Mechanism-based traps enable protease and hydrolase substrate discovery

Activity-based probes have confirmed the presence and selective reactivity of the catalytic serine or cysteine nucleophile for many hydrolase proteins in cells. Efforts to define hydrolase specificity have captured non-covalent interactors with hydrolases and investigated the substrates that accumulate in the absence of a hydrolase or the products that accumulate in the presence of a hydrolase. Current approaches to identifying protease substrates mostly aim to either co-immunoprecipitate substrates that are non-covalently bound to catalytically inactive protease variants, or identify the peptides resulting from the action of the protease from experiments with and without the protease. In the first approach, substrates may be lost in the washing steps, and bound proteins may not be substrates. The second approach typically underestimates the number of substrates, and the cleavages identified may be indirect. The identification of intramembrane protease substrates by current approaches is particularly challenging. The methodological challenges in defining hydrolase and protease specificity mean that the substrates of many proteases remain unknown or incomplete, and many hydrolases remain orphans—with known substrates, we captured the otherwise transient thioester or acyl-enzyme intermediates—resulting from the first step of the reaction of these enzymes with their substrates—as their stable amide analogues. We demonstrated the utility of this approach for structural studies of acyl-enzyme intermediates.

Here we demonstrate that genetically encoded pc-Dap coupled to mass spectrometry provides a powerful approach for discovering hydrolase substrates in complex mixtures and in live mammalian cells. We previously demonstrated the genetically encoded, site-specific incorporation of photocaged Dap ((2S)-2-amino-3-((2-[(2-[1-(6-nitrobenzoyl)phenyl]-5-ethyl]thiophen-2-yl)carbonoyl]amino)propanoic acid) (pc-Dap) into proteins expressed in Escherichia coli. We converted pc-Dap to Dap (2,3-diaminopropionic acid) in proteins in vitro by illuminating purified proteins followed by incubation for up to 2 days at pH 8. By incubating purified proteases or thioesterases—in which we had replaced the catalytic cysteine or serine with Dap—with known substrates, we captured the otherwise transient thioester or ester intermediates—resulting from the first step of the reaction of these enzymes with their substrates—as their stable amide analogues. We demonstrated the utility of this approach for structural studies of acyl-enzyme intermediates.

Hydrolase enzymes, including proteases, are encoded by 2–3% of the genes in the human genome and 14% of these enzymes are active drug targets. However, the activities and substrate specificities of many proteases—especially those embedded in membranes—and other hydrolases remain unknown. Here we report a strategy for creating mechanism-based, light-activated protease and hydrolase substrate traps in complex mixtures and live mammalian cells. The traps capture substrates of hydrolases, which normally use a serine or cysteine nucleophile. Replacing the catalytic nucleophile with genetically encoded 2,3-diaminopropionic acid allows the first step reaction to form an acyl-enzyme intermediate in which a substrate fragment is covalently linked to the enzyme through a stable amide bond; this enables stringent purification and identification of substrates. We identify new substrates for proteases, including an intramembrane mammalian rhomboid protease RHBDL4 (refs. 1,4). We demonstrate that RHBDL4 can shed luminal fragments of endoplasmic reticulum-resident type I transmembrane proteins to the extracellular space, as well as promoting non-canonical secretion of endogenous soluble endoplasmic reticulum-resident chaperones. We also discover that the putative serine hydrolase retinoblastoma binding protein 9 (ref. 5) is an aminopeptidase with a preference for removing aromatic amino acids in human cells. Our results exemplify a powerful paradigm for identifying the substrates and activities of hydrolase enzymes.
complex mixtures. Overall, our results demonstrate that we have developed an approach to define proteolytic fragments over time upon addition of the wild-type protease (Fig. 1d, Supplementary Fig. 3, Supplementary Tables 2, 3). Eighty-six percent of the candidates showed HtrA2 substrates by N-terminal proteomics24 (Supplementary Table 1). This approach identified the majority of the proteins previously identified as 17 times more potential HtrA2 substrates than previous work, and this respect to controls (Fig. 1c, Supplementary Fig. 2). Our study identified 37 proteins were significantly enriched (minimum fold change (S306Dap)–HA–Strep and its conjugates were enriched from cell lysate with anti-HA beads and detected with an anti-Strep antibody. Control experiments were performed with wild-type (WT) HtrA2 and the catalytically inactive S306A mutant. Input: HtrA2 variants in cell lysates before incubation. b, Venn diagram showing the number of proteins identified in HtrA2(S306Dap) elution compared with controls. Proteins identified in at least two of the three replicates were considered as positively identified. c, Volcano plot based on label-free quantification (LFQ) values for the proteins identified in HtrA2(S306Dap) and wild-type HtrA2 samples. The black line represents the cut-off curve for significance (S0 = 1, FDR < 0.01). Each data point is calculated in Perseus using n = 4 for each HtrA2 variant. The dot representing ornithine aminotransferase (OAT) is labelled in red. d, Wild-type HtrA2 or HtrA2(S306A) (1 μM) was added into Expi293 cell lysate and incubated for the indicated time at 37 °C. Red arrowhead, full-length OAT; blue arrowhead, wild-type HtrA2-dependent proteolytic fragments. GAPDH was used as a loading control. The experiment in a was performed in biological triplicate, and the experiment in d was performed in two biological replicates, both with similar results. For gel source data, see Supplementary Fig. 1.

Trapping substrates in live human cells

Next, we set out to extend our approach to capturing and identifying protease substrates in human cells (Extended Data Fig. 4a). It was unclear—on the basis of previous work1—whether pc-Dap could be converted to Dap in live cells to activate protease substrate traps. We set out to: encode pc-Dap in human cells, investigate the activation of a protease substrate trap through deprotection of pc-Dap to Dap in human cells, and demonstrate the covalent capture of protease substrates in human cells.

We demonstrated the efficient, site-specific, genetically directed incorporation of pc-Dap into proteins in human cells (Supplementary Figs. 4, 5). We produced TEV protease in which the catalytic cysteine was replaced by pc-Dap (TEV(C151pc-Dap)) in human cells (Supplementary Fig. 6). To deprotect pc-Dap we illuminated cells25,26 expressing TEV(C151pc-Dap) for 2 min. Mass spectrometry revealed that more than 60% of the TEV purified from cells immediately after illumination was fully deprotected to TEV(C151Dap); the remaining 40% of TEV contained the deprotection intermediate. Six hours after illumination, we detected only TEV(C151Dap) (Fig. 2a); this demonstrated that deprotection, from pc-Dap to Dap—to activate the protease substrate trap—proceeds rapidly in live cells.

To demonstrate that we can capture protease substrates in human cells, we co-expressed TEV(C151pc-Dap) and GFP-s (Fig. 2b). We illuminated HEK293T cells to generate TEV(C151Dap) and followed the formation of the TEV–GFP conjugate in cells by immunoblotting (Fig. 2c, Supplementary Fig. 7). The conjugate was observed 30 min...
Intramembrane proteases have diverse and important roles in biological regulation\footnote{32,38} and have been implicated in various diseases, including Alzheimer's disease\footnote{30}. However, defining the physiological substrates of intramembrane proteases in mammalian cells has proved exceptionally challenging\footnote{28,30}. Rhomboid proteases are an important class of intramembrane protease that use a catalytic serine for catalysis\footnote{32}. We focused on RHBDL4, a rhomboid protease located in the ER membrane of mammalian cells, which has been associated with multiple cellular pathways\footnote{31,32}, and potentially Alzheimer disease pathology\footnote{32}. However, physiologically relevant substrates of RHBDL4 remain essentially unknown. We set out to: encode pc-Dap in place of the catalytic serine in RHBDL4, activate the rhomboid protease substrate trap by illuminating cells, and demonstrate the covalent capture and identification of substrates for RHBDL4.

We produced RHBDL4(S144pc-Dap) at a similar level to a wild-type control and, confirmed that the protein localized predominantly in the ER membrane (Supplementary Fig. 9), as expected\footnote{35}. We co-expressed a model substrate pTa, which is efficiently cleaved by wild-type RHBDL4 at multiple sites\footnote{4} (Extended Data Fig. 5a, b). We illuminated cells containing RHBDL4(S144pc-Dap) and pTa to activate the protease substrate trap and followed conjugate formation by immunoblot against the N-terminal Flag tag on pTa after affinity purification using the twin-Strep-tag on RHBDL4 (Extended Data Fig. 5c). We detected Dap-specific conjugates 15 min after illumination of cells expressing RHBDL4(S144pc-Dap); these conjugates rapidly accumulated within 4 h (Extended Data Fig. 5d, Supplementary Fig. 10). Overall, these experiments demonstrated that we can express and optically activate a protease substrate trap for RHBDL4, and that the trap efficiently captures its model substrate.

Identifying RHBDL4 substrates

To identify RHBDL4 substrates, we illuminated human cells producing RHBDL4(S144pc-Dap). We collected the membrane fraction and purified the protease conjugates by Strep-tag affinity enrichment. Anti-Strep analysis of the affinity-enriched samples showed several bands of higher apparent molecular mass in the RHBDL4(S144Dap) experiments that were not observed in control experiments (Fig. 3a); this suggested that the Dap-containing protease had covalently captured RHBDL4 substrates.

