Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Structural basis of Ac/N-degron recognition by the CTLH substrate adapter FAM72A/RUMY1

Frank Sicheri

sicheri@lunenfeld.ca

Lunenfeld-Tanenbaum Research Institute https://orcid.org/0000-0002-9824-2117

April Yu

Lunenfeld-Tanenbaum Research Institute https://orcid.org/0009-0005-4973-208X

Leon Baronijan

University of Toronto

Sasha Morozova

Lunenfeld-Tanenbaum Research Institute

Chetan Chana

Lunenfeld-Tanenbaum Research Institute

Daniel Mao

Mount Sinai Hospital

Brendon Seale

Lunenfeld-Tanenbaum Research Institute https://orcid.org/0000-0002-7519-3156

Cassandra Wong

Sinai Health https://orcid.org/0009-0009-7292-3780

Matthew Wong

University of Toronto

Anne-Claude Gingras

Lunenfeld-Tanenbaum Research Institute, Sinai Health https://orcid.org/0000-0002-6090-4437

Diane Haakonsen

Lunenfeld-Tanenbaum Research Institute

Alberto Martin

University of Toronto https://orcid.org/0000-0002-1260-7710

Article

Keywords: C-terminal to LisH (CTLH), FAM72A/RUMY1, degron, protein degradation, ubiquitination, UNG2

Posted Date: November 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11655015/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: Yes there is potential Competing Interest. Frank Sicheri is a founder and consultant for Induxion Therapeutics.

Structural basis of Ac/N-degron recognition by the CTLH substrate adapter FAM72A/RUMY1.

April J.Y. Yu[1,2+], Leon Baronijan[3+], Sasha Morozova[2,4], Chetan K. Chana[1,2], Daniel Y.L. Mao[2], Brendon Seale[1], Cassandra Wong[1], Matthew K. Wong[3], Anne-Claude Gingras[2,4], Diane L. Haakonsen[2,4], Alberto Martin[3*], Frank Sicheri[1,2,4*]

[1] Department of Biochemistry, University of Toronto, Toronto, Ontario, M5S 1A8, Canada

[2] Lunenfeld-Tanenbaum Research Institute, Sinai Health System, Toronto, Ontario, M5S 1A8, Canada

[3] Department of Immunology, University of Toronto, Toronto, Ontario, M5S 1A8, Canada

[4] Department of Molecular Genetics, University of Toronto, Toronto, Ontario, M5S 1A8, Canada

[+] Shared first authors

[*] Corresponding authors: alberto.martin@utoronto.ca; sicheri@lunenfeld.ca

Key words: C-terminal to LisH (CTLH), FAM72A/RUMY1, degron, protein degradation, ubiquitination, UNG2.

Summary

Selective protein degradation relies on recognition of short substrate sequences, termed degrons, by E3 ubiquitin ligases. FAM72A/RUMY1 is a substrate adaptor for the CTLH–MKLN1 E3 ligase that recruits UNG2 for degradation. Here, we identify the minimal UNG2 degron as its three N-terminal residues, Ac-MIG, and demonstrate that N-terminal acetylation is essential for binding. Systematic mutagenesis defines a FAM72A-binding consensus of Ac-$(M)_1$-$(I/L/V)_2$-G_3. A 1.8 Å co-crystal structure reveals a zinc-stabilized binding pocket in which backbone hydrogen bonds anchor the degron, while steric constraints and hydrophobic interactions dictate sequence selectivity. Mutations disrupting the binding pocket abolish degron binding, UNG2 engagement, and cellular degradation. Applying the consensus to proteomic datasets identifies IST1, QKI, and HPF1 as novel FAM72A substrates. Together, these findings establish FAM72A as an Ac/N-degron reader with a substrate repertoire extending beyond UNG2.

HIGHLIGHTS

1. using deletion analysis in cells combined with peptide synthesis, we determined the minimal degron sequence in UNG2 that is recognized by FAM72A as acetylated-MIG.
2. using peptide mutants, we elucidated the FAM72A binding consensus.
3. we validated the degron consensus by mutagenesis using IP pull downs and a GFP-mCherry substrate stability assay
4. we performed AP-MS and whole proteome analysis to identify likely FAM72A substrates.
5. we validated substrates conforming to the consensus by pull down, mutagenesis and stability assays

Introduction

Selective protein degradation is a fundamental mechanism for controlling protein abundance and cellular state [1-5]. Specificity is typically achieved through recognition of short sequence elements, termed degrons, by substrate-recruiting components of ubiquitin ligases. Degrons can be encoded directly in the primary sequence or generated through post-translational modification, thereby coupling protein turnover to cellular signalling and enzymatic activity. Because individual substrate receptors can recognize families of related degrons, defining the sequence and chemical features that govern degron recognition is essential not only for understanding how known substrates are regulated, but also for establishing the broader substrate repertoires and biological functions of ubiquitin ligases.

