protein pyrophosphorylation in mammalian cells. In *Inositol Phosphates: Methods and Protocols* (ed Miller, G. J.) 93–105 (Springer, 2020).
- Penkert, M. et al. Unambiguous identification of serine and threonine pyrophosphorylation using neutral-loss-triggered electron-transfer/higher-energy collision dissociation. *Anal. Chem.* **89**, 3672–3680 (2017).
- Batth, T. S., Francavilla, C. & Olsen, J. V. Off-line high-pH reversed-phase fractionation for in-depth phosphoproteomics. *J. Proteome Res.* **13**, 6176–6186 (2014).
- Villén, J. & Gygi, S. P. The SCX-IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. *Nat. Protoc.* **3**, 1630–1638 (2008).

25. Yates, L. M. & Fiedler, D. Establishing the stability and reversibility of protein pyrophosphorylation with synthetic peptides. *ChemBioChem* **16**, 415–423 (2015).
26. Thingholm, T. E., Jensen, O. N., Robinson, P. J. & Larsen, M. R. SIMAC (sequential elution from IMAC), a phosphoproteomics strategy for the rapid separation of monophosphorylated from multiply phosphorylated peptides. *Mol. Cell. Proteom.* **7**, 661–671 (2008).
27. Ritorto, M. S., Cook, K., Tyagi, K., Pedrioli, P. G. A. & Trost, M. Hydrophilic strong anion exchange (hSAX) chromatography for highly orthogonal peptide separation of complex proteomes. *J. Proteome Res.* **12**, 2449–2457 (2013).
28. Winter, D., Seidler, J., Ziv, Y., Shiloh, Y. & Lehmann, W. D. Citrate boosts the performance of phosphopeptide analysis by UPLC-ESI-MS/MS. *J. Proteome Res.* **8**, 418–424 (2009).
29. Seidler, J. et al. Metal ion-mobilizing additives for comprehensive detection of femtomole amounts of phosphopeptides by reversed phase LC-MS. *Amino Acids* **41**, 311–320 (2011).
30. Monroe, M. Molecular weight calculator. <https://alchemistmatt.com/resume/mwtoverview.html> (Pacific Northwest National Laboratory, 2007).
31. Werner, J. K., Speed, T. & Bhandari, R. Protein pyrophosphorylation by diphosphoinositol pentakisphosphate (InsP<sub>5</sub>). In *Inositol Phosphates and Lipids: Methods and Protocols* (ed Barker, C. J.) 87–102 (Springer, 2010).
32. Thota, S. G., Unnikannan, C. P., Thampatty, S. R., Manorama, R. & Bhandari, R. Inositol pyrophosphates regulate RNA polymerase I-mediated rRNA transcription in *Saccharomyces cerevisiae*. *Biochem. J.* **466**, 105–114 (2015).
33. Pinna, L. A. Casein kinase 2: an ‘eminence grise’ in cellular regulation? *Biochim. Biophys. Acta.* **1054**, 267–284 (1990).
34. Obenaus, J. C., Cantley, L. C. & Yaffe, M. B. Scansite 2.0: proteome-wide prediction of cell signalling interactions using short sequence motifs. *Nucleic Acids Res.* **31**, 3635–3641 (2003).
35. Van Der Lee, R. et al. Classification of intrinsically disordered regions and proteins. *Chem. Rev.* **114**, 6589–6631 (2014).
36. Collins, M. O., Yu, L., Campuzano, I., Grant, S. G. N. & Choudhary, J. S. Phosphoproteomic analysis of the mouse brain cytosol reveals a predominance of protein phosphorylation in regions of intrinsic sequence disorder. *Mol. Cell. Proteom.* **7**, 1331–1348 (2008).
37. Worley, J., Luo, X. & Capaldi, A. P. Inositol pyrophosphates regulate cell growth and the environmental stress response by activating the HDAC Rpd3L. *Cell Rep.* **3**, 1476–1482 (2013).
38. Lafontaine, D. L. J., Riback, J. A., Bascetin, R. & Brangwynne, C. P. The nucleolus as a multiphase liquid condensate. *Nat. Rev. Mol. Cell Biol.* **22**, 165–182 (2021).
39. Thul, P. J. et al. A subcellular map of the human proteome. *Science* **356**, eaal3321 (2017).
40. Ide, S., Imai, R., Ochi, H. & Maeshima, K. Transcriptional suppression of ribosomal DNA with phase separation. *Sci. Adv.* **6**, eabb5953 (2020).
41. Yao, R.-W. et al. Nascent pre-rRNA sorting via phase separation drives the assembly of dense fibrillar components in the human nucleolus. *Mol. Cell* **76**, 767–783 (2019).
42. Carbon, S. et al. AmiGO: online access to ontology and annotation data. *Bioinformatics* **25**, 288–289 (2009).
43. *Inositol Phosphates and Lipids: Methods and Protocols* (ed Barker, C. J.) (Springer, 2010).
44. Shah, A. & Bhandari, R. IP6K1 upregulates the formation of processing bodies by influencing protein–protein interactions on the mRNA cap. *J. Cell Sci.* **134**, jcs259117 (2021).
45. Fridy, P. C., Otto, J. C., Dollins, D. E. & York, J. D. Cloning and characterization of two human VIP1-like inositol hexakisphosphate and diphosphoinositol pentakisphosphate kinases. *J. Biol. Chem.* **282**, 30754–30762 (2007).
46. Chabert, V. et al. Inositol pyrophosphate dynamics reveals control of the yeast phosphate starvation program through 1,5-IP<sub>3</sub> and the SPX domain of Pho81. *eLife* **12**, RP87956 (2023).
47. Gu, C. et al. KO of 5-InsP<sub>7</sub> kinase activity transforms the HCT116 colon cancer cell line into a hypermetabolic, growth-inhibited phenotype. *Proc. Natl Acad. Sci. USA* **114**, 11968–11973 (2017).
48. Kuleshov, M. V. et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).
49. Xie, Z. et al. Gene set knowledge discovery with Enrichr. *Curr. Protoc.* **1**, e90 (2021).
50. Wilson, M. S., Jessen, H. J. & Saiardi, A. The inositol hexakisphosphate kinases IP6K1 and -2 regulate human cellular phosphate homeostasis, including XPR1-mediated phosphate export. *J. Biol. Chem.* **294**, 11597–11608 (2019).
51. Padmanabhan, U., Dollins, D. E., Fridy, P. C., York, J. D. & Downes, C. P. Characterization of a selective inhibitor of inositol hexakisphosphate kinases. Use in defining biological roles and metabolic relationships of inositol pyrophosphates. *J. Biol. Chem.* **284**, 10571–10582 (2009).
52. Moritoh, Y. et al. The enzymatic activity of inositol hexakisphosphate kinase controls circulating phosphate in mammals. *Nat. Commun.* **12**, 4847 (2021).
53. Zhao, J. et al. HisPhosSite: a comprehensive database of histidine phosphorylated proteins and sites. *J. Proteom.* **243**, 104262 (2021).
54. Ruprecht, B. et al. Comprehensive and reproducible phosphopeptide enrichment using iron immobilized metal ion affinity chromatography (Fe-IMAC) columns. *Mol. Cell. Proteom.* **14**, 205–215 (2015).
55. Giansanti, P., Tsiatsiani, L., Low, T. Y. & Heck, A. J. R. Six alternative proteases for mass spectrometry-based proteomics beyond trypsin. *Nat. Protoc.* **11**, 993–1006 (2016).
56. Chen, H.-K., Pai, C.-Y., Huang, J.-Y. & Yeh, N.-H. Human Nopp140, which interacts with RNA polymerase I: implications for rRNA gene transcription and nucleolar structural organization. *Mol. Cell. Biol.* **19**, 8536–8546 (1999).
57. Valdez, B. C., Henning, D., So, R. B., Dixon, J. & Dixon, M. J. The Treacher Collins syndrome (*TCOF1*) gene product is involved in ribosomal DNA gene transcription by interacting with upstream binding factor. *Proc. Natl Acad. Sci. USA* **101**, 10709–10714 (2004).
58. Lin, C. I. & Yeh, N. H. Treacle recruits RNA polymerase I complex to the nucleolus that is independent of UBF. *Biochem. Biophys. Res. Commun.* **386**, 396–401 (2009).
59. Tuan, J. C., Zhai, W. & Comai, L. Recruitment of TATA-binding protein–TAF<sub>I</sub> complex SL1 to the human ribosomal DNA promoter is mediated by the carboxy-terminal activation domain of upstream binding factor (UBF) and is regulated by UBF phosphorylation. *Mol. Cell. Biol.* **19**, 2872–2879 (1999).
60. Lin, C. Y., Navarro, S., Reddy, S. & Comai, L. CK2-mediated stimulation of Pol I transcription by stabilization of UBF–SL1 interaction. *Nucleic Acids Res.* **34**, 4752–4766 (2006).
61. Sahu, S. et al. Nucleolar architecture is modulated by a small molecule, the inositol pyrophosphate 5-InsP<sub>7</sub>. *Biomolecules* **13**, 153 (2023).
62. Shears, S. B. Diphosphoinositol polyphosphates: metabolic messengers? *Mol. Pharmacol.* **76**, 236–252 (2009).
63. Shah, A., Ganguli, S., Sen, J. & Bhandari, R. Inositol pyrophosphates: energetic, omnipresent and versatile signalling molecules. *J. Indian Inst. Sci.* **97**, 23–40 (2017).
64. Sridharan, S. et al. Systematic discovery of biomolecular condensate-specific protein phosphorylation. *Nat. Chem. Biol.* **18**, 1104–1114 (2022).