Proteomic analysis identified 43 potential RHBDL4 substrates; 24 of these proteins were detected only in the RHBDL4(S144Dap) elution (Fig. 3b) and 19 of them were significantly enriched in the RHBDL4(S144Dap) samples relative to controls (Fig. 3c, d). We note that proteins enriched or uniquely identified in both the RHBDL4(S144Dap) and RHBDL4(S144A) samples with respect to the wild-type control may also be substrates (Supplementary Data File 2), but these were not investigated in our subsequent analysis. Twenty-five of the candidate substrates are ER-associated proteins, including 15 ER-resident proteins (5 of which are transmembrane proteins), 2 nuclear transmembrane proteins, and 8 proteins in secretory pathways (Supplementary Table 4). We also identified non-ER-associated proteins, primarily nuclear proteins (Supplementary Fig. II, Supplementary Note I). We investigated the RHBDL4-mediated cleavage of a subset of the ER-associated candidate substrates.

We first validated transmembrane protein candidates—the protein class that is conventionally considered as rhomboid protease substrates—using a cell-based rhomboid gain-of-function cleavage assay\footnote{4}. Over-expressed wild-type RHBDL4, but not the catalytically inactive S144A mutant, cleaved all tested transmembrane proteins identified by Dap conjugation (Fig. 3e, Extended Data Fig. 6a–c). RHBDL4 cleaved CCDC47 (ref. \footnote{32}) at three positions in the luminal domain, and the ectodomains resulting from proteolysis were secreted into extracellular media (Fig. 3e and Extended Data Fig. 6d, e). In addition, endogenous CCDC47 was cleaved by exogenously expressed wild-type RHBDL4 (Extended Data Fig. 6f). Our data collectively demonstrated that RHBDL4 can proteolyze the transmembrane proteins identified by our approach and, upon cleavage by RHBDL4, luminal fragments of CCDC47 are released into the extracellular medium.

Trapping a model substrate of RHBDL4

Intramembrane proteases have diverse and important roles in biological regulation\footnote{32} and can act as cell adhesion molecules\footnote{31,32}. However, defining the physiological substrates of RHBDL4 remains challenging. To address this, we developed a model system in which we expressed a twin-Strep-tag on RHBDL4 (Extended Data Fig. 5c). We detected Dap-specific conjugates 15 min after illumination of cells expressing RHBDL4(S144pc-Dap); these conjugates rapidly accumulated within 4 h (Extended Data Fig. 5d, Supplementary Fig. 10). Overall, these experiments demonstrated that we can express and optically activate a protease substrate trap for RHBDL4, and that the trap efficiently captures its model substrate.
Soluble proteins have not been reported as endogenous RHBDL4 substrates. However, a substantial percentage (67%) of the ER-resident proteins that we identified are soluble; the majority of these proteins (90%) are chaperones (Supplementary Table 4). We therefore focussed on further investigating the cleavage of ER soluble chaperones by RHBDL4.

We identified binding immunoglobulin protein (Bip) (encoded by HSPA5) as a potential substrate using our approach (Fig. 3c, d). We directly visualized the conjugates of RHBDL4 (Dap) and endogenous Bip in human cells (Fig. 3f, Supplementary Fig. 12). The RHBDL4 cleavage assay confirmed that Bip can be cleaved at its interdomain linker region and C-terminus by wild-type RHBDL4. The N-terminal fragments generated by these cleavages—which do not possess the ER retention motif—a部分 were secreted into the medium (Fig. 3g and Supplementary Fig. 13). Endogenous Bip was cleaved by exogenously expressed WT RHBDL4, and secretion of the proteolytic fragment into the medium was Brefeldin A (BFA) sensitive—cleavage by exogenously expressed WT RHBDL4, and secretion of the proteolytic fragment into the medium was Brefeldin A (BFA) sensitive.

Collectively, our results demonstrate that we can capture and identify substrates for an intramembrane protease. Moreover, we have discovered that RHBDL4 can act as a non-canonical secretase. Unlike conventional secretases—which cleave transmembrane substrates in transmembrane domains or juxtamembrane domains to release ectodomains—cleavage by RHBDL4 has the effect of removing the C-terminal ER-retention motif from a proportion of physiological
Determination of enzymatic function of RBBP9

Many serine hydrolases have been defined on the basis of their reactivity with activity-based probes, but their enzymatic function remains unknown. Defining the activity of these orphan hydrolases remains an outstanding challenge. We set out to address this challenge using Dap-mediated substrate trapping. We focussed on RBBP9, a tumour-associated putative serine hydrolase, which possesses a similar α/β hydrolase fold to RBBP9 and removes N-terminal hydrophobic residues—especially aromatic amino acids—from peptide sequences (Pept(Dap-X)).

Determining enzymatic function of RBBP9

To investigate the aminopeptidase activity of RBBP9, we performed a fluorescence-based hydrolysis assay for 19 amino acid substrate molecules, allowing the release of their N-terminal proteolytic fragments into the extracellular space.

RBBP9 is an aromatic aminopeptidase

To investigate the aminopeptidase activity of RBBP9, we performed a fluorescence-based hydrolysis assay for 19 amino acid...
7-amino-4-methylcoumarin (AA–AMC) compounds (Fig. 5a, Supplementary Fig. 18). RBBP9 showed a clear preference for hydrolysing aromatic residues, especially Phe and Tyr; this is generally consistent with the mass adducts identified on Dap from live cells. Moreover, RBBP9-mediated hydrolysis required a free α-amine, as acetylated Met–AMC (AcMet–AMC) and the dipeptide Glu-Phe–AMC were not hydrolysed; this explains why the aminopeptidase activity was not detected even though a subset of protease substrates (for example, Pro–pNA and Suc-Phe–pNA) were screened in previous studies40.

The aminopeptidase activity and selectivity of RBBP9 was further validated on peptides. As expected, wild-type RBBP9 but not the catalytically inactive S75A mutant removed the first Phe (Phe1) from nociceptin (Fig. 5b); this explains why a free α-amine is required on the N terminus of RBBP9 substrates. Overall, our results demonstrate that RBBP9 is an aminopeptidase with a preference for removing aromatic residues from the N terminus in human cells. To our knowledge, this is the first reported aminopeptidase in mammals that uses a catalytic serine nucleophile to remove the N-terminal amino acid from polypeptides. Future work should aim to identify the C-terminal portion of RBBP9 substrates and understand how the hydrolase activity we have discovered relates to tumour cell proliferation.

Discussion

We have demonstrated that adding a Dap-containing protease to a complex mixture facilitates discovery of protease substrates. Extensions of this approach should facilitate substrate discovery in systems where genetic manipulation is challenging, including primary tissue samples. We have demonstrated that we can directly express and rapidly optically activate hydrolase substrate traps in live mammalian cells. As the genetic code expansion methods used to express caged hydrolase substrate traps have now been developed in several model organisms41,44, including mice, future work may extend our approach to diverse physiological settings.

Hydrolases have arisen independently multiple times in evolution45,46. Proteases that proceed through an acyl-enzyme intermediate naturally divide into two classes on the basis of the stereochemistry of nucleophilic attack45. We have exemplified our approach for serine and cysteine proteases from both mechanistic classes. Proteases have also been classified into clans that have a common ancestry, as identified by structural homology47. We have exemplified our approach for cysteine proteases from the major clan within animals (Clan PA, 66%) and viruses (Clan CA and PA, 72%), and for the major serine proteases and hydrolase clans within animals (Clans PA and SC, 70%) and viruses (Clan PA, 48%)48 (Extended Data Fig. 9); we have also exemplified our approach for both soluble and intramembrane proteases, and provided biological insights (Supplementary Note 2). Thus, our results cover many classes of protease reaction that proceed through an acyl-enzyme intermediate and the majority of protease structural classes; this suggests that our approach will be broadly applicable.

Finally, we have demonstrated the utility of combining hydrolase substrate traps with the direct identification of Dap conjugates to define the molecular function of an orphan hydrolase. We anticipate that future work will extend the approaches we have developed to identify the activities and substrates of many other hydrolases.