The C-terminal to LisH (CTLH) complex is a conserved multisubunit E3 ubiquitin ligase that regulates diverse biological processes. CTLH complexes contain a catalytic core built around the RING proteins RMND5A/B and MAEA together with scaffolding and substrate-recruitment components that can vary between assemblies [6-8]. One well-characterized substrate receptor, GID4, recognizes short and relatively degenerate N-terminal degrons [9-11], illustrating how incorporation of a particular receptor can confer substrate specificity on the CTLH machinery. Work from our group and others have further shown that GID4-dependent substrate recognition can be manipulated chemically and that direct recruitment to GID4 is sufficient to promote CTLH-dependent protein degradation [12-14] or the inhibition of target protein function [15, 16] thereby establishing the CTLH complex as a modular platform with potential therapeutic utility.

FAM72A (recently renamed RUMY1) has emerged as a second substrate-recognition component of the mammalian CTLH machinery. FAM72A was initially identified through a genome-wide genetic screen as a regulator of antibody diversification [17, 18]. Loss of FAM72A in B cells and B cell lines increased the abundance of uracil-DNA glycosylase 2 (UNG2), resulting in more efficient removal of AID-generated uracils and reduced somatic hypermutation and class-switch recombination. FAM72A was shown to bind to UNG2 directly [19, 20] and mediate its degradation, but not UNG1, the mitochondrial isoform that differs in its amino terminal region [21]. These findings established a physiological FAM72A–UNG2 axis through which regulated protein turnover alters the balance between faithful and mutagenic DNA repair.

More recent work established the molecular machinery underlying this regulation. FAM72A functions as a substrate adaptor for a distinct CTLH complex containing MKLN1 together with RANBP9/10, TWA1, RMND5A/B and MAEA [19]. FAM72A bridges UNG2 to MKLN1, allowing the CTLH catalytic machinery to ubiquitinate UNG2 and promote its proteasomal degradation. Structural and biochemical analyses revealed a modular architecture in which FAM72A directly engages MKLN1 while independently recognizing a degron within the disordered N terminus of UNG2. Thus, FAM72A is not simply a regulator of UNG2 activity, but a bona fide substrate-recognition module of a CTLH E3 ubiquitin ligase.

This distinction raises the possibility that UNG2 represents only one member of a broader FAM72A substrate repertoire. Such a model would parallel GID4, for which recognition of a degenerate class of N-terminal degrons enables recruitment of multiple proteins rather than a single dedicated substrate. Indeed, affinity-purification mass spectrometry of FAM72A recovered proteins in addition to UNG2 and established CTLH components [19], providing candidate interactions whose relationship to FAM72A-dependent substrate recognition remains unexplored.

The potential breadth of FAM72A function is particularly relevant in human disease. Humans encode four highly related FAM72 paralogues, *FAM72A–D* (*RUMY1-4*), whose expression is generally low in normal tissues but elevated in multiple cancers [22]. FAM72A, FAM72B and FAM72D are upregulated in primary tumor tissues and FAM72 expression inversely correlates with UNG2 protein abundance. Despite their high sequence similarity and shared ability to associate with MKLN1, only FAM72A robustly binds UNG2 and promotes its degradation [22]. The widespread expression of FAM72 proteins in cancer, together with reported associations of FAM72 with proliferation, cell-cycle progression and mutational burden, raises the possibility that the biological consequences of FAM72A activity extend beyond regulation of UNG2 alone. Identifying additional substrates is therefore important for understanding both the normal physiological functions of FAM72A and the consequences of its dysregulation in cancer.

Despite these advances, the molecular rules governing substrate recognition by FAM72A remain poorly defined. FAM72A binds the unstructured N terminus of UNG2 [19, 20], but the minimal sequence required for recognition, the contribution of post-translational modifications, the broader

sequence preferences of the receptor, and the extent to which this recognition mechanism is used to target other cellular proteins are unknown.