65. Crooks, G. E, Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).

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

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the

source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024

## Methods

### Pyrophosphoproteomics sample preparation workflow

**Cell lysis and digestion.** HEK293T cells (American Type Culture Collection, CRL-3216) were cultured by seeding  $2 \times 10^6$  cells into eight 15-cm culture plates in DMEM (5% FBS, 1 mM L-glutamine, 50 U ml<sup>-1</sup> penicillin, 100  $\mu$ g ml<sup>-1</sup> streptomycin). Plates were cultured for 3 d, with medium exchange on day 3 and harvest on day 4 (80–90% confluency). To lyse, plates were washed twice with 5 ml of 0.9% NaCl, and then 2 ml of lysis buffer (8 M urea, 75 mM NaCl, 50 mM Tris (pH 8.2), 1 mM NaF, 1 mM  $\beta$ -glycerophosphate, 1 mM sodium orthovanadate, 10 mM sodium pyrophosphate, 1 mM PMSF, one cOmplete EDTA-free protease inhibitor tablet (Roche) per 10 ml)<sup>24</sup> was added to seven plates. Plates were incubated at 4 °C for 10–15 min. The remaining plate was trypsinized, and cells were counted using a Bio-Rad cell counter (average cell number in HEK293T triplicate:  $6.6 \times 10^6$  cells per milliliter).

Lysed plates were scraped using a spatula, and the combined lysate was transferred into a 50-ml Falcon tube. The lysate was subjected to sonication at 4 °C (on ice, 50% output, 0.5 cycle rate,  $5 \times 30$  s, 30-s rest between pulses). The lysate was centrifuged (3,200g, 10 min, 4 °C); aggregated insoluble material on the surface was removed; and the supernatant was decanted from the cell pellet and retained. Supernatant protein concentration was determined by BCA assay (commercial kit by Thermo Fisher Scientific, average 112 mg).

To reduce and alkylate lysate proteins, DTT in Milli-Q water (5 mM concentration in sample) was added, and the sample was incubated at 37 °C for 1 h. After incubation, the sample was cooled to room temperature. Iodoacetamide in Milli-Q water (14 mM concentration in sample) was added, and the sample was incubated at room temperature in the dark for 30 min. Remaining iodoacetamide was quenched with a second aliquot of DTT (10 mM final concentration in sample), and the sample was incubated for 15 min at room temperature in the dark.

To digest, sequencing-grade modified trypsin (Promega, V5111) was used. On ice, the lysate was diluted approximately 1:5 with 25 mM Tris (pH 8.0). CaCl<sub>2</sub> (1 mM concentration in sample) was added. Trypsin was added at a ratio of approximately 1:50 protease to protein. The sample was incubated for 16 h at 37 °C with 500-r.p.m. agitation.

**Lysate desalting.** Neat TFA (approximately 0.4% of sample volume) was added (approximately pH 2), and the sample was centrifuged for 10 min at 3,200g. The supernatant of the centrifuged protein solution was desalted using four SepPak tC18 3-cc 500-mg cartridges (Waters, WAT043425; loading of approximately 20 mg per cartridge). Air pressure was used to speed up washing steps but was not used during the loading or elution steps. To prepare, columns were washed with MeCN (9 ml) and then with 3 ml of 50% MeCN in water with 0.5% AcOH. The columns were equilibrated with 0.1% TFA in water (9 ml) and then each loaded with the digested peptide samples in 0.4% TFA in water. The columns were washed with 0.1% TFA in water (9 ml) to desalt the peptides. The counter-ion was exchanged by a final wash with 0.5% AcOH in water (1 ml). Peptides were eluted with 6 ml of 50% MeCN and 0.5% AcOH in water. The eluent was lyophilized overnight to isolate tryptic digest as a white solid (average yield 84 mg)

**$\lambda$ -phosphatase treatment.** Lyophilized tryptic digest (50 mg) was dissolved in 10 ml of phosphatase reaction buffer (50 mM HEPES, 100 mM NaCl, 1 mM MnCl<sub>2</sub>, 2 mM DTT, 0.01% Brij 35). The solution was then treated with 50,000 units of  $\lambda$ -phosphatase and incubated at 37 °C with 300-r.p.m. agitation for 5 h. Neat TFA was added (approximately 0.4% of sample volume, to pH 2), and the sample was centrifuged (10 min, 3,200g) to remove any cellular debris. The samples were then desalted as described in the previous step, except that only three SepPak cartridges were used. The eluent was again lyophilized to isolate the tryptic digest as a white solid (average yield 45 mg).

**SIMAC enrichment.** A High Select Fe-NTA Phosphopeptide Enrichment Kit from Thermo Fisher Scientific was used to enrich the pyrophosphopeptides using a protocol adapted from the manufacturer's instructions to facilitate SIMAC<sup>26</sup>.

Tryptic digest (40 mg) treated with  $\lambda$ -phosphatase from the previous step was dissolved in 800  $\mu$ l of SIMAC loading buffer (0.1% TFA, 50% MeCN). Four columns were washed twice with 200  $\mu$ l of SIMAC loading buffer and then closed with a plug supplied with the kit and loaded with 200  $\mu$ l of the digest solution (10 mg per column). The columns were incubated for 30 min at room temperature. Every 10 min, the columns were gently tapped for 10 s to resuspend the Fe-NTA resin in the peptide solution.

After this binding step, columns were washed three times with 200  $\mu$ l of SIMAC loading buffer, once with 200  $\mu$ l of Milli-Q water, twice with 100  $\mu$ l of SIMAC buffer A (1% TFA, 20% ACN), to selectively elute monophosphopeptides) and once again with 200  $\mu$ l of Milli-Q water.

Elution was performed twice with 100  $\mu$ l of SIMAC buffer B (0.5% NH<sub>4</sub>OH in water) into clean 2-ml Protein LoBind Eppendorf tubes. The eluent from four columns was then combined into a single 2-ml Protein LoBind Eppendorf tube. This yielded an 800- $\mu$ l peptide solution that was directly frozen using liquid nitrogen and lyophilized overnight.

**Fractionation.** After enrichment, the material was dissolved in 100  $\mu$ l of high-performance liquid chromatography (HPLC) buffer A, spun at 21,000g for 10 min to pellet any insoluble components and transferred to an HPLC vial for fractionation. Separation was achieved using a decreasing pH gradient to exploit pKa differences among acidic amino acid side chains, phosphoryl groups and pyrophosphoryl groups, with lower pKa peptides expected to retain longer. Additionally, a decreasing acetonitrile gradient was employed with the aim to rapidly elute peptides with low negative charge in the early fractions of the separation. Fractionation was performed on an Agilent 1260 Infinity HPLC, equipped with an IonPac AS24 2-mm analytical ultra-hydrophilic SAX column (Thermo Fisher Scientific) and the corresponding IonPac AG24 guard column. HPLC buffers were prepared freshly.

HPLC buffer A: 20% MeCN in distilled water, 1% formic acid, pH 9.0 set with NH<sub>4</sub>OH. Buffer B: 1% MeCN, 1% NH<sub>4</sub>OH, pH 2.8 set with formic acid. Flow rate: 0.2 ml min<sup>-1</sup>. Gradient: 0–5 min, 100% A; 5–50 min, gradient increase to 100% B, 50–60 min, 100% B. Sample injection volume was 100  $\mu$ l. Fractions were collected every 5 min, 121-ml fractions in total.

Fractions were transferred to 2-ml Protein LoBind Eppendorf tubes and lyophilized overnight. The resulting solids were redissolved in 200  $\mu$ l of 20% MeCN and transferred to glass LC-MS vials and then dried using a vacuum centrifuge for 2 h at room temperature. Samples were stored at –20 °C until LC-MS analysis.

### LC-MS

Fractions were re-suspended with 50 mM citric acid in 3% acetonitrile solution before submitting to the LC-MS analysis. At this concentration, citric acid was tolerated by the LC-MS system and 50 mM was never exceeded. Each biological replicate was injected in two technical replicates, and the identified sites were combined. Sample separation was achieved by reverse-phase HPLC on a Thermo Fisher Scientific Dionex UltiMate 3000 system coupled online to an Orbitrap Fusion mass spectrometer (Thermo Fisher Scientific), operated with the Xcalibur software package (Thermo Fisher Scientific) version 4.4.16.14. For sample loading, a PepMap C18 trap column (Thermo Fisher Scientific) of 0.075 mm ID  $\times$  50 mm length, 3- $\mu$ m particle size and 100- $\text{Å}$  pore size was used. Reversed-phase separation was performed using a 50-cm analytical column (in-house packed with Poroshell 120 EC-C18, 2.7  $\mu$ m, Agilent Technologies) with mobile phase A containing 0.1% formic acid in water and mobile phase B containing 0.1% formic acid in acetonitrile. The gradient started with 4% mobile phase B reaching 80% mobile phase B in 101 min, with total run time of 120 min, including column wash and equilibration.