Online content

Any methods, additional references, Nature Research 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/s41586-022-04414-9.
31. Urban, S. E., Boeyo, M. Activity-based profiling of proteases. Annu. Rev. Biochem. **83**, 249–273 (2014).
32. De Cesare, V. et al. Deubiquitinating enzyme amino acid profiling reveals a class of ubiquitin esterases. Proc. Natl Acad. Sci. USA **118**, 20206947118 (2021).
33. Lopež-Otin, C. & Overall, C. M. Protease degradomics: a new challenge for proteomics. Nat. Rev. Mol. Cell Biol. **3**, 509–519 (2002).
34. Rei Liao, J. Y. & van Wijk, K. J. Discovery of AAA+ protease substrates through trapping approaches. Trends Biochem. Sci. **44**, 528–545 (2019).
35. Saita, S. et al. PARL mediates Smac proteolytic maturation in mitochondria to promote apoptosis. Nat. Cell Biol. **19**, 318–328 (2017).
36. Knopf, J. D. et al. Intramembrane protease RHBDL4 cleaves oligosaccharyltransferase subunits to target them for ER-associated degradation. J. Cell Sci. **133**, jcs245790 (2020).
37. Dix, M. M., Simon, G. M. & Cravatt, B. F. Global mapping of the topography and magnitude of proteolytic events in apoptosis. Cell **134**, 679–691 (2008).
38. Zhou, J. et al. Deep profiling of protease substrate specificity enabled by dual random and scanned human protease substrate phage libraries. Proc. Natl Acad. Sci. USA **117**, 25464–25475 (2020).
39. Luo, S. Y., Araya, L. E. & Julien, O. Protease substrate identification using N-terminomics. ACS Chem. Biol. **14**, 2361–2371 (2019).
40. O’Brien, R. J. & Wong, P. C. Amyloid precursor protein processing and Alzheimer’s disease. Annu. Rev. Neurosci. **34**, 185–204 (2011).
41. Crawford, E. D. & Wells, J. A. Caspase substrates and cellular remodeling. Annu. Rev. Biochem. **80**, 1055–1087 (2011).
42. Wolfe, M. S. Intramembrane proteolysis. Chem. Rev. **109**, 1599–1612 (2009).
43. Beard, H. A., Barniol-Xicota, M., Yang, J. & Verhelst, S. H. L. Discovery of cellular roles of intramembrane proteases. ACS Chem. Biol. **14**, 2372–2388 (2019).
44. Tint, N. & Hilvert, D. Trapping transient protein species by genetic code expansion. ChemBioChem **22**, 92–94 (2021).
45. Suzuki, Y. et al. A serine protease, HtrA2, is released from the mitochondria and interacts with XIAP, inducing cell death. Mol. Cell. **8**, 613–621 (2001).
46. Chao, J. R. et al. Haax1-mediated processing of HtrA2 by Parl allows survival of lymphocytes and neurons. Nature **462**, 98–102 (2008).
47. Martins, L. M. et al. Binding specificity and regulation of the serine protease and PDZ domains of HtrA2/Omi. J. Biol. Chem. **278**, 48417–49427 (2003).
48. Vande Walle, L. et al. Proteome-wide identification of HtrA2/Omi substrates. J. Proteome Res. **6**, 1006–1015 (2007).
49. Courtney, T. & Deiters, A. Recent advances in the optical control of protein function through genetic code expansion. Curr. Opin. Chem. Biol. **46**, 99–107 (2018).
50. Wang, J. et al. Time-resolved protein activation by proximal decaging in living systems. Nature **569**, 509–513 (2019).
51. Brown, M. S., Ye, J., Rawson, R. B. & Goldstein, J. L. Regulated intramembrane proteolysis: a control mechanism conserved from bacteria to humans. Cell **100**, 391–398 (2000).
52. Wolfe, V. S. & Kopan, R. Intramembrane proteolysis: theme and variations. Science **305**, 1115–1123 (2004).
53. Muller, S. A., Sciblarb, S. D. & Lichtenhaler, S. F. Proteomic substrate identification for membrane proteases in the brain. Front. Mol. Neurosci. **9**, 96 (2016).
54. Dickey, S. W., Baker, R. F., Cho, S. & Urban, S. Proteolysis inside the membrane is a rate-governed reaction not driven by substrate affinity. Cell **156**, 1270–1281 (2013).
55. Urban, S., Lee, J. R. & Freeman, M. Drosophila hrombid-1 defines a family of putative intramembrane serine proteases. Cell **107**, 173–182 (2001).
**Methods**

**Plasmid construction**

Standard molecular biology techniques, including PCR, restriction cloning, Gibson assembly, Golden gate assembly, and quik-change mutagenesis were applied to assemble plasmids. To generate plasmids for protein expression in *E. coli*, the DNA fragment encoding human HtrA2 (134–458), RBP9 or Scov2-PLpro was synthesized as a double stranded DNA (Integrated DNA Technologies (IDT)), and UL36USP (UL36(39–285)) and UL36USP (C65S) were amplified from UL36USP containing plasmids. The encoding sequence was cloned into pHD vector with a C-terminal HA–Strep-tag for HtrA2 and RBP9, and a C-terminal twin-Strep-tag for UL36USP and Scov2-PLpro. To convert catalytic serine/cysteine to alanine or amber stop codon, site-directed mutagenesis was completed using quick-change primers (Agilent primer design). To generate vectors for protein expression in mammalian cells, DapRST2, TEV2, human RHBDL4 (ref. 49) or RBP9 encoding sequence was amplified and introduced into the previously reported pcDNA3.1 plasmid backbone with 4 × [U6-PyltRNA_{UCA}] , in which an ER leader peptide was followed by a V5-tag and BiP encoding sequence. An additional HA-tag was introduced at the C-terminus of CCDC47, or before the ER retention motif sequence of ERP44. The DNA fragments encoding MFN2, LEMD2, EMD, HNRPNPH1 and HNRNPNM were amplified from HEK293T cDNA and cloned into a previously reported pcDNA3.1-based BiP expressing plasmid, which contains an ER leader peptide followed by a V5-tag and BiP encoding sequence. After transfection to achieve pc-Dap incorporation. Cells were incubated at room temperature for 10 min before adding to the cell culture.

**Western blot**

Samples were separated by SDS–PAGE (note that NuPAGE 4–12%, 10% Bis-Tris or 3–8% Tris-Acetate gels running in MES or MOPS buffer were applied to achieve the optimal separation of proteins (protein fragments) of interest) and transferred to polyvinylidene difluoride (PVDF) membrane by iBlot 2 dry blotting system (Thermo Fisher Scientific). Membrane was blocked by Odyssey blocking buffer in PBS (catalogue (cat.) no. 927-40000, Li-Cor) at room temperature for 30 min. Membrane was incubated in primary antibody solution (dilution according to manufacturer’s instructions in Odyssey T20 (PBS) antibody diluent (927-75001, Li-Cor)) at 4 °C overnight. All incubations were carried out on a platform shaker. The membrane was washed three times with PBST (PBS supplemented with 0.1% Tween-20 (v/v)), and incubated with the secondary antibody solution (I:15,000 (v/v) in PBS blocking buffer supplemented with 0.2% Tween-20 (v/v), and 0.01% SDS) at room temperature for 1 h. After washing 3 times with PBST and once with PBS, the immunoreactive proteins were visualized by the Odyssey CLx imaging system (Li-Cor) by scanning at 700 nm and/or 800 nm channels. Revert 700 Total Protein Stain (926–11015, Li-Cor) was used for total protein staining. The data were analysed by Image Studio Lite (version 5.2.5). For primary and secondary antibodies used in this study, see ‘Antibodies’ in Supplementary Methods.

**Deprotection of pc-Dap containing proteins in buffer**

To activate protease(pc-Dap), proteins were illuminated (365 nm, 4 mW cm⁻²) for 1 min in Tris buffer (5 mM DTT, pH 8.0) and incubated at 4 °C or 37 °C overnight to generate protease(Dap). MIC-LED-365 (500 mA, Prizmatix collimated modular Mic-LED light source, Supplementary Fig. 2) was used for illuminating proteins in solution. This apparatus was also used for illuminating suspension cells (Expi293 cells) in tissue culture hood.

**HtrA2 substrate trapping in cell lysate**

Thirty micrograms HtrA2–HA–Strep variant (wild type, Ala or Dap) was added to 1 ml of Expi293 cell lysate (3 μg ml⁻¹) and incubated at 30 °C for 3 h. The reaction was shaken 10 s every 10 min. Fifty microlitres of anti-HA agarose slurry (A2095, Merck) was added to the reaction and mixed at 4 °C on an end-over-end rocker for 2 h. The mixture was transferred to a Bio-spin column. The resin was washed with RIPA buffer 3 times and PBST buffer 3 times using a vacuum pump, followed by centrifugation at 5,000g for min to remove the residual buffer. Then, beads were incubated in 100 μl of 1× LDS loading buffer and boiled at 95 °C for 5 min. Twenty microlitres of eluate was analysed by SDS–PAGE or western blot. Twenty microlitres of eluate was separated in a Bolt 10% Bis-Tris Plus gel for 3 min at 200 V. Gel slices containing all proteins were cut and analysed by LC–MS/MS as described in ‘Electrospray ionization tandem mass spectrometry’.

**Validation of HtrA2 substrates in cell lysate**

Wild-type HtrA2 or HtrA2(S306A) (1 μM) was added to 1.2 ml Expi293 cell lysate. At indicated time points, 300 μl of reaction was quenched by mixing with 100 μl of 1× LDS loading buffer and boiled at 95 °C for 5 min. The samples were analysed by western blot with primary antibodies listed in Supplementary Tables 2, 3. GAPDH was used as a loading control.

**Incorporation of pc-Dap in HEK293T cells**

HEK293T cells were purchased from European Collection of Cell Cultures (authenticated by STR DNA profiling) and were tested negative for *Mycoplasma* contamination.

HEK293T cells were cultured in Dulbecco’s Modified Eagle Medium (DMEM) (Gibco) supplemented with 10% fetal bovine serum (FBS) (Gibco) and Penicillin-Streptomycin (Pen/Strep, 100 IU ml⁻¹ penicillin and 100 μg ml⁻¹ streptomycin) at 37 °C in a humidified incubator supplied with 5% CO₂. Cells were passaged every 2–3 days by detaching with trypsin–EDTA solution, resuspended in DMEM with 10% FBS, and seeded into cell culture flasks.

For transfection in a 24-well plate: 0.75 μl of Lipofectamine 3000 (Thermo Fisher) was diluted in 25 μl Opti-MEM (Gibco) and vortexed briefly. DNA solution was prepared by mixing 500 ng DNA mixture (substrate:DapRST:protease, 1:1:3 or empty vector:DapRST:Protease, 1:1:3) in 25 μl Opti-MEM, followed by adding 1 μl of P3000 reagent. Then, diluted Lipofectamine was added to DNA solution (1:1 v/v). The mixture was incubated at room temperature for 10 min and the DNA–lipid complexes were added to cells. Indicated concentrations in figure legends (or 0.3 mM) of pc-Dap was added to the culture medium 30 min after transfection to achieve pc-Dap incorporation. Cells were incubated at 37 °C for 40–48 h before further analysis.