Here, we systematically define the molecular basis of substrate recognition by FAM72A. We identify the minimal FAM72A-binding degron within UNG2 as the N-terminal acetyl-MIG sequence. Systematic mutagenesis of this peptide establishes the sequence determinants that govern FAM72A binding and allows us to derive a degron consensus of Ac-$(M)_1$-$(I/L/V)_2$-G_3. These sequence requirements were validated in cells by mutagenesis, interaction assays and measurements of substrate stability, demonstrating that the biochemical binding preferences of FAM72A predict substrate recognition and degradation in a cellular context. This degron definition was then used to identify additional FAM72A substrates. Combining affinity-purification mass spectrometry with whole-proteome sequence analysis, we identified candidate proteins containing compatible N-terminal degrons and subsequently validated three substrates through direct binding and protein-stability assays. Together, these results establish FAM72A as the reader of a broader class of acetylation-dependent N-terminal degrons, define the sequence rules underlying substrate selection by the CTLH–MKLN1–FAM72A ligase, and expand its known substrate repertoire beyond UNG2. Our findings are consistent with and extend recent independent studies from the Schulman laboratory [23] and from the Yang and Wang laboratories (https://doi.org/10.64898/2026.09.19.752822) that similarly identify N-terminal acetylation as a key determinant of substrate recognition by CTLH–MKLN1–FAM72A. More broadly, our work provides a framework for predicting FAM72A substrates across the proteome and for understanding how this specialized CTLH substrate receptor contributes to selective protein turnover in normal physiology and disease.

Results

FAM72A/RUMY1 recognizes an N-terminal acetylated degron in UNG2

To define the sequence requirements for recognition of UNG2 by FAM72A, we first examined the contribution of residues within the UNG2 N-terminus. Previous studies demonstrated direct binding of both human and murine FAM72A to the first 25 residues of UNG2 [19, 20]. We recapitulated this result in insect cells using Strep-tagged full length human FAM72A and a GST-His fusion to the first 25 residues of UNG2 (UNG2 (1-25)-GST-His) (**Fig. 1a**). Next, we introduced single point substitutions within the first 13 residues of UNG2$^{1-25}$-GST-His and observe that only substitutions within the first three residues strongly disrupted FAM72A interaction (**Fig. 1a**). These results further localized the FAM72A binding element to the extreme N-terminus of UNG2.

To investigate whether post-translational modification of UNG2 contributes to FAM72A binding, we performed mass spectrometric analysis following Arg-C digestion of the UNG2$^{1-25}$-GST-His band excised from Strep-tag FAM72A pulldowns. The N-terminal Arg-C peptide of UNG2 (MIGQKTLYSFFSPSPAR) was identified predominantly in its N-terminally acetylated form (**Fig. 1b, Supplementary Table 1**). To test whether N-terminal acetylation contributes directly to FAM72A recognition we used a fluorescence polarization assay. Both FITC-labelled, N-terminally acetylated peptides corresponding to residues 1 to 25 and residues 1 to 13 of UNG2 bound FAM72A in a concentration-dependent and saturable manner (K_D=55.1±8.0 and 53.9 ±5.9 nM, respectively), whereas the corresponding non-acetylated peptides showed no detectable binding (**Fig. 1c**). These results demonstrate that N-terminal acetylation is required for high-affinity recognition of UNG2 by FAM72A.

To further map the binding determinants within the UNG2 N-terminus, we generated a series of C-terminal deletions to the Ac-UNG2 (1-13) peptide and assessed whether these could compete off the high affinity FITC-labelled Ac-UNG2(1-13) peptide. Acetylated UNG2 peptides comprising residues 1 to 13, 1 to 10, 1 to 7 and 1 to 4 competed with comparable potency, with IC_{50} values of 333 ± 31, 380 ± 5, 466 ± 12 and 401 ± 31 nM, respectively (**Fig. 1d**). Further truncations markedly reduced binding with Ac-UNG2 1 to 3 and Ac-UNG2 1 to 2, displaying 7 and >10 fold losses in binding activity (IC_{50} = 2.33 ± 0.66 μM and >5 μM respectively), while the acetylated N-terminal methionine alone was completely inactive. Consistent with our mutational

analysis on UNG2$^{1-25}$-GST-His (**Fig. 1a**), an I2D substitution in Ac-UNG2(1-13) also abolished binding (**Fig. 1d**). Together, these data identify Ac-MIG as the minimal FAM72A-recognition element, with the fourth residue contributing substantially to binding affinity, and establish the extreme N terminus of UNG2 as an acetylation-dependent FAM72A degron.