MS1 scans were acquired in the Orbitrap with a mass resolution of 120,000. MS1 scan range was set to 380–1,400  $m/z$ , with standard AGC target of  $4 \times 10^5$  and maximum injection time of 50 ms. Precursor ions with charge states +2 to +4 were isolated with an isolation window of 1.6  $m/z$  and dynamic exclusion of 15 s. Precursor ions were selected with precursor priority to the higher charge state. MS2 scans were acquired in the Orbitrap with AGC target of  $1 \times 10^4$  and maximum injection time of 100 ms. Precursor ions were fragmented using CID with a normalized collision energy of 25%. If neutral losses of 177.9432 Da from precursor ions were measured above a threshold of 15% of relative intensity in the CID scan, an additional spectrum of the same precursor ions was acquired using EThcD. EThcD spectra were measured in the Orbitrap with a resolution of 15,000 with AGC target of  $1 \times 10^5$ , maximum injection time of 2 s and normalized collision energy of 30%. Cycle time was set to 3 s.

### Sample preparation and LC–MS for intensity-based absolute quantification calculation

HEK293T cells were grown and harvested, and proteins were reduced, alkylated and digested with trypsin as described above. Then, 150  $\mu\text{g}$  of peptides was dissolved in 10 mM  $\text{NH}_4\text{OH}$  and loaded onto a Gemini 3- $\mu\text{m}$  C18 110- $\text{\AA}$  100  $\times$  1-mm high pH suitable column. Peptides were separated using an 85-min gradient with 10 mM  $\text{NH}_4\text{OH}$  in ACN as the eluent. All 24 fractions were dried in the SpeedVac, resuspended in 1% ACN/0.05% TFA and subjected to LC–MS. Online separation of the samples was achieved by reverse-phase HPLC on a Thermo Fisher Scientific Dionex UltiMate 3000 system connected to a PepMap C18 trap column (Thermo Fisher Scientific) and an in-house packed C18 column for reverse-phase separation at 300  $\text{nl min}^{-1}$  flow rate over a 120-min gradient. Samples were analyzed on an Orbitrap Fusion mass spectrometer with Instrument Control Software version 3.4. Data were acquired in DDA mode. MS1 scans were acquired in the Orbitrap with a mass resolution of 120,000. MS1 scan range was set to 375–1,500  $m/z$ , 100% normalized AGC target, 50-ms maximum injection time and 40-s dynamic exclusion. MS2 scans were acquired in the ion trap in rapid mode. The normalized AGC target was set to 100%, 35-ms injection time and an isolation window of 1.6  $m/z$ . Only precursors at charged states +2 to +4 were subjected to MS2. Peptides were fragmented using NCE 30%. Raw files were searched using MaxQuant (version 1.6.2.6) using the following parameters: MS1 mass tolerance, 10 ppm; MS2 mass tolerance, 0.5 Da; maximum number of missed cleavages, 2; minimum peptide length, 7; peptide mass, 500–4,600 Da; protein and PSM false discovery rate (FDR) 1%. Carbamidomethylation (+57.021 Da) on cysteines was used as a static modification. Oxidation of methionine (+15.995 Da) and acetylation of the protein N-terminus (+42.011 Da) were set as a variable modification. Data were searched against the human proteome retrieved from UniProt with intensity-based absolute quantification (iBAQ) calculation and match between runs activated. Data visualization was performed in R.

### Data analysis

Raw files were analyzed using Proteome Discoverer (Thermo Fisher Scientific) version 2.4.

EThcD spectra were selected by spectrum selector, accepting precursor masses of 350–5,000 Da.

Non-fragment filter was used as follows: Precursor ions were removed within a 1-Da window offset, and charge-reduced precursors and neutral losses were removed within a 0.5-Da window offset.

The SEQUEST HT node was used for peptide identification, searching against a full human proteome, digested by trypsin, with two missed cleavages allowed. Precursor mass tolerance was 10 ppm, and fragment mass tolerance was 0.02 Da. Carbamidomethylation (C) was set as a static modification, and phosphorylation (S,T,Y), pyrophosphorylation (S,T) and oxidation (M) were searched as dynamic modifications.

PSMs were filtered using the fixed value PSM validator, with a maximum accepted delta Cn of 0.5.

ptmRS was used for automated PTM assignment. PhosphoRS mode was set to false to detect phosphorylation and pyrophosphorylation in parallel.

### Validation of individual workflow steps

A set of five synthetic pyrophosphopeptides with a variety of sequence characteristics was produced as previously published<sup>66</sup>. In brief, a phosphorimidazole reagent was used to install a benzyl-protected  $\beta$ -phosphate onto a phosphoserine or phosphothreonine residue in a peptide, followed by Pd-catalyzed hydrogenolysis of the protecting group and purification by C8 reverse-phase HPLC. These standard peptides were used to validate individual steps of the sample preparation workflow. The sequences were the following:

- ppT-1: DAVTY-ppT-EHAK
- ppS-2: SQYHVDG-ppS-LEK
- ppS-3: LD-ppS-EEDSAWPTNEK
- ppS-4: NEEDEGH-ppS-NSSPR
- ppS-5: AQWTQE-ppS-FQSNNTNR

**$\lambda$ -phosphatase treatment.** Five synthetic pyrophosphopeptides (20 pmol each) and 0.5  $\mu\text{g}$  of HCT116 tryptic digest were spiked into  $\lambda$ -phosphatase buffer (50 mM HEPES, 100 mM NaCl, 1 mM  $\text{MnCl}_2$ , 2 mM DTT, 0.01% Brij 35). Total peptide amount was approximately 0.7  $\mu\text{g}$ .  $\lambda$ -phosphatase was added (1 U  $\mu\text{g}^{-1}$  as in the proteomics workflow), and samples were incubated at 37 °C and 300 r.p.m. for 5 h in a heater shaker.

Four replicates containing PP peptides, four negative control samples without enzyme, four replicates containing the corresponding monophosphopeptides and four control samples with monophosphopeptides, but without enzyme, were treated in this way.

All samples were desalted using C18 stage tips and subjected to LC–MS/MS analysis as described above.

Extracted total ion counts (TICs) of +2 precursor masses of the synthetic phosphopeptides and pyrophosphopeptides were integrated on MS1 level and normalized against total background to achieve a relative quantification of peptide abundance in samples versus controls.

**SIMAC enrichment.** Four replicates containing 5 mg of a  $\lambda$ -phosphatase-treated tryptic digest from HCT116 cells and five synthetic pyrophosphopeptides (20 pmol each) and four controls containing only the HCT116 digest were dissolved in 200  $\mu\text{l}$  of SIMAC loading buffer per sample (0.1% TFA, 50% MeCN) and subjected to SIMAC enrichment as in the pyrophosphoproteomics workflow described above. After enrichment, 20 pmol of each synthetic peptide was added to each control sample (positive control). Eluates were lyophilized, cleaned using C18 stage tips and dried by vacuum centrifugation. Peptides were redissolved in 6  $\mu\text{l}$  of injection buffer (50 mM sodium citrate, 3% MeCN) and subjected to LC–MS analysis as described above.

Extracted TICs of +2 precursor masses of the synthetic phosphopeptides and pyrophosphopeptides were integrated on MS1 level and normalized against total background to achieve a relative quantification of peptide abundance in samples versus controls.

**Use of citrate resuspension buffer.** Five synthetic pyrophosphopeptides (2  $\mu\text{l}$  of peptide mix, 20 pmol each) and 0.5  $\mu\text{g}$  of HCT116 tryptic digest (1  $\mu\text{l}$ ) were mixed with either 3  $\mu\text{l}$  of water or 3  $\mu\text{l}$  of citrate buffer (100 mM sodium citrate, 6% MeCN) and subjected to LC–MS analysis as described above.

Ion chromatograms of the most abundant free and iron-bound species  $[\text{M}+\text{Fe(III)-H}]^{2+}$  and  $[\text{M}+2\text{H}]^{2+}$  were extracted, integrated and normalized against total background to achieve a relative quantification of free versus iron-bound pyrophosphopeptides. The same experiment was performed with the corresponding monophosphopeptides.

## Expression constructs

The plasmids employed for expression in mammalian cells are as follows: N-terminally SFB (S-protein/FLAG/SBP)-tagged human TCOF1 (GenBank ID [NM\\_000356.4](#); gift from Maddika Subba Reddy, Centre for DNA Fingerprinting and Diagnostics); human UBF1 cDNA (GenBank ID [NM\\_014233.4](#); gift from Solomon Snyder, Johns Hopkins School of Medicine) subcloned into pCMV-Myc-N plasmid for expression with an N-terminal myc tag; UBF1 $\Delta$ 629-764 cloned into N-terminal myc-tagged destination vector using the Gateway cloning strategy (Thermo Fisher Scientific); human IWS1 (GenBank ID [NM\\_017969.3](#)) amplified using cDNA from HepG2 cells and cloned into N-terminal GFP or N-terminal SFB-tagged destination vectors; and N-terminally myc-tagged NOLC1 cDNA (GenBank ID [NM\\_001284388.2](#)) plasmid obtained from Sino Biological (HG16317-NM). The generation of catalytically active and inactive versions of C-terminally V5-tagged human IP6K1 (GenBank ID [NM\\_001242829.2](#)) was described previously<sup>44</sup>.