**Incorporation of pc-Dap in Exp293 cells**

Exp293 cells were purchased from Thermo Fisher (authenticated by STR DNA profiling) and were tested negative for *Mycoplasma* contamination.

Exp293 cells were cultured in Exp293 media (Gibco) and shaken at 125 rpm in incubator at 37 °C with 8% CO₂. Cells were passaged every 2–3 days, starting with the cell density around 0.5 × 10⁶ cells per ml. Transfection was performed at cell density around 2.5 × 10⁶ cells per ml. Transfection of 100 ml Exp293 cells: 300 μl of polyethylenimine molecular mass 40,000 (PEI, 1 mg ml⁻¹). Polyscienes) was diluted in 3.3 ml Exp293 medium. 100 μg DNA mixture (substrate:DapRST:protease, 1:1:3 or empty vector:DapRST:protease, 1:1:3) was diluted in 3.3 ml Exp293 media. Diluted DNA and PEI solution were mixed and incubated at room temperature for 15 min before adding to the cell culture. 0.5 mM (or indicated concentrations in figure legends) of pc-Dap was added to incubation medium.
added 30 min after transfection for pc-Dap incorporation. Forty to forty-eight hours after transfection, the cells were collected and photo-activated for further analysis.

**Photoactivation of protease(pc-Dap) and substrate trapping in mammalian cells**

Forty to forty-eight hours after transfection, cell culture medium was replaced with fresh medium, and cells were illuminated for 2 min. The apparatus for illuminating adherent mammalian cells was built in-house (Supplementary Fig. 7). LuxiGen 365 nm UV LED Emitter (LZ4-04UVOR-0000, Mouser Electronics) was used for illumination. The UV intensity at the well plate was set at 4 mW cm⁻². After illumination, cells were incubated at 37 °C for indicated period of time (Proteasome inhibitor MG132 (2 μM) was added if needed). For adherent cells, at each time point, cells in a 6-well plate were washed with PBS and lysed in 400 μl RIPA lysis buffer (89900, Thermo) supplemented with Halt Protease Inhibitor Cocktail (78429, Thermo Fisher) and the Universal Nuclease at 4 °C. Cell lysates were centrifuged at 21,000 g for 5 min. The supernatant was flash frozen and stored at −80 °C. Then, cell pellets were lysed in 1 ml RIPA lysis buffer supplemented with protease inhibitors and the Universal Nuclease at 4 °C. Cell lysates were centrifuged at 21,000 g for 5 min. The cleared lysates were used for SDS–PAGE and western blot analysis or affinity enrichment by MagStrep type3 XT beads (2-4090-002, IBA).

**Trapping endogenous substrates to RHBDL4(S144Dap)**

RHBDL4 variants were expressed in Expi293 cells as described in ‘Incorporation of pc-Dap in Expi293 cells’. Forty hours after transfection, 50 ml cell culture was resuspended in fresh Expi293 media and illuminated (365 nm, 4 mW cm⁻²) for 2 min. Cells were incubated at 37 °C for 4 h in the presence of 2 μM MG132 before collection. After pelleting, cells were resuspended in HEPES buffer (50 mM HEPES, pH 7.4, 150 mM NaCl, 1 mM MgCl₂, 5% glycerol, 1 mM DTT) supplemented with protease inhibitors and the Universal Nuclease. The suspension was cleared by centrifugation at 21,000 g for 5 min and the supernatant was flash frozen and stored at −80 °C. For suspension cells, at each indicated time point, 5 ml cell culture was resuspended in fresh Expi293 media and illuminated (365 nm, 4 mW cm⁻²) for 2 min. Cells were incubated at 37 °C for 3 h. Then, cell pellets were lysed in 1 ml RIPA lysis buffer supplemented with protease inhibitors and the Universal Nuclease at 4 °C. Cell lysates were centrifuged at 21,000 g for 5 min. The cleared lysates were used for SDS–PAGE and western blot analysis or affinity enrichment by MagStrep type3 XT beads.

**RHBDL4 cleavage assay**

Empty vector, wild-type (WT) RHBDL4 or RHBDL4(S144A) plasmid was co-transfected with candidate substrate containing plasmid or empty vector (for endogenous substrates) into HEK293T or Expi293 cells. The amount of candidate substrate-containing plasmid was optimized for expression level. Forty to forty-eight hours after transfection, the cells were collected and lysed in RIPA buffer supplemented with protease inhibitors and the Universal Nuclease. The lysate was cleared and analysed by western blot.

To analyse proteins in the extracellular medium, FBS-containing medium for HEK293T cells was replaced with hybridoma serum free medium (12045076, Thermo Fisher) 24 h before collection. Expi293 medium, which is serum-free and protein-free, can be directly collected for further analysis. The medium was collected and filtered through a 0.22 μm polyethersulfone membrane. To obtain total proteins in the medium, 1/10 volume of 100% ice-cold TCA solution (T0699, Sigma) was added at 4 °C for protein precipitation. To obtain proteins in the supernatant, the medium was centrifuged at 200,000 g for 1 h to separate supernatant from microvesicles. The SN was collected and added 1/10 volume of ice-cold TCA to precipitate proteins. After centrifuging at 21,000 g for 10 min, the precipitate was washed once with acetone, and dissolved in 1× LDS loading buffer. The microvesicles pellet after ultra-centrifugation was washed with PBS once and dissolved in equal volume of 1× LDS loading buffer.

Deglycosylation was performed by adding 1/10 volume of 10% NP40 and PNGase (P07045, NEB) or DeGlycosylation mix II (P6044S, NEB) to proteins dissolved in the 1× LDS loading buffer. The reaction was incubated at 37 °C for 1 h before analysis.

**Brefeldin A inhibitory assay**

Twenty-four hours after transfection, Expi293 cells were split into two halves treated separately with DMSO or BFA (5 μg ml⁻¹). The BFA treatment was performed in two ways: (1) BFA was directly added into medium culture; (2) the medium culture was replaced with fresh medium supplemented with BFA. Sixteen hours after BFA treatment, the cells and extracellular medium were collected and analysed as described in ‘RHBDL4 cleavage assay’.

**Knockout of RHBDL4**

HCT116 cells were purchased from American Type Culture Collection (authenticated by STR DNA profiling) and were tested negative for Mycoplasma contamination.

HCT116 cells were cultured in McCoy’s 5A (modified) Media (Gibco) supplemented with 10% fetal bovine serum (FBS) (Gibco) and Pen/Strep at 37 °C in humidified incubator with 5% CO₂. Cells were passaged every 2–3 days.

HCT116 cells in 6-well plates were transfected by Lipofectamine LTX (15338100, Thermo Fisher) with 2.5 μg of pX330-puro plasmid containing gRNA (5′-TCCAGTAAGTACAGAAAATG-3′) and Cas9 for RHBDL4 knockout. Twenty-four hours after transfection, cells were trypsinized and plated in a 10 cm petri dish. After 24 h, the cells were treated with puromycin (1 μg ml⁻¹). The puromycin selection stopped after 48 h. Cells were trypsinized and limited dilution was performed to generate single clones, which were expanded and analysed by western blot (anti-RHBDL4) and genotyped by sequencing the genomic DNA region targeted by the gRNA.

To detect the proteolytic fragments from endogenous substrates generated by endogenous RHBDL4, 10 million wild-type or RHBDL4 knockout HCT116 cells were cultured in hybridoma serum free medium for 40 h. The medium was collected, filtered and concentrated by a 30 kDa cut-off concentrator. Proteins were precipitated by TCA and dissolved in 1× LDS loading buffer for immunoblotting analysis.

**Trapping endogenous substrates to RBBP9(S75Dap)**

To identify X attached to Pept(Dap), RBBP9 variants were expressed in HEK293T cells as described in ‘Incorporation of pc-Dap in HEK293T cells’. pc-Dap (0.1 mM) was added to cells to produce RBBP9(S75pc-Dap). To characterize the entire masses of RBBP9 variants, RBBP9 variants were produced in 100 ml Expi293 cells as described in ‘Incorporation of pc-Dap in Expi293 cells’. RBBP9(S75pc-Dap) was expressed in the presence of 0.5 mM pc-Dap. 40 h after transfection, cells were illuminated (365 nm, 4 mW cm⁻²) for 2 min, and incubated at 37 °C for 3 h. Cells were then collected and lysed in Tris buffer (50 mM Tris, pH 8.0, 150 mM NaCl, 1 mM EDTA and Universal Nuclease) by sonication. Note that protease inhibitors were not added in lysis buffer. The lysate was cleared by centrifuging at 21,000 g for 20 min. RBBP9 species in the supernatant were enriched using MagStrep type3 XT beads. Proteins attached to beads were eluted in 50 μM Biotin in Tris buffer for mass characterization.
Aminopeptidase assay of RBBP9

**Fluorescence-based assay.** Two micromolar RBBP9 was incubated with each AA–AMC over a range of different substrate concentrations in Tris buffer (100 mM Tris, 150 mM NaCl, pH 7.3). Fluorescence intensity (due to the release of the AMC fluorophore by hydrolysis of AA–AMC by RBBP9) was measured every 20 s over a 10-min period (MARS Data Analysis Software (version 3.20 R2)). For each substrate, the rate of fluorescence increase was converted to rate of product formation using standard curves. At substrate concentrations of greater than 10 μM, intermolecular quenching of AMC fluorescence by AA–AMC was found to be significant. Therefore, for all substrates other than Phe-AMC and Tyr-AMC, AA-AMC concentrations between 0 and 4 μM were used, and pseudo-first-order kinetics were employed to calculate specificity constants. For Phe-AMC and Tyr-AMC, which showed significantly faster rates of hydrolysis when compared to the other substrates, a concentration range of 0 to 160 μM was used and converted rates were fitted to Michaelis–Menten kinetics in order to obtain specificity constants.