FAM72A recognizes a defined N-terminal sequence consensus that determines UNG2 binding and degradation

Next, we sought to define the sequence preferences for the three dominant N-terminal positions in the UNG2 degron peptide that governs interaction with FAM72A. We systematically substituted each of the first three residues of an N-terminally acetylated UNG2 peptide to every other amino acid and measured binding to FAM72A by fluorescence polarization (**Fig. 2a** and **Supplementary Fig. 1a**). FAM72A displayed distinct preferences at each position. At position 1, methionine and leucine were well tolerated, whereas substitution with cysteine substantially weakened binding, and all other substitutions extinguished binding. Position 2 showed broader tolerance for small or aliphatic residues, with isoleucine and valine supporting the strongest binding and cysteine and alanine also being well tolerated; leucine, serine, threonine and glycine progressively reduced affinity. In contrast, position 3 was highly constrained, with a strong requirement for glycine. Notably, peptides bearing the N-terminal sequences Ac-MVG bound at least as well as the natural UNG2 Ac-MIG sequence, whereas Ac-MLG, Ac-MSG, Ac-MTG and Ac-MGG showed progressively weaker binding. Together, these data define a sequence preference centred on an N-terminally acetylated hydrophobic residue followed by a small/aliphatic residue and an invariant glycine.

Although the *in vitro* binding assays revealed broad tolerance for small or aliphatic residues at position 2, N-terminal processing could restrict recognition in cells. In human cells, methionine aminopeptidases can co-translationally remove the initiator methionine when followed by a small residue, particularly alanine, glycine or serine, with processing efficiency depending on the surrounding sequence [24]. Thus, variants such as MAG or MSG would be expected to undergo initiator-methionine removal before N-terminal acetylation, preventing formation of the acetylated methionine-containing degron and engagement with CTLH-MKLN1-FAM72A.

We next asked whether the binding preferences derived from synthetic peptides were maintained in the context of full-length UNG2 in cells. Selected mutations spanning a range of peptide-binding affinities were introduced into UNG2 and their interaction with FAM72A was examined by reciprocal HA- and FLAG-pulldown assays (**Fig. 2b**). The degron variations MVG and MLG predicted to preserve high-affinity recognition retained robust interaction with FAM72A, whereas MAIG (predicted to be processed to AIG), MCIG (predicted to be processed to CIG), MCG, MSG, MTG and MDG abolished detectable binding (**Fig. 2b**). These results indicate that cellular recognition is more restrictive than binding to the corresponding N-terminally acetylated peptides, consistent with a potential contribution from N-terminal processing.

We further determined whether differences in FAM72A binding translated into altered UNG2 stability. We used a bicistronic reporter in which UNG2 was fused at its C-terminus to GFP, preserving its native N-terminus, while mCherry was expressed from an internal ribosome entry site (IRES). The GFP/mCherry fluorescence ratio therefore provided a normalized readout of UNG2 reporter abundance. Loss of FAM72A in HEK293T cells increased the GFP/mCherry ratio of wild-type UNG2, whereas re-expression of FAM72A reduced it, consistent with FAM72A-dependent degradation (**Fig. 2c**). Treatment with the proteasome inhibitor carfilzomib stabilized the wild-type reporter, whereas the lysosomal inhibitor bafilomycin had little effect (**Supplementary Fig. 1b**), supporting a proteasome-dependent degradation of UNG2.

We then tested substitutions at the first, second and third positions of the UNG2 degron using the same reporter assay in FAM72A KO HEK293T cells and following re-expression of FAM72A. Among the variants tested, MIG (wild-type UNG2), MLG and MVG strongly supported FAM72A-dependent degradation, with MIG showing the strongest FAM72A-dependent effects, followed by MLG and MVG (**Fig. 2c**). Intriguingly, MCIG- (predicted to be processed to CIG-) and MTG also showed subtle but reproducible FAM72A-dependent degradation. This suggests that although these degron variants exhibit lower affinity for FAM72A in the *in vitro* FP assay (K_D=3338 nM and 1185 nM, respectively), rendering their interactions undetectable by co-IP, they retain sufficient affinity to transiently engage FAM72A in cells and promote its degradation. Taken together, the fully concordant co-IP and reporter results support Ac-$(M)_1$-$(I/L/V)_2$-G_3 as the stringent consensus N-terminal degron motif for FAM72A-dependent recognition in cells, whereas inclusion of the more subtle variants detected by the GFP reporter assay supports a broader consensus represented by Ac-$(M/C)_1$-$(I/L/V/T)_2$-G_3.