## Cell lines and transfection

Cell lines were grown in DMEM supplemented with 10% FBS, 1 mM L-glutamine, 100 U ml<sup>-1</sup> penicillin and 100  $\mu$ g ml<sup>-1</sup> streptomycin in a humidified incubator with 5% CO<sub>2</sub>. Cell culture reagents were from Thermo Fisher Scientific. Cells were transfected using polyethylenimine (PEI) (Polysciences) at a ratio of 1:3 (DNA:PEI) and harvested 48 h after transfection for further analyses. WT and *IP6K1*<sup>-/-</sup>*IP6K2*<sup>-/-</sup> DKO HCT116 cells were a gift from Adolfo Saiardi (University College London)<sup>50</sup>. *IP6K1*<sup>-/-</sup> HEK293T cell line was generated using the CRISPR-Cas9 strategy. An sgRNA sequence targeting exon 5 of IP6K1 (Supplementary Table 7) was designed using the Benchling tool (<https://www.benchling.com>) and cloned into pU6-2A-GFP-2A-Puro plasmid (a gift from P. Chandra Shekar, Centre for Cellular & Molecular Biology), which co-expresses Cas9. The plasmid was transfected into HEK293T cells, and, 48 h after transfection, cells were selected using 2  $\mu$ g ml<sup>-1</sup> puromycin for 3 d. Serial dilution was performed to isolate single-cell-derived colonies, which were screened for frameshift mutations by genotyping using a 3500xL Genetic Analyzer (Applied Biosystems). Primers used for genotyping are listed in Supplementary Table 7. IP6K1 knockout was confirmed by immunoblot analysis

## Analysis of cellular inositol pyrophosphates

HEK293T WT and *IP6K1*<sup>-/-</sup> cells expressing either active or kinase-dead (K226A) IP6K1 were seeded in 60-mm dishes and labeled with [<sup>3</sup>H]-inositol as described previously<sup>67</sup>. Upon attaining 30% confluence, cells were transferred to inositol-free DMEM (MP Biomedicals, D9802-06.25) containing 10% dialyzed FBS and 30  $\mu$ Ci myo-2-[<sup>3</sup>H] inositol (American Radiolabeled Chemicals, ART 0116B) for 2.5 d. The media were removed, and fresh media containing myo-2-[<sup>3</sup>H] inositol (30  $\mu$ Ci) were added for another 2.5 d, when plasmids expressing IP6K1-V5 or IP6K1-V5 K226A were transfected into *IP6K1*<sup>-/-</sup> cells. Upon achieving isotopic labeling, cells were collected in chilled PBS. Soluble inositol phosphates were extracted by the addition of 350  $\mu$ l of extraction buffer (0.6 M HClO<sub>4</sub>, 2 mM EDTA, 0.2 mg ml<sup>-1</sup> phytic acid) for 15–20 min, followed by centrifugation at 21,000g for 10 min. The supernatant containing soluble inositol phosphates was collected, and lipid inositides in the pellet were extracted with 1 ml of lipid extraction buffer (0.1 N NaOH, 0.1% Triton X-100) at room temperature and counted in a liquid scintillation counter (Tri-Carb 2910 TR, PerkinElmer). The soluble inositol phosphate extract was mixed with approximately 120  $\mu$ l of neutralization solution (1 M K<sub>2</sub>CO<sub>3</sub>, 5 mM EDTA). Tubes were left open on ice for 1 h, followed by centrifugation at 21,000g for 10 min at 4 °C. The extracted inositol phosphates were resolved by HPLC (515 or 5125 HPLC pumps, Waters) on a PartiSphere SAX column (4.6 mm  $\times$  125 mm, HiChrome) using a gradient of buffer A (1 mM EDTA) and buffer B (1 mM EDTA and 1.3 M (NH<sub>4</sub>)<sub>2</sub>HPO<sub>4</sub> (pH 3.8)) as follows: 0–5 min, 0% B; 5–10 min, 0–20% B; 10–70 min, 20–100% B; 70–80 min, 100% B. Then, 1-ml fractions containing soluble inositol phosphates were mixed

with 3 ml of scintillation cocktail (Ultima-Flo AP) and counted. The soluble inositol phosphate in each fraction was normalized to total lipid inositide in the sample.

## Immunofluorescence, protein pulldown and immunoblotting

The antibodies used in this study are listed in Supplementary Table 8. For immunofluorescence, cells seeded on glass coverslips were fixed with 4% paraformaldehyde for 15 min and permeabilized using PBS containing 0.15% Triton X-100 (PBST) for 15 min at room temperature. Cells were incubated with blocking solution (5% BSA in PBST) for 1 h at room temperature and then with primary antibodies diluted in the blocking solution overnight at 4 °C. Cells were washed three times with PBST and incubated with fluorophore-conjugated secondary antibodies for 1 h at room temperature. After incubation, the cells were washed with PBST and mounted on glass slides using mounting medium with DAPI (Vector Laboratories). Images in Fig. 4b were captured on a Leica TCS SP8 confocal microscope equipped with 405-nm, 488-nm, 514-nm, 561-nm and 633-nm lasers using a  $\times$ 63 1.4 NA oil immersion objective. Images in Fig. 4c,d were captured using the Elyra 7 structured illumination microscopy (SIM) module of the Zeiss LSM 980 confocal microscope, equipped with 405-nm, 488-nm, 561-nm and 642-nm lasers using a  $\times$ 63 1.4 NA oil immersion objective. All immunofluorescence images are in z-stacks and are shown as maximum intensity projections using LAS X (Fig. 4b) or ZEN (Fig. 4c,d) software.

For protein pulldown, cells were harvested 48 h after transfection and lysed for 1 h at 4 °C in 1% lysis buffer (50 mM HEPES, pH 7.4, 150 mM NaCl, 1% Nonidet P-40, 1 mM EDTA, protease and phosphatase inhibitor cocktail) followed by sonication for 10 s at 30% amplitude (SONICS, VCX 750). To the lysate, the specific antibody was added and incubated overnight at 4 °C. The antibody-bound protein was pulled down using Protein A or Protein G Sepharose beads (GE Healthcare) for 1 h. The beads were washed with lysis buffer and used for [<sup>32</sup>P]5PP-InsP<sub>5</sub>-mediated pyrophosphorylation. For SFB-tagged proteins, streptavidin sepharose beads (GE Healthcare) were added to the cell lysate and incubated at 4 °C before washing and pyrophosphorylation with [<sup>32</sup>P]5PP-InsP<sub>5</sub>. For MS analysis of overexpressed proteins (Fig. 5a), 4% of cell lysate was used for immunoblotting, and 90% was pulled down on streptavidin sepharose beads. The beads were washed with lysis buffer and boiled in 1 $\times$  Laemmli buffer, and proteins were resolved by SDS-PAGE. The bands were visualized with 0.2% Coomassie brilliant blue R-250, excised, digested in-gel with trypsin and subjected to neutral-loss-triggered EThcD MS as described above. For immunoblotting, proteins were transferred to a PVDF membrane and detected using standard western blotting techniques with protein-specific or tag-specific antibodies. To monitor protein stability, *IP6K1*<sup>-/-</sup> HEK293T cells were transfected to express either active IP6K1-V5 or kinase-dead IP6K1-V5 (K226A), and, 24–30 h after transfection, cells were treated with cycloheximide (100  $\mu$ g ml<sup>-1</sup>) for the indicated time. Cells were lysed for 1 h at 4 °C in lysis buffer (50 mM HEPES, pH 7.4, 100 mM NaCl, 0.5% Nonidet P-40, 1 mM EDTA, containing protease and phosphatase inhibitor cocktail), and lysates were subjected to immunoblotting. For all immunoblotting experiments, chemiluminescence was detected using a GE ImageQuant LAS 500 imager, and protein bands were quantified using Fiji software.

## In-gel digestion after pulldown

Coomassie-stained bands were cut out and washed with 200  $\mu$ l of wash buffer (50 mM TEAB in 1:1 water:MeCN) until destained. In case of strongly colored bands, this process was repeated to fully remove Coomassie. The wash buffer was discarded; gel slices were equilibrated for 10 min at 30 °C in 50 mM TEAB; and the supernatant was discarded. Then, 200  $\mu$ l of MeCN was added to shrink and dry gel pieces. This step was repeated once more.

IWS1 samples were then reduced and alkylated. NOLC1 and TCOF1, having no Cys residues, were not. Then, 100  $\mu$ l of 5 mM DTT was added to each tube and incubated for 45 min at 56 °C, 300 r.p.m. on a shaker,

and the supernatant was discarded. Next, 100  $\mu\text{l}$  of 40 mM CAA in 50 mM TEAB was added and incubated for 30 min in the dark at room temperature, and the supernatant was discarded. Gel slices were incubated in 200  $\mu\text{l}$  of wash buffer for 5 min at 30 °C, 300 r.p.m., and then in 200  $\mu\text{l}$  of MeCN for 5 min. The MeCN step was repeated once more.

For digestion, 0.2  $\mu\text{g}$  of trypsin in 30  $\mu\text{l}$  of 50 mM TEAB was added to each gel slice. In some cases, more buffer was added until the gel piece was covered. Samples were incubated overnight at 37 °C and 300 r.p.m.