**Peptide-based assay.** Peptides (100 μM) were dissolved in Tris buffer (100 mM Tris pH 7.4, 150 mM NaCl, 1 mM EDTA). Two micromolar wild-type RBBP9 or RBBP9(S75A) was added to start the hydrolysis reaction. The reaction was stopped by quenching with acetic acid and monitored by mass spectrometry. Selected ion mass (SIM) mode was applied for detection of peptide substrates and the desired products.

**Protein crystallization and data collection**

Human RBBP9 with a C-terminal His-tag<sup>60</sup> (LEHHHHHHH) was expressed in BL21(DE3) cells and purified by a two-step protocol consisting of HisTrap enrichment and gel filtration (Superdex 75) chromatography. Pure fractions of RBBP9 (> 98% purity determined by SDS–PAGE) were concentrated with a 10 kD MWCO Vivapin 20 concentrator (Sartorius) to 10 mg ml<sup>−1</sup> in buffer containing 10 mM Tris (pH 7.5), 100 mM NaCl, 5 mM DTT and 5 mM Phe. Prior to crystallization, samples were cleared by centrifugation for 15 min at 10,000 g. Crystallization trials with multiple commercial crystallization kits were performed in 96-well sitting-drop vapor diffusion plates (Molecular Dimensions) at 18 °C and set up with a mosquito HTS robot (TTP Labtech). Drop ratios of 0.2 μl protein solution plus 0.2 μl reservoir solution were used for coarse and fine screening. Initial hits were obtained under multiple conditions and required no further optimization. Data was collected from crystals collected from following conditions: 30% w/v PEG 4K, 0.1 M MES sodium salt, pH 6.5.

To ensure cryo-protection, crystal-containing drops were mixed with 25% glycerol in reservoir solution prior to picking and flash frozen in liquid nitrogen. Diffraction data was collected at the Diamond Light Source (DLS, UK) on beamline I04. Datasets were auto-processed with XIA2 DIALS (version 0.7.90), scaled using Aimless and Refmac5 (version 5.8.0258) in the CCP4 suite (version 7.0.078) of programs. Structure refinement and manual model building were performed with Refmac5 and COOT (version 0.8.9.2). Colour figures were prepared with PyMol (version 2.5).

**Mass characterization**

**Electrospray ionization mass spectrometry.** Mass spectra of all protein samples were acquired on an Agilent 1200 LC-MS system equipped with a 6130 Quadrupole spectrometer. A Phenomenex Jupiter C4 column (150 × 2 mm, 5 μm) was used to elute proteins. Buffer A (0.2% formic acid in H<sub>2</sub>O) and buffer B (0.2% formic acid in acetonitrile) was used for RP-HPLC. Mass spectra were acquired with the pyOpenMS package (version 2.5). In brief, collected spectra were centroided and all MS2 spectra with a precursor mass lower than that of the unconjugated Dap-containing tryptic peptide from RBBP9 (Pept(Dap)) were filtered out. For each filtered MS2 spectrum, the ten most abundant peaks in each 100 Th mass interval were extracted. For substrate-conjugated Pept(Dap), (Pept(Dap-X)). This list contained substrate-conjugated Pept(Dap), (Pept(Dap-X)). For each filtered MS2 spectrum, the ten most abundant peaks in each 100 Th mass interval were extracted. For substrate-conjugated Pept(Dap), (Pept(Dap-X)). This list contained substrate-conjugated Pept(Dap), (Pept(Dap-X)).

**LC–MS/MS data analysis by Venn diagram.** LC–MS/MS data were searched against an in-house protein sequence database containing Swiss-Prot and the protein constructs specific to the experiment, using the Mascot search engine program (Matrix Science, version 2.4). Database search parameters were set with a precursor tolerance of 5 p.p.m. and a fragment ion mass tolerance of 0.8 Da. Variable modifications for oxidized methionine, carbamidomethyl cysteine, pyrogallol acid, and deamination of glutamine/asparagine were included. MS/MS data were validated using the Scaffold program (version 5, Proteome Software Inc.).

**LC–MS/MS data analysis by volcano plot.** For quantitative analysis, MS raw files were processed by MaxQuant software (version 1.6.3.4) and searched with the embedded Andromeda search engine against the corresponding database (Uniprot). The required FDR was set to 1% or 5% at peptide and protein levels. The maximum number of allowed missed cleavages was set to two. Protein quantification was done by LFQ with default settings. The MaxQuant ProteinGroups output file was further processed with Perseus (version 1.6.2.3)<sup>15</sup>. Contaminations and reverse hits were removed by filtering. The resulting protein quantifications were log<sub>2</sub>-transformed.

**Determination of X attached to Dap in RBBP9**

LC–MS/MS files (in RAW format) were first converted to mzML format<sup>4<sup>4</sup> using ProteoWizard (version 3.0.11252)<sup>60</sup>. Data preparation and processing were then performed using custom Python (version 3.8.1) scripts written with the pyOpenMS package (version 2.4.0)<sup>69</sup>. In brief, collected spectra were centroided and all MS2 spectra with a precursor mass lower than that of the unconjugated Dap-containing tryptic peptide from RBBP9 (Pept(Dap)) were filtered out. For each filtered MS2 spectrum, the ten most abundant peaks in each 100 Th mass interval were extracted.

Based on the peptide sequence of Pept(Dap) and the precursor mass for each MS2 spectrum, a list of theoretical ion masses was calculated; these corresponded to the MS2 fragmentation of a substrate-conjugated Pept(Dap), (Pept(Dap-X)). This list contained the monocationic b- and y-ions, the dicaticonic b- and y- ions, and ions (http://www.peptidesynthetics.co.uk/tools/) with the molecular mass of non-canonical amino acids.
corresponding to water or ammonium losses from the side-chains of b- or y-ions. Peaks in the MS2 spectrum were matched against this list, and a score for this matching was calculated as previously described. This score was ten times the negative logarithm of the approximate probability that at least k out of n masses have been matched by chance, where k is the number of matches and n is the number of masses in the list.

To extract the top-scoring spectra, the family-wise error rate for the probability values was controlled at 0.05 using the Bonferroni correction. The mass difference between Pept(Dap) and the precursor ion for each Pept(Dap-X) spectrum was calculated to determine the molecular mass of each conjugate. For each mass shift, representative top-scoring spectra were manually interrogated to verify the assignment.

**Reporting summary**

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

**Data availability**

The structure of RBBP9 in complex with Phe is available in the Protein Data Bank under accession code 7OEX. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the accession number PXD030381. All other datasets and materials generated or analysed in this study are available from the corresponding authors upon reasonable request. The data used to analyse serine and cysteine proteases clones were downloaded from the MEROPS database (https://www.ebi.ac.uk/merops/). Source data are provided with this paper.

**Code availability**

The code used for RBBP9 substrate identification by proteomic analysis is available at https://doi.org/10.5281/zenodo.5768340.

48. MacDonald, C., Winstorfer, S., Pope, R. M., Wright, M. E. & Piper, R. C. Enzyme reversal to explore the function of yeast E3 ubiquitin-ligases. Traffic 18, 465–484 (2017).

49. Ikeda, K. N. & Freeman, M. Spatial proteomics reveal that the protein phosphatase PTP1B interacts with and may modify tyrosine phosphorylation of the rhomboid protease RHBDL4. J. Biol. Chem. 294, 11486–11497 (2019).

50. Schmied, W. H., Elsasser, S. J., Uttamapinant, C. & Chin, J. W. Efficient multisite unnatural amino acid incorporation in mammalian cells via optimized pyrrolysyl tRNA synthetase/tRNA expression and engineered eFlpJ. Am. Chem. Soc. 136, 15577–15583 (2014).

51. Cesaranito, F. et al. BiP/GRP78 mediates ERAD targeting of proteins produced by membrane bound ribosomes stalled at the STOP-codon. J. Mol. Biol. 431, 123–141 (2019).

52. Maule, G. et al. Allele specific repair of splicing mutations in cystic fibrosis through AsCas12a genome editing. Nat. Commun. 10, 3556 (2019).

53. Tyanova, S. et al. The Perseus computational platform for comprehensive analysis of proteomics data. Nat. Methods 13, 731–740 (2016).

54. Martens, L. et al. mzML-a community standard for mass spectrometry data. Mol. Cell Proteomics 10, R110.001533 (2011).

55. Chambers, M. C. et al. A cross-platform toolkit for mass spectrometry and proteomics. Nat. Biotechnol. 30, 918–920 (2012).

56. Rost, H. L., Schmitt, U., Aeberson, R. & Malmstrom, L. pyOpenMS: a Python-based interface to the OpenMS mass-spectrometry algorithm library. Proteomics 14, 74–77 (2014).