Structural basis of degron recognition by FAM72A

To define the molecular basis of substrate recognition by FAM72A, we determined a 1.8 Angstrom X-ray crystal structure of zebrafish FAM72A (zFAM72A) in complex with the N-terminally acetylated UNG2 degron peptide Ac-MIGQ-NH2 (**Table 1**). While attempts to crystallize human and mouse FAM72A proved unsuccessful, zFAM72A (64% identify with human) crystallized readily in the presence of the Ac-MIGQ-NH2 degron peptide. Compared to hFAM72A, sequence alignment shows that zFAM72A contains two fewer residues at its extreme N-terminus (**Supplementary Fig. 2a**). Accordingly, residue numbering is shifted by -2 relative to hFAM72A.

The crystal structure revealed that the degron binds deep within a pocket in zFAM72A lined by residues conserved invariantly across evolution (**Supplementary Fig. 2a**). The well-defined electron density of the bound degron peptide in unbiased maps allowed unambiguous modeling of its conformation with the exception of the Gln4 sidechain, which lacked density and hence was truncated to the C_γ position (**Fig. 3a**). The map quality also allowed unambiguous modelling of the zFAM72A protein chain from residue 6 to residue 126, with the exception of a loop following β8 (residue 113 to 119), where ambiguous difference (Fo-Fc) density suggests that alternative loop conformations may exist. Overall, FAM72A adopts a β-tent fold, consisting of two four-stranded beta sheets packed roughly at right angle like a tent. At the apex of the tent lies the first of two zinc binding sites. The two beta sheets, formed by β4, 5, 6, and 7 on the left side and β3, 2, 1, and 8 on the right side creates a cradle onto which the degron engages. Distinct from other β-tent fold proteins, such as the CULT domain of CRBN, FAM72A harbors a second zinc binding site. Whereas the first zinc binding site of FAM72A consists of two CxxC motifs (C16, C19, C72 and C75) that tetrahedrally coordinates a Zn ion, the second zinc binding site, which is situated adjacent to the degron binding pocket, consists of a CxC motif (C63, C65) paired with CxxC motif (C87, C90).

The degron binding pocket of FAM72A is formed mainly by three flexible loops linking the secondary structure paired elements β2/β3, β4/β5 and α1/β7 (**Fig. 3a**). Within the pocket, an extensive hydrogen-bond network anchors both the degron and the flexible loops of FAM72A in place. Intriguingly, the degron engages FAM72A through direct, backbone-mediated hydrogen bonds between $M1^{degron}/L33^{FAM72A}$, $I2^{degron}/N96^{FAM72A}$, $G3^{degron}/V31^{FAM72A}$ and $Q4^{degron}/C94^{FAM72A}$ (**Fig. 3b, top left panel** and **Supplementary Fig. 2b**). Notably, by hydrogen

bonding with V31 on β2 and N96 and C94 on the α1/ β7 loop, the degron engages FAM72A in a manner reminiscent of an additional β-strand that locks the structure in place. Degron binding likely rigidifies the overall fold of FAM72A, as crystals only formed in the presence of peptide, despite the degron peptide not participating directly in crystal packing interactions. Furthermore, the high-resolution electron density map allowed modelling of water molecules present in the binding pocket. The direct interface between FAM72A and the degron is extended by additional water bridged hydrogen bonds between $M1^{degron}$/$A34^{FAM72A}$ and $C90^{FAM72A}$, $I2^{degron}$/$N96^{FAM72A}$ and $H98^{FAM72A}$ (**Fig. 3b, bottom left panel**). Lastly, the binding pocket is supported by an intricate set of intra-protein hydrogen bonds between S42, L91, S93, N95, N96, G97 and H98 of FAM72A that appears to lock the flexible loop between α1/β7 into a conformation that supports degron binding (**Fig. 3b, right panel)**.