The next day, samples were briefly centrifuged, and digestion was stopped by adding 30  $\mu\text{l}$  of 0.5% TFA in MeCN. The supernatant was transferred to an MS vial. Next, 20  $\mu\text{l}$  of MeCN was added to the gel piece to shrink and dry it and then was added into the same vial. The combined solutions were dried in a vacuum centrifuge and stored at -20 °C until LC-MS analysis.

Samples were then redissolved in 10  $\mu\text{l}$  of citrate injection buffer (50 mM sodium citrate, 3% MeCN) and sonicated for 5 min. Next, 2  $\mu\text{l}$  was injected and analyzed in the same manner as the pyrophospho-proteomics samples described above.

### Protein pyrophosphorylation using [ $^{32}\text{P}$ ]5PP-InsP<sub>5</sub>

[ $^{32}\text{P}$ ]ATP was procured from JONAKI/BRIT. Radiolabeled [ $^{32}\text{P}$ ]5PP-InsP<sub>5</sub> synthesis was conducted as described previously<sup>68</sup>, with a few modifications. In brief, for a 200- $\mu\text{l}$  reaction, 200  $\mu\text{M}$  InsP<sub>6</sub> (SiChem) was incubated for 6 h at 37 °C with 3 mCi [ $^{32}\text{P}$ ]ATP and 50  $\mu\text{M}$  unlabeled ATP in the presence of 80 ng  $\mu\text{l}^{-1}$  purified *Entamoeba histolytica* InsP<sub>6</sub> kinase (IP6KA) in buffer containing 100 mM MES, pH 6.8, 30 mM MgSO<sub>4</sub>, 250 mM NaCl and 5 mM DTT. [ $^{32}\text{P}$ ]5PP-InsP<sub>5</sub> was purified by strong anion exchange HPLC (Waters) as described previously<sup>68</sup>.

Proteins were subjected to phosphorylation by CK2 and pyrophosphorylation by [ $^{32}\text{P}$ ]5PP-InsP<sub>5</sub> as described previously<sup>16,20,31</sup>. For in vitro pyrophosphorylation assays, immunoprecipitated proteins on beads were first pre-phosphorylated with CK2 (New England Biolabs) in protein kinase buffer (New England Biolabs) and 0.5 mM Mg<sup>2+</sup>-ATP for 30 min at 30 °C. Beads were washed in cold PBS and incubated in pyrophosphorylation buffer (25 mM HEPES, pH 7.4, 50 mM NaCl, 6 mM MgCl<sub>2</sub>, 1 mM DTT) containing 3–5  $\mu\text{Ci}$  [ $^{32}\text{P}$ ]5PP-InsP<sub>5</sub> at 37 °C for 15 min. Samples on beads were mixed with LDS sample buffer (Thermo Fisher Scientific), heated at 95 °C for 5 min, resolved on a 4–12% NuPAGE Bis-Tris gel (Thermo Fisher Scientific) and transferred to a PVDF membrane (GE Life Sciences). Pyrophosphorylation was detected using a phosphorimager (Typhoon FLA-9500), and proteins were detected by immunoblotting. To improve visualization, the phosphorimager scan and immunoblots were subjected to uniform ‘Levels’ adjustment in Adobe Photoshop.

Back-pyrophosphorylation was conducted as described previously<sup>19–21</sup>, and the method is explained schematically in Fig. 5b. Overexpressed or endogenous UBF1 was immunoprecipitated from *IP6KI*<sup>-/-</sup> HEK293T cells expressing either active or kinase-dead IP6K1. The immunoprecipitated proteins were subjected to pyrophosphorylation as described above. Radiolabeled protein as a fraction of total immunoprecipitated protein was quantified using Fiji software<sup>69</sup>.

### RT-qPCR

Cells were lysed using TRIzol reagent (Thermo Fisher Scientific), and total RNA was extracted using a kit (HiMedia). Where indicated, cells were treated with 10  $\mu\text{M}$  TNP<sup>51</sup> (Merck Millipore), 1  $\mu\text{M}$  SC-919 (synthesized as described by Kröber et al.<sup>70</sup>) or DMSO for 5 h before RNA isolation. cDNA was prepared by reverse transcription with SuperScript Reverse Transcriptase III (Thermo Fisher Scientific) using random hexamers. Two sets of 45S pre-rRNA specific primers were designed to amplify the region between sites 01 (also called A') and A0 on the human 47S pre-rRNA transcript<sup>71</sup>. GAPDH transcript was used as an internal control. The sequences of the primers are provided in Supplementary Table 7. qPCR was conducted with SYBR Green PCR Master Mix on a CFX96 Touch Real-Time PCR Detection System (Bio-Rad). All samples

were run in technical duplicates. The fold change ( $\Delta\Delta C_T$ ) method<sup>72</sup> was used to calculate the difference in transcript levels.  $\Delta C_T$  refers to the  $C_T$  value for the target pre-rRNA normalized to the  $C_T$  value for GAPDH in the same sample.  $\Delta\Delta C_T$  values were determined as a relative change in  $\Delta C_T$  in HCT116 DKO compared to WT cells or drug-treated compared to control cells.  $2^{-\Delta\Delta C_T}$  was used to represent fold changes.

### Statistical analysis

Statistical analyses and graph preparation were done using GraphPad Prism 8 software. The number of independent experimental replicates for each analysis or image is provided in the respective figure legend. Quantified data are presented as mean  $\pm$  s.e.m. for the indicated number of biological replicates ( $n$ ).  $P$  values were calculated using a one-sample  $t$ -test.  $P \leq 0.05$  was considered statistically significant.

### Reporting summary

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

### Data availability

The reported proteomics data are publicly available on the jPOST repository<sup>73</sup> under accession numbers JPST001935 / [PXD038962](https://doi.org/10.1038/s41589-024-01613-5), JPST001934 / [PXD038963](https://doi.org/10.1038/s41589-024-01613-5) and JPST002429 / [PXD048031](https://doi.org/10.1038/s41589-024-01613-5). Source data are provided with this paper.

### References

- Marmelstein, A. M., Yates, L. M., Conway, J. H. & Fiedler, D. Chemical pyrophosphorylation of functionally diverse peptides. *J. Am. Chem. Soc.* **136**, 108–111 (2014).
- Jadav, R. S. et al. Deletion of inositol hexakisphosphate kinase 1 (IP6K1) reduces cell migration and invasion, conferring protection from aerodigestive tract carcinoma in mice. *Cell. Signal.* **28**, 1124–1136 (2016).
- Azevedo, C., Burton, A., Bennett, M., Onnebo, S. M. N. & Saiardi, A. Synthesis of InsP<sub>7</sub> by the inositol hexakisphosphate kinase 1 (IP6K1). In *Inositol Phosphates and Lipids: Methods and Protocols* (ed Barker, C. J.) 73–85 (Springer, 2010).
- Schindelin, J. et al. Fiji—an open platform for biological image analysis. *Nat. Methods* **9**, 676–682 (2009).
- Kröber, T., Bartsch, S. M. & Fiedler, D. Pharmacological tools to investigate inositol polyphosphate kinases—enzymes of increasing therapeutic relevance. *Adv. Biol. Regul.* **83**, 100836 (2022).
- Mullineux, S.-T. & Lafontaine, D. L. J. Mapping the cleavage sites on mammalian pre-rRNAs: where do we stand? *Biochimie* **94**, 1521–1532 (2012).
- Schmittgen, T. D. & Livak, K. J. Analyzing real-time PCR data by the comparative  $C_T$  method. *Nat. Protoc.* **3**, 1101–1108 (2008).
- Okuda, S. et al. jPOSTrepo: an international standard data repository for proteomes. *Nucleic Acids Res.* **45**, D1107–D1111 (2017).

### Acknowledgements

The authors thank the Leibniz-Forschungsinstitut für Molekulare Pharmakologie peptide facility (I. Kretzschmar) for assistance with peptide synthesis and B. Bodganov for providing the R script used in Fig. 3.

For important preliminary experiments, we would like to thank R. K. Harmel (Leibniz-Forschungsinstitut für Molekulare Pharmakologie). We also thank the Centre for DNA Fingerprinting and Diagnostics microscopy facility for support; R. Manorama for radiolabeled PP-InsP<sub>5</sub> synthesis; J. Ladke for generation of IP6K1 knockout cell lines; M. S. Reddy for sharing Gateway cloning vectors; and members of the Laboratory of Cell Signalling at the Centre for DNA Fingerprinting and Diagnostics for their valuable feedback.



R.B. acknowledges support from the Department of Biotechnology, Ministry of Science and Technology, Government of India (BT/PR29960/BRB/10/1762/2019 and IC-12025(11)/2/2020/ICD-DBT); the Science and Engineering Research Board, Ministry of Science and Technology, Government of India (CRG/2019/002597); and Centre for DNA Fingerprinting and Diagnostics core funds. A.S. and S.S. are recipients of research fellowships from the Department of Biotechnology, Ministry of Science and Technology, Government of India.

J.A.M.M. and L.K. gratefully acknowledge funding from the Deutsche Forschungsgemeinschaft (DFG) (grant 278001972 – TRR186).

### Author contributions

J.A.M.M., A.S., R.B. and D.F. designed and conceived the study. J.A.M.M. developed the proteomics platform and the data analysis workflow. L.K. conducted pyrophosphoproteomics experiments and workflow validation. A.S. carried out in vitro pyrophosphorylation assays and microscopy studies. M.R., M.N.-H., M.P., E.K. and F.L. provided guidance and assistance for proteomics experiments. S.G. and S.S. conducted measurements to determine cellular PP-InsP levels and protein stability. J.A.M.M. and A.S. contributed equally. J.A.M.M., A.S., R.B. and D.F. wrote the paper. Figures and schematics were crafted by L.K. and A.S. All authors engaged in discussions and provided feedback on the final paper.