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

58. Kattenhorn, L. M., Korbel, G. A., Kessler, B. M., Spooner, E. & Ploegh, H. L. A deubiquitinating enzyme encoded by HSV1 belongs to a family of cysteine proteases that is conserved across the family Herpesviridae. Mol. Cell 19, 547–557 (2005).

59. Shn, D. et al. Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. Nature 587, 657–662 (2020).

60. Patchett, S. et al. A molecular sensor determines the ubiquitin substrate specificity of SARS-CoV-2 papain-like protease. Cell Rep. 36, 109754 (2021).

61. Klemm, T. et al. Mechanism and inhibition of the papain-like protease, Plpro, of SARS-CoV-2. EMBO J. 39, e106275 (2020).

62. Liu, G. et al. ISG15-dependent activation of the sensor MDAS is antagonized by the SARS-CoV-2 papain-like protease to evade host innate immunity. Nat Microbiol. 6, 467–478 (2021).

Acknowledgements This work was supported by the UK Medical Research Council (MRC; MC_U105180009 and MC_UP_A024_1008) to J.W.C.; MC_U105192713 to M. Bienz (M. Fiedler’s supervisor), and Wellcome Trust Investigator Awards (101035/Z/13/Z and 220887/Z/20/Z) to M. Freeman. S.T. was supported by a European Molecular Biology Organization Long-Term Fellowship (ALTF 825-2016). L.K. (no. 884511) and G.P. (no. 897663) were supported by Marie Skłodowska-Curie European Postdoctoral Fellowships. We thank M. K. Lemberg for providing the pTα construct; N. Gomez-Navarro for sharing the UL36USP containing plasmids; O. R. Burrowe for providing the BiP-expressing plasmid; R. Hegde and S. Munro (MRC-LMB) for helpful comments on previous versions of the manuscript; O. Persic, Y. Christova, S. Grazier, C. Morgan and M. Spinck for helpful discussions; K. Lilley for advice on RHBDL4 proteomics analysis, the FACS facility (MRC-LMB); K. Heesom, M. Skehel, F. Begum, M. Kyte and A. Howie and the MRC-LMB Mechanical Workshop for support; and B. Cravatt (Scripps Research) for suggesting RBBP9 as a target for investigation.

Author contributions S.T. and J.W.C. wrote the paper. J.W.C. set the overall direction of research. S.T. and N.H.-D. developed the hydrolyase substrate identification approaches in cell lysate (for TEV and HtrA2, and S.T. established the hydrolyase substrate discovery approaches in mammalian cells (for TEV, UL36USP, SCoV2-PLpro, RHBDL4 and RBBP9). N.H.-D. expressed and purified TEV and HtrA2 variants from E. coli. S.T. expressed and purified UL36USP and SCoV2-PLpro variants from E. coli and performed the conjugation assays with UBL-AMC molecules. S.T. performed HtrA2 substrate validation. S.T. developed the capture of RHBDL4 model substrates with contributions from L.K. S.T. performed RHBDL4 substrate capture and identification experiments. S.T., L.K. and G.P. performed RHBDL4 substrate validation. S.T. and G.P. generated the RHBDL4 knockout cell line. M. Freeman supervised L.K. and contributed to the planning and interpretation of the RHBDL4 experiments. A.T.B. developed the computational pipeline for identifying the conjugates on RBBP9(75Dap). A.T.B. and S.T. characterized the aminopeptidase activity of RBBP9. S.T. and M. Fiedler solved the structure of RBBP9 in complex with Phe. All authors provided critical feedback and helped shape the research, analysis and manuscript.

Competing interests The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41586-022-04444-9.

Correspondence and requests for materials should be addressed to Jian W. Chen or Jason W. Chen.

Peer review information Nature thanks Michael Bollong and the other, anonymous reviewers for their contribution to the peer review of this work.

Reprints and permissions information is available at http://www.nature.com/reprints.
Extended Data Fig. 1 | Strategies for Dap-mediated hydrolase substrate discovery. Genetically encoded pc-Dap in place of the catalytic serine or cysteine in the active site of a hydrolase enables the deprotection of pc-Dap to Dap and the activation of a hydrolase substrate trap that can covalently capture substrates. Purified recombinant and pre-activated Dap-containing hydrolase can be added to lysate or extract and used to capture substrates. Hydrolase containing pc-Dap can also be expressed directly in mammalian cells. The hydrolase substrate trap can then be optically activated in live cells to capture substrates. Covalent conjugates can be enriched by immunoprecipitation with stringent washing to remove non-covalent binders and the conjugates can be visualized and identified by immunoblot and mass spectrometry-based methods. Control experiments use the wild-type (WT) enzyme that does not form a stable covalent acyl-enzyme conjugate and the catalytically inactive mutant (Ala, catalytic serine/cysteine is mutated to alanine) of the hydrolase that does not react with substrates. We note that not all the Dap-containing hydrolase will necessarily be found in conjugates; in general, we expect the fraction of Dap-containing hydrolase in conjugates to be a function of hydrolase abundance, substrate abundance, effects on the rate of acyl-intermediate formation resulting from replacing the catalytic nucleophile with Dap, and the stability of the acyl-intermediate.
Extended Data Fig. 2 | TEV(C151Dap) specifically traps its substrate in Expi293 cell lysate. TEV variants (WT, Ala, Dap, 1 μM) were incubated with control GFP-c (a polypeptide “Gly-Gly-Gly-Ser-Gly-Gly-Gly-His6” was attached at the C-terminus of GFP) or substrate GFP-s (a TEV cleavage sequence “Glu-Asn-Leu-Tyr-Phe-Gln-Gly-His6” was attached at the C-terminus of GFP. The cleavage between Gln and Gly is underlined) in Expi293 cell lysate at 37 °C for 3 h. The concentration of GFP-s or GFP-c was (a) 1 μM or (b) 5 μM. Dap*: reaction of TEV(C151Dap) with GFP-s in Tris buffer. The input (reaction before IP) and Strep-tag enriched TEV species were analyzed by SDS-PAGE (Coomassie staining) and WB (anti-Strep for TEV and anti-GFP). (a) and (b) were repeated in three biological replicates with similar results. For gel source data, see Supplementary Fig. 1.
Extended Data Fig. 3 Dap-containing proteases selectively capture known substrates of their parent enzymes. (a) The change of AMC fluorescence resulting from cleavage from the C-terminus of Ubiquitin (Ub) or ubiquitin-like molecules (SUMO, NEDD8, ISG15) was followed upon mixing with UL36USP. WT UL36USP from human herpesvirus 1 specifically hydrolyses Ub-AMC, but not other ubiquitin-like protein-AMC (UBL-AMC) molecules. The deubiquitination activity of UL36USP is lost when Cys65 is mutated to Ser. This data is consistent with the previously reported specificity of UL36 USP58. (b) UL36USP(C65Dap) specifically reacts with Ub-AMC to form the UL36USP(Dap)-Ub conjugate. In contrast, no conjugates were observed between UL36USP(Dap) and SUMO1, NEDD8 or ISG15. (c) WT SCoV2-PLpro selectively hydrolyses ISG15-AMC in preference to other Ub/UBL-AMC molecules; this data is consistent with the previously reported specificity of SCoV2-PLpro59. The hydrolysis of ISG15-AMC is abrogated when the catalytic Cys111 of SCoV2-PLpro is mutated to Ala. (d) SCoV2-PLpro(C111Dap) specifically reacts with ISG15-AMC, generating the SCoV2-PLpro(Dap)-ISG15 conjugate. *: UBL-AMC independent higher MW bands resulting from PLpro(C111Dap) self-reaction. (a) and (c) were generated using n = 3 independent measurements. The line represents the means of three measurements. (b) was repeated twice and (d) was performed in three biological replicates with similar results. For gel source data, see Supplementary Fig. 1.
Extended Data Fig. 4 | UL36USP(C65Dap) and SCoV2-PLpro(C111Dap) form conjugates with endogenous ubiquitin in live cells. (a) A protease containing pc-Dap in place of its catalytic serine or cysteine is produced in mammalian cells by genetic code expansion, creating a photocaged protease trap. The trap is activated by illuminating cells. The activated trap covalently and specifically captures substrate fragments in acyl-enzyme complexes linked through stable amide bonds. (b) Immunoblotting analysis of UL36USP variants enriched from an equal number of cells after optical activation and substrate capture. The conjugates formed between UL36USP(C65Dap) and endogenous proteins were detected by an anti-Strep antibody and an anti-Ub antibody; this demonstrates that UL36USP(C65Dap) captures Ub and di-ubiquitin (Ub$_2$) in cells. Catalytically inactive UL36USP(C65S) non-covalently associated with Ub chains to co-IP them (lanes 2 and 6), while UL36USP(C65Dap) formed conjugates with endogenous Ub molecules (lanes 7–9). Input: cell lysates before IP probed with an anti-Strep antibody. These data are consistent with previous work demonstrating that UL36USP is a deubiquitinase in cells

(c) Immunoblotting analysis of SCoV2-PLpro(C111Dap) variants enriched from an equal number of cells after optical activation and substrate capture. The conjugates formed between SCoV2-PLpro(C111Dap) and endogenous proteins were detected by an anti-Strep antibody and an anti-Ub antibody; this demonstrates that SCoV2-PLpro(C111Dap) captures Ub and Ub$_2$ in cells. Catalytically inactive SCoV2-PLpro(C111A) and SCoV2-PLpro(C111pc-Dap) without illumination non-covalently associated with Ub chains to co-IP them (lanes 2, 4 and 6), while SCoV2-PLpro(C111Dap) also formed conjugates with endogenous Ub molecules (lanes 8 and 9). Input: cell lysates before IP probed with an anti-Strep antibody. Note that SCoV2-PLpro has previously been shown to cleave poly-Ub chains in cells

UL36USP is a deubiquitinase in cells.