Degron recognition by FAM72A also involves extensive hydrophobic interactions particularly with the degron side chains (**Supplementary Fig. 2b**). At position 1 of the degron **(Fig. 3c)**, the hydroxyl of the acetyl group forms a hydrogen bond with the hydroxyl group of Y81 of FAM72A, providing a structural explanation for the strong dependence on N-terminal acetylation for UNG2 binding. The methyl of the acetyl group and the methionine side chain pack against a hydrophobic surface formed by zinc binding site 2 and the hydrophobic residues L32, L33, V83, P86 and F99. At position 2 of the degron **(Fig. 3d)**, the Ile side chain sits in a hydrophobic cavity lined by FAM72A residues A30, N95, L33, L40, and M101. At position 3 of the degron **(Fig. 3e)**, the Gly backbone tunnels through a constriction point formed by FAM72A residues A30, V31, L33, C94, and F6, with little to no room to accommodate a side chain. As noted above, the Gly backbone of the degron forms two hydrogen bonds with the backbone of V31, forcing the residue into a highly strained, unusual region of the Ramachandran plot not accessible to other residues (**Supplementary Fig. 2c**). At position 4 of the degron (**Fig. 3e**), as noted, the main-chain amide NH hydrogen-bonds to the main-chain carbonyl oxygen of C94 of FAM72A, while the Gln4 side chain, which is disordered in the crystal structure, is oriented towards solvent.

Rationalizing the degron consensus

The degron-bound structure of FAM72A tentatively explains the strict requirement for N-terminal acetylation and the observed degron consensus *in vitro*. The acetyl group is recognized through a

side-chain-mediated hydrogen bond between its carbonyl oxygen and the Y81 hydroxyl of FAM72A, explaining why the unacetylated peptide fails to bind. At degron position 1, only Met, Leu, and Cys are well tolerated. These residues fit a hydrophobic binding pocket without steric conflict, whereas similarly sized but differently branched residues, such as Ile, or bulkier residues, such as Phe, introduce clashes. Position 2 is more permissive, but binding affinity follows a clear trend: intermediate sized hydrophobic residues, such as Val and Ile, bind best, followed by smaller hydrophobic residues, small polar residues, and finally Gly. This pattern is consistent with an intermediately sized hydrophobic pocket in which optimal side-chain packing enhances affinity. Position 3 is restricted to Gly because of both steric and conformational constraints. Modelling Ala at this position produces clashes with the FAM72A backbone near C94 and N95. In addition, Gly3 adopts a backbone conformation that is disallowed for other amino acids. Lastly, the restriction of favourable contacts to only the backbone and not the side chain of position 4 could explain why replacing Gln4 with glycine had no effect on the FAM72A–UNG2 (1–25) interaction in pulldown assays (**Fig. 1a**), yet truncation of the degron peptide from a tetramer to a trimer reduced binding ~8-fold (**Fig. 1d**). Together, these observations allow rationalization of the acetylation dependence and the Ac-$(M)_1$-$(I/L/V)_2$-G_3 consensus recognized by FAM72A *in vitro* and *in vivo*.

Structure validation

We next tested whether structural contacts observed in the crystal structure are required for degron recognition. Residues lining the peptide-binding pocket of zFAM72A were individually mutated and binding to the acetylated UNG2 degron was measured by fluorescence polarization (**Fig. 4a**). Consistent with the presence of a conserved degron binding pocket, wild-type zFAM72A bound the degron peptide with a K_D of 231 ± 35 nM. Substitution of Y81 to Phe or Ala to abolish its hydrogen bond with the N-terminal acetyl group of the degron completely abrogated peptide binding. Mutation of V83, which contacts the first Met of the degron, to an Arg or Phe residue completely abolished degron binding. Similarly, mutation of A30, which contacts both the second and third residue of the degron to a bulkier His also abolished degron binding. Mutation of Phe6, which contributes to the restrictive tunnel through which Gly3 of the degron exits, to Arg greatly attenuated degron binding, though some binding is apparent at highest concentration tested. Lastly,

disrupting the intramolecular hydrogen bonding network within FAM72A by mutating N95 and N96 to Arg, also completely abolished degron binding. Notably, all mutant proteins eluted as mono-dispersed peaks with unaffected retentions times indicating that the protein remained fully folded (**Fig. 4b**). Together, these findings highlight that the crystal structure reported here reflects the degron binding function of FAM72A in solution.

We then asked whether disruption of this structurally defined binding pocket also prevented substrate recognition in cells. Consistent with expectations, FAM72A binding pocket mutants that were defective in peptide binding (F8R, A32H, Y83A, Y83F, V85F, V85R, N97R, N98R) showed correspondingly reduced interaction with full-length UNG2 in cellular pulldown assays (**Fig. 4c**).

We next examined UNG2 degradation using our stability reporter assay in FAM72A KO cells following re-expression of wild-type or mutant human FAM72A (**Fig. 4d top**). V85F, V85R, N98R, A32H and Y83A impaired UNG2 degradation, whereas F8R, N97R and Y83F retained substantial activity despite all maintaining comparable protein expression levels (**Fig. 4e bottom**). This may reflect residual interactions sufficient to support degradation in cells, while differences between zebrafish and human FAM72A could also contribute to the discrepancy between peptide binding and cellular activity.