### Funding

Open access funding provided by Leibniz-Forschungsinstitut für Molekulare Pharmakologie im Forschungsverbund Berlin e.V. (FMP).

### Competing interests

The authors declare no competing interests.

### Additional information

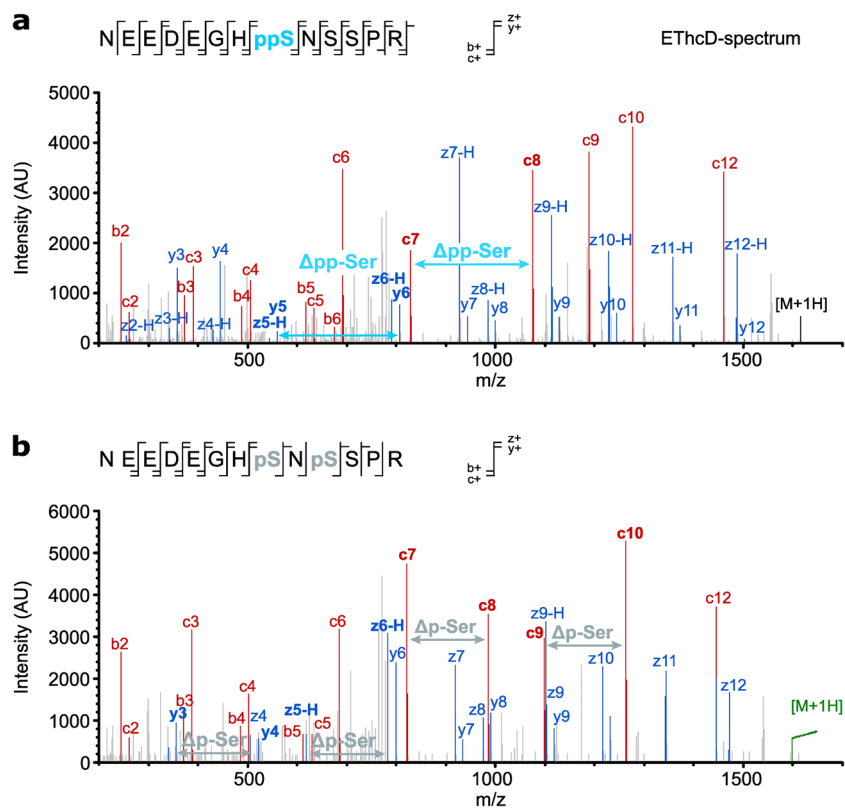
**Extended data** is available for this paper at <https://doi.org/10.1038/s41589-024-01613-5>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41589-024-01613-5>.

**Correspondence and requests for materials** should be addressed to Rashna Bhandari or Dorothea Fiedler.

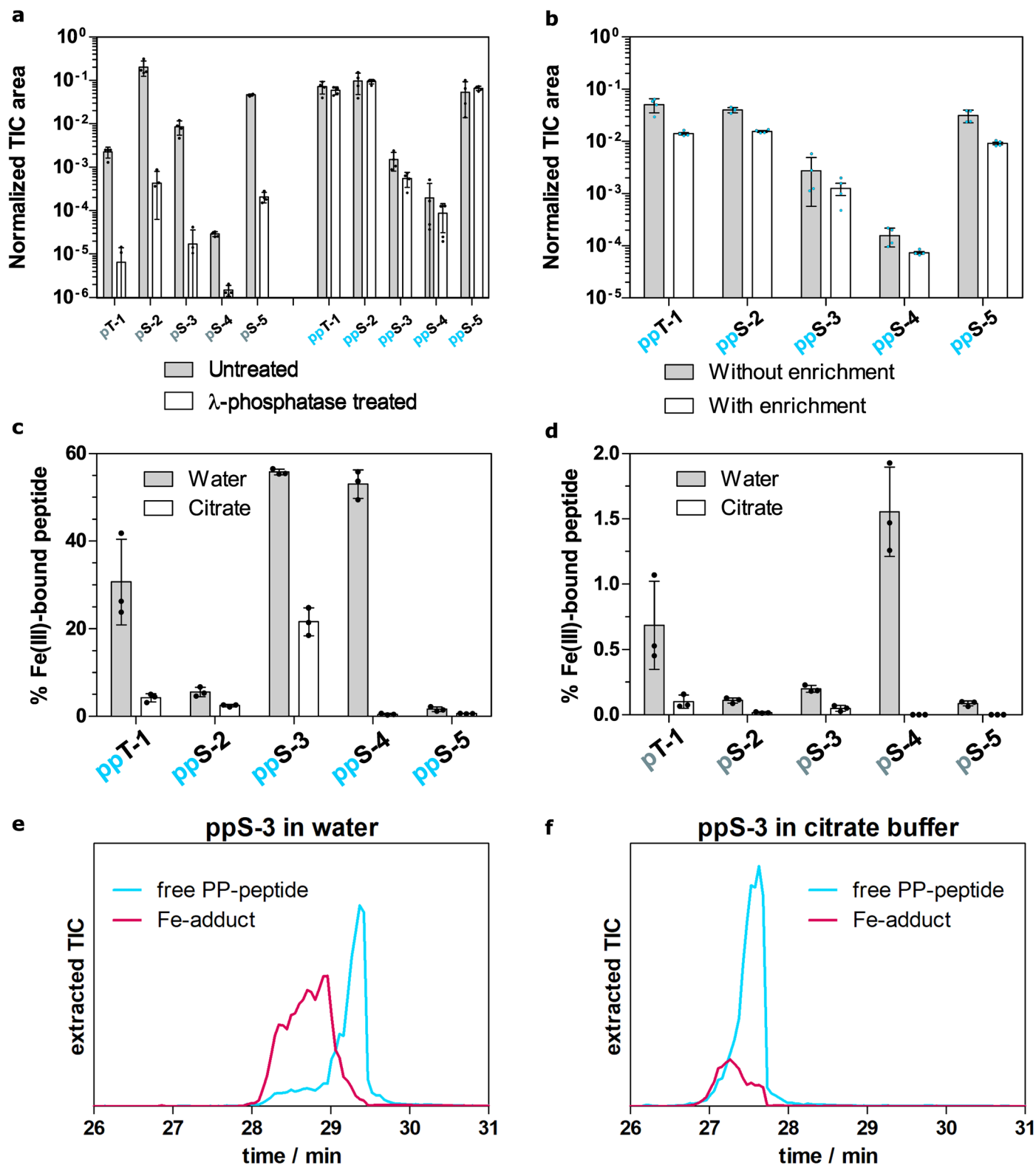
**Peer review information** *Nature Chemical Biology* thanks Eiichiro Nagata and the other, anonymous reviewers for their contribution to the peer review of this work.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).



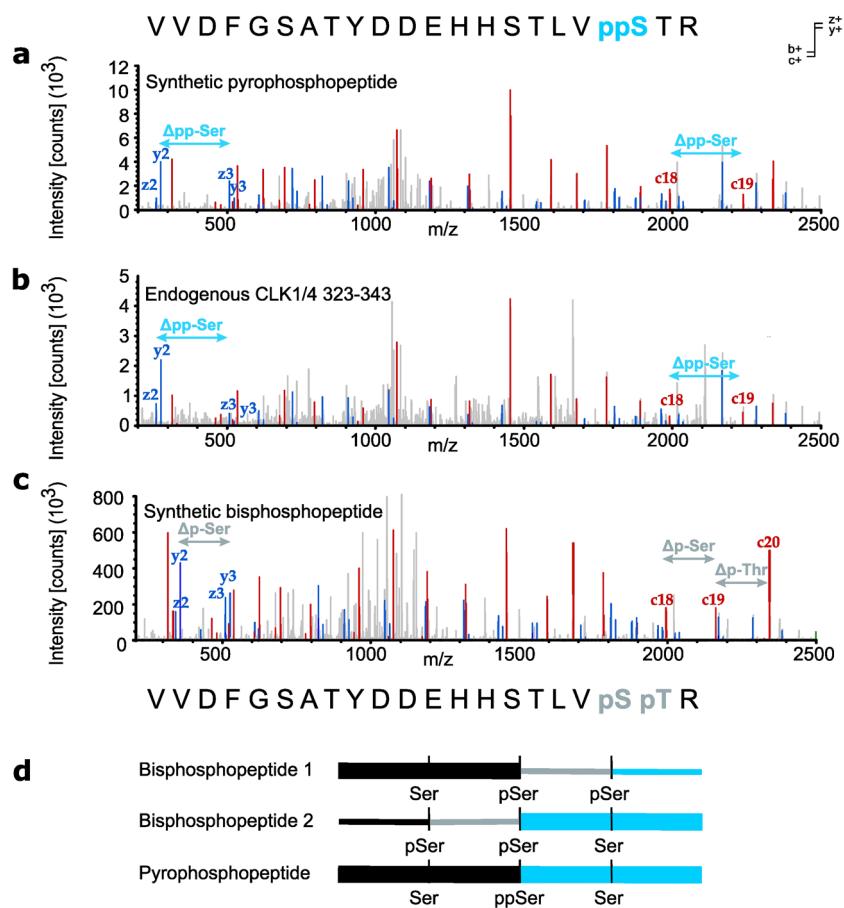
**Extended Data Fig. 1 | EThcD spectra of model peptides. a)** EThcD fragmentation spectrum of the pyrophosphopeptide from Fig. 1c. **b)** EThcD fragmentation spectrum of the corresponding bisphosphopeptide. N-terminal fragments are labeled in red, C-terminal fragments in blue. Arrows

indicate the characteristic fragments, which enable a distinction between pyrophosphorylated and bisphosphorylated peptide species in spite of their equal MS1 mass.