(b) and (c) were repeated in two biological replicates with similar results. For gel source data, see Supplementary Fig. 1.
Extended Data Fig. 5 | Formation of RHBDL4(Dap)-pTα conjugates in Expi293 cells. (a) Schematic representation of the model substrate pTα. SP: signal peptide; FLAG: FLAG-tag; TMH: transmembrane helix; V5: V5-tag. Red arrows: RHBDL4 cleavage sites; black arrow: SPase cleavage site. (b) Co-expression of C-terminal Strep-tagged RHBDL4 variants and pTα in Expi293 cells for 40 h. pc-Dap was added to produce full-length RHBDL4(S144pc-Dap). Cell lysates were analysed by anti-Strep (for RHBDL4), anti-FLAG and anti-V5 antibodies. Blue triangles: N-terminal proteolytic fragments of pTα; Red triangles: C-terminal proteolytic fragments of pTα. (c) Detection of RHBDL4(Dap)-pTα conjugates after Strep-tag enrichment. Samples without illumination (lane 4) and with illumination (lanes 1–3 and 5–11) were collected at indicated time points. Input: cell lysate before IP analysed by anti-Strep and anti-FLAG antibodies. (d) The formation of conjugates was monitored for a longer period of time. RHBDL4(Dap)-pTα conjugates were gradually degraded 4 h after UV irradiation. RHBDL4 variants were also gradually degraded over time pTα*, the deglycosylated form of pTα. The experiments were repeated in two biological replicates with similar results. For gel source data, see Supplementary Fig. 1.
Extended Data Fig. 6 | RHBDL4 cleaves transmembrane substrate candidates. RHBDL4 cleavage assays for (a) MFN2, (b) EMD, (c) LEMD2 and (d) CCDC47. The putative cleavage sites – the approximate positions of which were estimated based on the MW of the proteolytic bands – are indicated by red arrows in schematic representations. HA: HA-tag; GFP: GFP-tag; SP: signal peptide; V5: V5-tag. The transmembrane helices are labeled in grey. (a) WT RHBDL4 cleaved MFN2, a multiple membrane-spanning protein resident in mitochondria or the ER. Red triangle: the C-terminal proteolytic fragment. (b) WT RHBDL4 cleaved EMD, a type II transmembrane protein in the nuclear inner membrane. Red triangle: the C-terminal proteolytic fragment. (c) WT RHBDL4 cleaved LEMD2, a multiple membrane-spanning protein in the nuclear inner membrane. Two minor cleavages (red triangles) would be in the perinuclear space, but the major cleavage site (blue triangle) in the domain resident in nucleus might be cut by the inverted RHBDL4 (see Supplementary Note 1). (d) RHBDL4 cleaved both full-length (FL) CCDC47 and CCDC47(21–135), generating the same MW N-terminal proteolytic fragment (lanes 2 and 5, blue triangle); this suggested that RHBDL4-mediated cleavage is independent of the transmembrane helix of CCDC47. Blue triangle: the N-terminal fragment; Red triangle: C-terminal proteolytic fragments. (e) Comparison of the proteolytic fragments to truncated CCDC47 standards confirmed that the cleavages are in the luminal domain of CCDC47. WT + FL: FL CCDC47 cleaved by WT RHBDL4. (f) Over-expressed WT RHBDL4 cleaved endogenous CCDC47. Red triangles: proteolytic fragments detected by an anti-CCDC47 antibody. The C-terminal proteolytic fragments of MFN2, LEMD2 and CCDC47 were probed with an anti-HA antibody, while the proteolytic fragment of EMD was detected by an anti-GFP antibody. The N-terminal proteolytic fragments were probed with an anti-V5 antibody. RHBDL4 variants were detected by an anti-Strep antibody. (a–d) were repeated in three biological replicates with similar results. (e) was performed once. (f) was repeated in two biological replicates with similar results. For gel source data, see Supplementary Fig. 1.
Extended Data Fig. 7 | RHBDL4 cleaves endogenous BiP in mammalian cells and facilitates the secretion of N-terminal proteolytic fragments. (a) BFA inhibitory assay was performed for endogenous BiP. The secretion of BiP proteolytic fragment into extracellular media was inhibited by Brefeldin A (BFA) compared to cells treated with DMSO (lanes 5 vs 2, 11 vs 8, anti-BiP). Especially when the medium was changed, no BiP proteolytic fragment was detected in the medium (lane 11, dashed red arrow). Treatment of BFA greatly increased the endogenous level of BiP (lanes 4–6 vs 1–3, anti-BiP in cell lysate). In comparison, the expression of endogenous BiP was slightly increased when WT RHBDL4 or RHBDL4(S144A) was expressed (lanes 2 and 3 vs 1, anti-BiP in cell lysate). BFA treatment or RHBDL4 variant expression facilitated the secretion of full-length BiP (lane 4 vs 1 or 2 vs 1), but not the proteolytic fragment generated by RHBDL4 cleavage. RHBDL4 expression was detected by an anti-Strep antibody in cell lysate. Revert 700 total protein staining was used as loading control. (b) Endogenous BiP was cleaved by endogenously expressed RHBDL4. WT RHBDL4-dependent BiP proteolytic fragment in the media was detected by an anti-BiP antibody in WT HCT116 cells (red arrow), but not in RHBDL4 knockout HCT116 cells. Transferrin: loading control; additionally, the same samples run in different lanes were stained by Revert 700 for total protein loading control. Endogenous BiP was detected by an anti-BiP antibody and RHBDL4 was detected by an anti-RHBDL4 antibody in the cell lysate. All experiments were repeated in two biological replicates with similar results. For gel source data, see Supplementary Fig. 1.
Extended Data Fig. 8 | See next page for caption.
**Extended Data Fig. 8 | RHBDL4 cleaves other ER-resident proteins before ER retention motifs and facilitates the secretion of N-terminal proteolytic fragments.** RHBDL4 cleavage assays for V5-tagged (a) PDIA6, (b) Calreticulin and (e) ERP44. The putative cleavage sites are indicated by blue arrows in the schematic representations. (a) WT RHBDL4 cleaved PDIA6 at the C-terminus, resulting in secretion of the N-terminal fragment (blue arrow) into the medium. (b) WT RHBDL4 cleaved Calreticulin at the internal region (minor cleavage) and the C-terminus (major cleavage), resulting in secretion of the N-terminal fragments (blue arrows) into the SN. (c) WT RHBDL4 cleaved ERP44 before RDEL sequence. The secreted proteins were deglycosylation mix II sensitive. Because we could not source an antibody to specifically detect the RDEL sequence, we inserted a HA tag four amino acids before the RDEL sequence for detection. Anti-V5 detected proteolytic fragments 1 and 2 (p1 and p2), and p1 was also detected by the anti-HA antibody, indicating that one cleavage might happen after the HA tag. The expression of RHBDL4 was detected by an anti-Strep antibody. RHBDL4 cleavage assays for (d) endogenous PDIA6 and (e) endogenous Calreticulin. (d) WT RHBDL4 cleaved endogenous PDIA6. The proteolytic fragment of PDIA6 (red arrow) without the KDEL sequence was secreted into the SN. (e) WT RHBDL4 cleaved endogenous Calreticulin. The proteolytic fragment of Calreticulin (red arrow) without the KDEL sequence was secreted into the SN. Expression of RHBDL4 did not increase the endogenous level of PDIA6 and Calreticulin (lanes 2 and 3 vs 1). The expression of RHBDL4 is shown in Fig. 3g. BFA inhibitory assays for (f) PDIA6 and (g) Calreticulin. The secretion of proteolytic fragments of PDIA6 and Calreticulin was inhibited compared to cells treated with DMSO (lanes 5 vs 2, 11 vs 8). Endogenous PDIA6 and Calreticulin were detected by an anti-PDIA6 and an anti-CALR antibody, respectively. The expression of RHBDL4 is shown in Extended Data Fig. 7a. (h) Endogenous RHBDL4 cleaves endogenous Calreticulin. WT RHBDL4-dependent Calreticulin proteolytic fragments in the extracellular media were detected by an anti-CALR antibody in WT HCT116 cells (red arrows), but not in RHBDL4 KO HCT116 cells. The secretion of the major proteolytic fragment (near full-length Calreticulin, low intensity) and the minor proteolytic fragment (high intensity) were both detected. Transferrin: loading control. The endogenous RHBDL4 is shown in Extended Data Fig. 7b. (a–h) were repeated in two biological replicates with similar results. For gel source data, see Supplementary Fig. 1.
Extended Data Fig. 9 | The distribution of cysteine and serine protease clans. The clan distribution of cysteine proteases from (a) animals and (b) viruses. The clan distribution of serine proteases from (c) animals and (d) viruses. We have demonstrated the Dap-mediated protease substrate capture method for multiple cysteine proteases (TEV (Clan PA), UL36USP (Clan CA) and SCoV2-PLpro (Clan CA)) and serine proteases (HtrA2 (Clan PA), RHBDL4 (Clan ST) and RBBP9 (Note that RBBP9 has not been included in the MEROPS database by September 30, 2021. RBBP9 possesses the “alpha-beta hydrolase” fold⁴⁰, suggesting that it belongs to Clan SC). The data used to generate this figure were downloaded from the MEROPS database⁴⁰.
Extended Data Table 1 | Summary of crystallographic information

| Data Collection | R88P9_PHE (7OEX) |
|-----------------|------------------|
| Space Group     | P 1 2 1          |
| Cell dimensions |                  |
| a, b, c (Å)     | 36.96, 130.13, 38.99 |
| α, β, γ (*)     | 90.0, 115.24, 90.0 |
| Resolution (Å)  | 29.75 - 1.51     |
| *Rmerge         | 0.068 (0.625)    |
| *Mean I/I(I)    | 14.93 (2.44)     |
| *Completeness (%)| 95.9 (81.6)      |
| *Multiplicity   | 7 (5.7)          |
| Complexes in A.U.| 2               |