Together, these results provide a structural explanation for the sequence specificity established in **Figures 1** and **2** and identify a discrete N-acetyl-degron-binding pocket in FAM72A that is required for substrate engagement and degradation.

Identification of novel FAM72A substrates

To identify additional FAM72A substrates, we combined the newly defined FAM72A degron-binding consensus with mass spectrometry-based approaches. Here we analyzed data from three complementary datasets: our previously reported HA- affinity purification mass spectrometry (AP-MS) analysis of on CH12F3-2 B-cells (hereafter referred to as CH12 cells) stably expressing HA-tagged FAM72A, which identified FAM72A-associated proteins [17], and two newly generated whole-proteome analyses of mouse LLC1 carcinoma and mouse CH12 cells to identify proteins whose abundance is altered by perturbation of the FAM72A-MKLN1 pathway (**Fig. 5a and 5b**). To create matched comparison cell lines, we ectopically overexpressed mouse FAM72A in LLC1

carcinoma cells that display low endogenous FAM72A expression (**Supplementary Fig. 3a**) and used FAM72A- and MKLN1-deficient mouse CH12 cells which display high basal levels of FAM72A expression [17]. To refine our search, we applied two criteria to select potential substrates for follow-up validation: proteins had to contain an N-terminal sequence matching the FAM72A degron consensus and exhibit differential abundance in at least one of our whole proteome datasets. Using the stricter, cell-validated FAM72A degron consensus, we identified three candidate substrates, namely IST1, HPF1, and QKI in addition to UNG2 (**Fig. 5c**). Like UNG2, all three candidates exhibited reduced abundance in LLC1 cells over-expressing FAM72A (**Supplementary Fig. 3b**), consistent with FAM72A-dependent regulation of their abundance. Furthermore, flow-cytometry based reporter assays showed that HPF1 and QKI undergo FAM72A-dependent destabilization that requires an intact N-terminal degron (**Supplementary Fig. 3c**). We also tested three additional candidate substrates that conform only to the broader *in-vitro* FAM72A degron consensus, namely Tmem50b, Nup62, and VPS37b. None of these three proteins exhibited detectable FAM72A-dependent destabilization in the reporter assay (**Supplementary Fig. 3d**).

We decided to further validate IST1 (Increased Sodium Tolerance 1) and HPF1 (Histone PARylation Factor 1) since these two proteins were also increased in abundance in MKLN1- and MAEA-deficient cells in previously published whole proteome screens [25, 26]. IST1 is a member of the endosomal sorting complex required for transport (ESCRT)-III which functions during cytokinesis, mitotic spindle disassembly, and endosomal tubulation [27]. To test whether IST1 protein abundance depends on FAM72A-MKLN1-CTLH, we performed Western blotting for IST1 in mouse and human cell lines that overexpress FAM72A and are deficient in FAM72A, respectively. IST1 levels were reduced in FAM72A over-expressing mouse LLC1 cells (**Fig. 5d**) but increased in *FAM72A*$^{-/-}$ and *MAEA*$^{-/-}$ human Jurkat T cells (**Fig. 5e**), consistent with FAM72A-dependent degradation. To assess whether IST1 physically associates with FAM72A, HA-tagged FAM72A was immunoprecipitated in *FAM72*$^{-/-}$ human HCT116 cells transfected with 3x-HA-FAM72A. Endogenous IST1 and UNG2 co-immunoprecipitated with HA-tagged FAM72A (**Fig. 5f**). Stability analysis of IST1 by cycloheximide (CHX) chase in WT and *FAM72A*$^{-/-}$ Jurkat cells revealed that IST1 is stabilized in the absence of FAM72A (**Fig. 5g**).