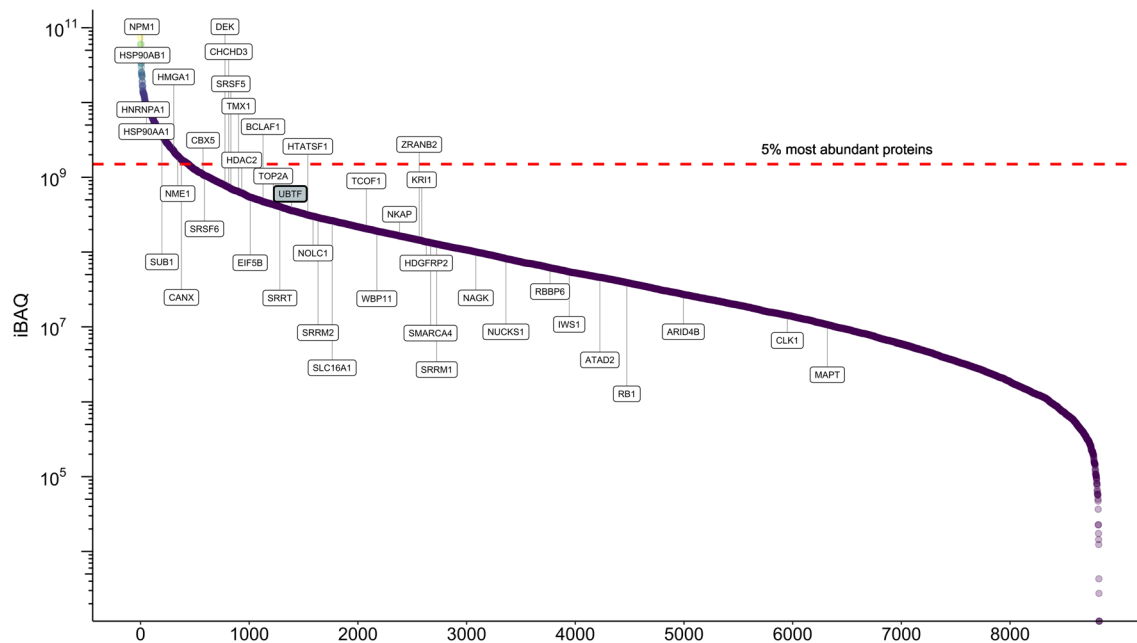
**Extended Data Fig. 2 | Validation of workflow steps.** **a)** Pyrophosphopeptide stability to  $\lambda$ -phosphatase. Five synthetic pyrophosphopeptides and the corresponding monophosphopeptides were treated with the enzyme for 5 h. Data are presented as mean  $\pm$  SD of four replicates. **b)** Showing retention of pyrophosphopeptides during SIMAC enrichment. Five synthetic pyrophosphopeptides were spiked into a background of HCT116 tryptic digest before or after enrichment and detected by LC-MS. Data are presented as mean  $\pm$  SD of four replicates.

**c-f)** Improvement of LC-peak shape and reduction of Fe-adducts by citrate resuspension buffer (50 mM sodium citrate, 3% MeCN). **c)/d)** Proportion of Fe-adducts of synthetic pyrophosphopeptides (**c)** or monophosphopeptides (**d)** detected by LC-MS after sample resuspension in water vs. citrate buffer. Data are presented as mean  $\pm$  SD of three replicates. **e)/f)** Representative total ion chromatograms (three replicates) of a synthetic pyrophosphopeptide in free or  $\text{Fe}^{3+}$ -bound form after resuspension in water (**e)** or citrate buffer (**f**).



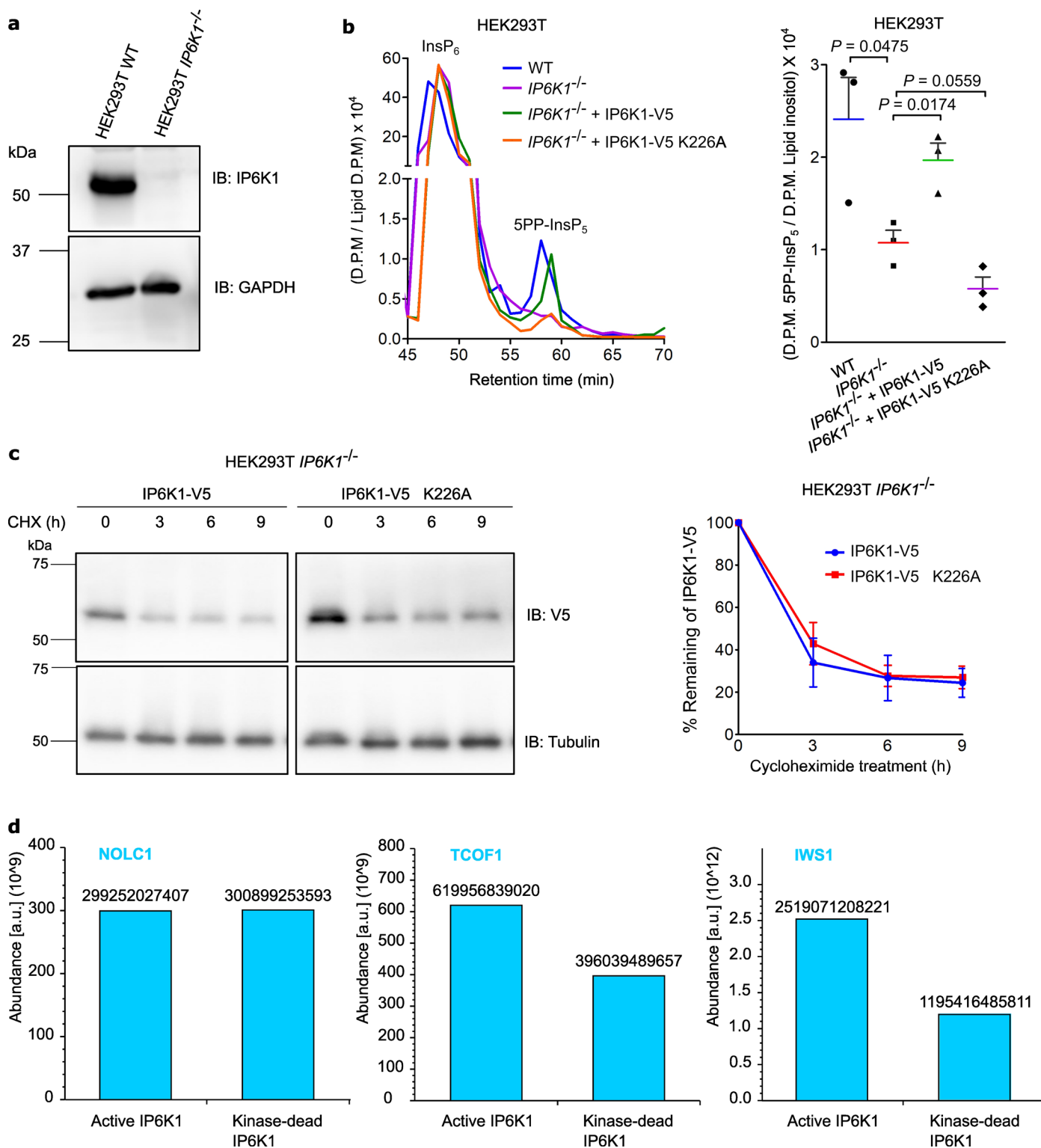
**Extended Data Fig. 3 | Fragmentation of pyrophospho- vs. bisphosphopeptides. a-c)** EThcD spectra of an endogenous pyrophosphopeptide (CLK1, residues 323-343) (**a**) and the corresponding synthetic pyrophospho- (**b**) and bisphosphopeptides (**c**), highlighting the

fragment ions essential for distinguishing the two modifications. **d)** Schematic demonstrating how a mixture of bisphosphopeptides may contain all fragments typical of the corresponding pyrophosphopeptide and lead to false-positive assignment of a pyrophosphopeptide.



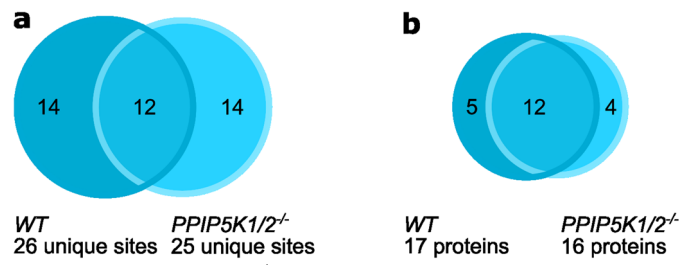
**Extended Data Fig. 4 | Expression levels of pyrophosphoproteins.** iBAQ abundance scores of 8821 identified proteins from HEK293T cells (after removing decoys and contaminants) were plotted against their rank in the list of iBAQs. All identified pyrophosphoproteins from a HEK293T biological triplicate are

highlighted in white boxes. UBF1, another substrate of pyrophosphorylation that was detected by radiolabeling, but not mass spectrometry, ranks among the more abundant proteins and is highlighted in grey.