Refinement

| Resolution (Å) | 29.75 - 1.51 |
| No. Reflections| 49633       |
| Rwork / Rfree (%)| 16.7 / 16.4 |
| No. Atoms      | 3322        |
| Wilson B-factor (Å²) | 16.9 |
| Anisotropy     | 0.668       |
| Average B, all atoms (Å²) | 20.0 |
| B-factors      |             |
| Protein        | 20.5        |
| Ligands        | 21.77       |
| Water          | 29.06       |
| R.m.s. deviations |           |
| Bond length (Å) | 0.012      |
| Bond angle (*) | 1.42        |

* Highest resolution shell (in Å) shown in parenthesis.
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**
- Odyssey CLX imaging system (Li-Cor); MS Chemstation (Rev.C01.06[61]); MARS Data Analysis Software (version 3.20.2.2); XIA2 (version 0.7.90); DIALS (version 3.1.3), scaled using Aimless and Refmac (version 5.8.0258) in the CCP4 suite (version 7.0.078) of programs for crystal data processing

**Data analysis**
- Image Studio Lite (version 5.2.5) for western blot analysis; Mascot Search Engine Program (version 2.4) and Scaffold Proteome Software (version 4) for protein identification (Venn Diagrams); MaxQuant Software (version 1.6.3.4) and Perseus (version 1.6.2.3) for LFQ analysis (Volcano Plots); FCS Express Flow Cytometry software (version 7) for FACS data analysis; Fiji (ImageJ2) for IP analysis; GraphPad Prism (version 8) for graphs; Refmac (version 5.8.0258) and CDOT (0.8.9.2) for structure refinement and manual model building; Pymol (version 1.0.2) for protein structural visualization; ProteoWizard (version 3.0.11252) for converting LC-MS/MS RAW files to mZML format; Custom Python (version 3.8.1) scripts with the pyOpenMS package (version 2.4.0) for MS data preparation and processing. The code used for RBBP9 substrate identification by LC-MS/MS analysis is available at https://doi.org/10.5281/zenodo.5768340.

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 structure of R8BP9 in complex with Phe is available in the Protein Data Bank under accession code 7OEX. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the accession number PXD030381. All other datasets and materials generated or analyzed in this study are available from the corresponding author upon reasonable request. The data used to analyze serine and cysteine proteases clans were downloaded from the MEROPS database (https://www.ebi.ac.uk/merops/)

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.

- [x] 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

Life sciences study design

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

Sample size

No sample size calculations were performed. These are biochemical experiments, not animal experiments, and there are therefore no individuals to sample. Sample size is not a relevant parameter, but the number of replicates is relevant. The number of replicates is indicated in the relevant figure legend.

Data exclusions

No data was excluded

Replication

Three replicates were performed for protein identification by LC-MS/MS; Two replicates were performed to obtain the entire masses. All attempts at replicates were successful. Three or two replicates (annotated in the figure legends) were performed for immunoblotting analysis. All immunoblotting analysis were reproducible. Two replicates were performed for fluorescence-based aminopeptidase activity analysis. All fluorescent measurements were reproducible. All experiments were performed in the number of replicates indicated in the figure legends. All replicates were successful.

Randomization

These are biochemical experiments, where different components need to be added to different reactions. A single investigator needs to perform defined distinct and skilled operations on different samples to make the experiment meaningful and therefore randomization does not make sense.

Blinding

These are biochemical experiments, where different components need to be added to different reactions. A single investigator needs to perform defined distinct and skilled operations on different samples to make the experiment meaningful and therefore blinding does not make sense.

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 & experimental systems

| n/a | Involved in the study |
|-----|------------------------|
| ☐   | Antibodies            |
| ☒   | Eukaryotic cell lines  |
| ☒   | Palaeontology and archaeology |
| ☒   | Animals and other organisms |
| ☒   | Human research participants |
| ☒   | Clinical data         |
| ☒   | Dual use research of concern |

Methods

| n/a | Involved in the study |
|-----|------------------------|
| ☒   | Chip-seq               |
| ☒   | Flow cytometry         |
| ☒   | MRI-based neuroimaging |

March 2021
Validation

Validations are based on the data from the manufacturers.

Anti-Strep (ab76949) was validated by detecting recombinant proteins with strep-tag in cell lysate. https://www.abcam.com/strep-tag-ii-antibody-ab76949.html

Anti-Calreticulin (ab292516) was knockout validated in HAPI cell line. https://www.abcam.com/calreticulin-antibody-epr3924-er-marker-ab292516.html

Anti-PDIA6 (ab154820) was validated on IC were from Bio-Rad. Anti-MDN1 (A304-739A-T) and Anti-WDCP (A303-337-T) were from Bethyl Laboratories. Anti-Bip (GTX11340) and Anti-Mcm7 (GTX110278) and Anti-Fam120A (GTX120824) were from GenTex. Anti-MTMR3 (H00008897-B01P) was from Novus Biologicals. Anti-RHDL4 (A0869-1-A) and Anti-Cainexin (A0472-2-A) and Anti-LaminB1 (A20967-1-A), Anti-Transferin (66171-1-g) and Anti-SNRNP200 (23875-1-A) were from Progenesis. Anti-DSP (MA8908) and Anti-ub (MA88595) were from R&D Systems. Anti-Rps5 (A2535) was from Abclonal. Anti-GAPDH (TAB02519) was from Origene. Anti-goat-Mouse (925-68070; 925-32210) secondary antibodies and Goat anti-Rabbit (925-68071; 926-32211) secondary antibodies were from Li-Cor.

Regions used for the validation of antibodies are:

- Anti-Strep (ab76949): was validated by detecting recombinant proteins with strep-tag in cell lysate. https://www.abcam.com/strep-tag-ii-antibody-ab76949.html
- Anti-Calreticulin (ab292516) was knockout validated in HAPI cell line. https://www.abcam.com/calreticulin-antibody-epr3924-er-marker-ab292516.html
- Anti-PDIA6 (ab154820) was validated on IC were from Bio-Rad. Anti-MDN1 (A304-739A-T) and Anti-WDCP (A303-337-T) were from Bethyl Laboratories. Anti-Bip (GTX11340) and Anti-Mcm7 (GTX110278) and Anti-Fam120A (GTX120824) were from GenTex. Anti-MTMR3 (H00008897-B01P) was from Novus Biologicals. Anti-RHDL4 (A0869-1-A) and Anti-Cainexin (A0472-2-A) and Anti-LaminB1 (A20967-1-A), Anti-Transferin (66171-1-g) and Anti-SNRNP200 (23875-1-A) were from Progenesis. Anti-DSP (MA8908) and Anti-ub (MA88595) were from R&D Systems. Anti-Rps5 (A2535) was from Abclonal. Anti-GAPDH (TAB02519) was from Origene. Anti-goat-Mouse (925-68070; 925-32210) secondary antibodies and Goat anti-Rabbit (925-68071; 926-32211) secondary antibodies were from Li-Cor.
Eukaryotic cell lines

Policy information about cell lines

- **Cell line source(s)**: HEK293T cells were purchased from European Collection of Cell Cultures (ECACC). Expi293 cells (A14527) were purchased from Thermo Fisher. HCT116 cells were purchased from American Type Culture Collection (ATCC).

- **Authentication**: HEK293T cell line was authenticated by ECACC, Expi293 cell line was authenticated by Thermo Fisher, and HCT116 cell line was authenticated by ATCC using STR DNA profiling analysis to rule-out intra- and inter-species contamination.

- **Mycoplasma contamination**: All cell lines tested negative for Mycoplasma contamination.

- **Commonly misidentified lines** (See ICAC register): None

Flow Cytometry

Plots

- Confirm that:
  - The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
  - The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a ‘group’ is an analysis of identical markers).
  - All plots are contour plots with outliers or pseudocolor plots.
  - A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

- **Sample preparation**: 48 h after transfection, cells in 24-well plates were washed with PBS, detached by trypsin/EDTA solution, and resuspended in growth medium. Cells were pelleted and resuspended in PBS supplemented with 3% FBS.

- **Instrument**: Becton Dickinson LSRFortessa (407 nm violet laser V-450) for BFP excitation, 488 nm blue laser (5-525) for eGFP excitation, 561 nm yellow green laser (Y6-610) for mCherry excitation.

- **Software**: FCS Express 7 software (De Novo software)

- **Cell population abundance**: 20,000 intact cells (gated by from scatter and side scatter) were analyzed for each sample.

- **Gating strategy**: The front scatter and side scatter were used to identify intact cells. Non-transfected cells, cells only transfected with mCherry or eGFP plasmid and BFP positive cells were used to define the gate.

- Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.