**Extended Data Fig. 5 | Controls for pulldown experiments.** **a**) Representative immunoblots demonstrating the absence of IP6K1 in HEK293T *IP6K1*<sup>-/-</sup> knockout cell line (n = 4). **b**) Left: representative HPLC profiles of [<sup>3</sup>H]-inositol labeled HEK293T wild type cells, *IP6K1*<sup>-/-</sup> knockout cells, and *IP6K1*<sup>-/-</sup> knockout cells expressing V5-epitope tagged active or kinase-dead IP6K1. Soluble inositol phosphate counts were normalized to the total lipid inositol count for each sample. Peaks corresponding to InsP<sub>6</sub> and 5PP-InsP<sub>5</sub> are indicated. Right: Level of 5PP-InsP<sub>5</sub> normalized to the total lipid inositol in the four cell lines from a (mean ± SEM, n = 3 independent experiments) analyzed using a two-tailed

unpaired Student's t test. **c**) Protein stability of active and kinase-dead IP6K1 expressed in *IP6K1*<sup>-/-</sup> HEK293T cells subjected to treatment with cycloheximide (100 µg/mL) to block protein synthesis for the indicated time. α-tubulin was used as a loading control. Graphs show the levels of IP6K1 at each time point, normalized to the untreated sample (mean ± SEM, n = 5 independent experiments). Loss of kinase activity did not alter the stability of IP6K1. **d**) Abundance of target proteins in pulldown samples. Comparison between samples from a background with high (IP6K1-kinase active, KA) or low (IP6K1-kinase dead, KD) PP-InsP levels.



**Extended Data Fig. 6 | Pyrophosphoproteomics on HCT116 *PPIP5K1/2<sup>-/-</sup>* cells. a) Overlap of manually validated pyrophosphorylation sites in wild type vs knockout cells. b) Overlap of modified proteins.**

## Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

#### Data collection

The Orbitrap Fusion mass spectrometer was operated using the Xcalibur software package (ThermoFisher Scientific) version 4.4.16.14. 3500xL Genetic Analyzer (Applied Biosystems) was used for the genotyping of IP6K1-/- HEK293T cell line. For immunofluorescence, images were taken on Leica TCS SP8 confocal microscope using LAS X software version 3.7.4.23463 or Elyra 7 structured illumination microscopy (SIM) module of the Zeiss LSM 980 confocal microscope using ZEN software version 3.2. RT-qPCR data were collected on a CFX96 Touch Real-Time PCR Detection System (Biorad) using CFX Maestro software version 1.1. Autoradiographs were detected using a phosphorimager Typhoon FLA-9500 (GE Healthcare). Chemiluminescence was detected using the GE ImageQuant LAS 500 imager (GE Healthcare).

#### Data analysis

MS raw data was analyzed using Proteome Discoverer (ThermoFisher Scientific) version 2.4.0.305, FreeStyle (ThermoFisher Scientific) version 1.7.73.12, Molecular Weight Calculator (Matthew Monroe <https://alchemistmatt.com/resume/mwtoverview.html>) version 6.50 and MaxQuant version 2.0.3.0. IuPred2A was used for disorder prediction. Scansite 4.0 was used to predict motifs that are likely to undergo phosphorylation by specific protein kinases. Gene ontology term analysis was done using Enrichr. Radiolabeled protein as a fraction of total immunoprecipitated protein was quantified using Fiji software version 1.53t. To improve visualization, the phosphorimager scan and immunoblots were subjected to uniform "Levels" adjustment in Adobe Photoshop version 10.0. Statistical analyses and graph preparation were done using GraphPad Prism 8.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.



## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The reported proteomics data is publicly available on the jPOST repository under the accession numbers: JPST001935 / PXD038962, JPST001934 / PXD038963 and JPST002429 / PXD048031. Source Data are provided with this paper.

## Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender

Population characteristics

Recruitment

Ethics oversight

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences  Behavioural & social sciences  Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

Data exclusions

Replication

Randomization

Blinding

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials &amp; experimental systems

## Methods

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Antibodies

## Antibodies used

Primary antibodies used in this study for immunofluorescence (IF), immunoblotting (IB), or immunoprecipitation (IP), along with their respective antibody dilutions were: anti-IP6K1 (Merck Millipore, HPA040825, C118624, IF 1:500; IB 1:4000); anti-UBF1 (F:9, Santa Cruz, sc-13125, I1509, IF 1:300; IB 1:500; IP 1 µg); anti-FBL (38F3, Santa Cruz, sc-56676; IF 1:50); anti-FLAG (M2, Merck Millipore, F1804, SLBN2445V, IB 1:10,000); anti-GFP (Thermo Fisher Scientific, A11122, 57204A, IB 1:7000; IP 2 µg); anti-myc (9E10, Merck Millipore, M4439, 84292, IB 1:10,000; IP 1 µg); anti-V5 tag (Thermo Fisher Scientific, R960-25, 2129366, IB 1:7000); anti-GAPDH (Merck Millipore, G8795, 104M4751V, IB 1:10,000); anti- $\alpha$ -tubulin (Merck, T9026; IB 1:10,000). HRP-conjugated secondary goat anti-rabbit (4010-05, A4211-ZB66D) and goat anti-mouse (1031-05, D2719-N941C) IgG antibodies for IB (1:10,000) were obtained from Southern Biotech. Fluorophore-conjugated secondary Alexa Fluor 488 (goat anti-mouse IgG; A-11029) and Alexa Fluor 568 antibodies (goat anti-rabbit IgG; A-11011) for IF (1:500) were from Thermo Fisher Scientific.

## Validation

As per the manufacturer's website, anti-IP6K1 (Merck Millipore, HPA040825) is validated in the human cell line U-2 OS for IF and human cell line PC-3 for IB; anti-UBF1 (F:9, Santa Cruz, sc-13125) is validated in human cell lines HL-60, mouse cell lines WEHI-231 and NIH/3T3 nuclear extracts for IB and human cell line SW480 for IF; anti-FBL (38F3, Santa Cruz, sc-56676) is validated in the human cell line SH-SY5Y for IF and human cell line Hep G2 whole cell lysates for IB; anti-GAPDH (Merck Millipore, G8795) is validated in several human cell lines including HeLa, A-431 for IB. anti-myc (9E10, Merck Millipore, M4439) is validated in the human cell line HEK293 overexpressing C-terminal c-Myc tagged fusion protein for IF and whole extract of *E. coli* cells expressing LacZ tagged with c-Myc for IB. anti-FLAG (M2, Merck Millipore, F1804) is validated for the detection of FLAG fusion proteins in mammalian, plant, and bacterial expression systems for IB; anti-V5 tag (Thermo Fisher Scientific, R960-25) is validated in HEK-293 cells transiently overexpressing V5-His-LacZ for IF and whole cell extracts of HEK-293 transiently overexpressing V5-His-LacZ for IB; anti-GFP (Thermo Fisher Scientific, A11122) is validated using H3-GFP construct transfected in HEK-293E cells for IF and GFP-transfected HeLa cell lysates for IF; anti- $\alpha$ -tubulin (Merck, T9026) is validated in many cell line including HeLa, JURKAT, COS7, NIH-3T3 for IB.

## Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

## Cell line source(s)

HEK293T cells were from ATCC, CRL-3216.  
 HCT 116 cells used in Fig. 2e were from ATCC, CCL-247.  
 U-2 OS cells, used in Fig 4b, c and 5g, were from ATCC, HTB-96.  
 HEK293T cells, used in Fig 4b, f, g, h and i, were from ATCC, CRL-3216, and a gift from Solomon Synder, Johns Hopkins School of Medicine, Baltimore, USA.  
 HCT116 cells, used in Fig 5 e and f, were a gift from Adolfo Saiardi, University College London, UK.  
 HEK293T IP6K1/- cells, used in Fig 5a, c and d, were generated in the lab of Rashna Bhandari, Centre for DNA Fingerprinting and Diagnostics, Hyderabad, India.

## Authentication

For data shown in Fig 4b, c, f, g, h, i and 5g, U-2 OS and HEK293T cells were authenticated using STR profiling. For authentication, the total cellular DNA was isolated using our standard protocol. For STR profiling, total ten Genetic loci viz: TH01, D21S11, D5S818, D13S317, D7S820, D16S539, CSF1PO, AMEL, vWA and TPOX were PCR amplified by labelled primers specific to respective loci. The labelled amplicons were resolved by capillary electrophoresis and alleles were called. The sample genotypes were compared with reference STR genotypes available in ATCC®, DSMZ® and Biosample databases to authenticate sample identity and check for cross contamination. The remaining cell lines were not authenticated.

## Mycoplasma contamination

All cell-lines tested negative by PCR.

Commonly misidentified lines  
(See [ICLAC](#) register)

None of the cell lines used in this study are listed as commonly misidentified lines.