HPF1 is a PARP1/2 co-factor which functions during the DNA damage response [28]. Importantly, HPF1 appeared in both the FAM72A-overexpressing and FAM72A-deficient proteomic screens

(**Fig. 5a-c**). These findings were validated by Western blotting in FAM72A over-expressing LLC1 cells (**Supplementary Fig. 3e**) and CH12 cells deficient in FAM72A, MKLN1, and MAEA (**Supplementary Fig. 3f**), consistent with HPF1 being degraded by FAM72A-MKLN1-CTLH. The increased levels of HPF1 in *Fam72a*$^{-/-}$ CH12 cells were also observed in an UNG-deficient background, suggesting the increased abundance of HPF1 is independent of UNG2 accumulation (**Supplementary Fig. 3f**). To test whether HPF1 physically associates with FAM72A, we stably introduced HA-tagged mFAM72A into *Fam72a*$^{-/-}$ CH12 cells with two sizes of flexible linkers ($G4S_1$ and $G4S_3$) separating *Fam72a* cDNA and the HA tag. HPF1 protein levels returned close to wildtype levels in lysates from *Fam72a*$^{-/-}$ cells re-introduced with FAM72A, but interestingly not IST1 (**Supplementary Fig. 3g**). Endogenous HPF1, but not IST1, co-immunoprecipitated with HA-FAM72A and MKLN1 in CH12 cells (**Supplementary Fig. 3h** and **data not shown**). Together, these data suggest that the FAM72A-MKLN1-CTLH complex targets the ESCRT protein IST1 and DNA damage response co-factor HPF1 for degradation in a cell type-specific context.

Discussion

We previously characterized FAM72A as a substrate adaptor that recruits UNG2 to the $CTLH^{MKLN1}$ E3 ligase complex for polyubiquitylation and subsequent proteasomal degradation during antibody maturation in B cells. However, the expression of FAM72A in other lymphoid populations, including thymocytes [22], together with its elevated expression and association with poor prognosis across a broad range of cancers [19, 22], suggests that its biological functions extend beyond regulation of UNG2 and that additional FAM72A substrates likely remain to be identified. Here, we define the molecular basis of substrate recognition by FAM72A. We biophysically reconstituted the interaction between FAM72A and the N-terminal degron of UNG2 and used systematic mutagenesis to establish a three-residue N-terminal consensus motif, Ac-$(M)_1$-$(I/L/V)_2$-G_3, while demonstrating an essential role for N-terminal acetylation in substrate recognition. The 1.8 Å X-ray crystal structure of zebrafish FAM72A in complex with the UNG2 degron corroborates and rationalizes the sequence preferences identified by mutagenesis.

An important consideration is how the degron preferences defined in vitro relate to the processing of protein N termini in cells. Our biochemical data indicate that N termini conforming to the sequence Ac-$[MLC]_1$-$[VCALSTG]_2$-G_3 can, in principle, bind FAM72A. However, whether a protein ultimately presents such a sequence depends on co-translational N-terminal processing. As nascent polypeptides emerge from the ribosome, initiator methionine can be removed by methionine aminopeptidases (MetAPs), particularly when the second residue is small and uncharged. N-terminal acetyltransferases (NATs) subsequently modify protein N termini according to sequence-dependent substrate preferences. Importantly, N-terminal acetylation and methionine excision are kinetically competing co-translational processes, such that acetylation of the initiator methionine can occur before its removal [29, 30]. We therefore tested whether the sequence preferences derived from purified components also predict substrate recognition in cells using reciprocal pulldown experiments and GFP-based stability reporters containing variant UNG2 degrons. In comparison to our *in vitro* findings, only the MLG and MVG degron mutants co-immunoprecipitated with FAM72A. The processing of the N-terminus within cells may differ from what is predicted *in vitro*, potentially explaining this discrepancy. For example, MetAPs and NATs may compete for access to the nascent N-terminus, potentially resulting in a processed N-terminal sequence that differs from what is expected. Together, these experiments provide an

important link between the biochemical consensus and the more complex constraints imposed by cellular N-terminal processing.

Degron sequence alone, however, is unlikely to determine whether a candidate protein is recognized by FAM72A. The N terminus must also remain physically accessible. A degron buried within the protein fold or occluded by a protein–protein interaction would be expected to be inaccessible to FAM72A and therefore protected from FAM72A-dependent degradation. Conversely, changes in protein conformation, assembly state, or localization could potentially regulate degron exposure and thereby provide an additional layer of substrate control.

Our findings demonstrate the utility of the degron consensus as a predictive framework for identifying FAM72A substrates and provide an initial indication that FAM72A-dependent degradation extends beyond UNG2. More broadly, identifying these substrates will be important for understanding why FAM72A is expressed outside activated B cells and how its dysregulation may contribute to cancer and other biological contexts. Having defined and validated a FAM72A degron consensus as Ac-$(M)_1$-$(I/L/V)_2$-G_3, we next considered the potential prevalence of FAM72A substrates within the proteome. A ScanProsite search for the consensus motif at protein N termini identified only a small fraction of sequences as potential FAM72A substrates (0.024%, 141 out of 575,748 total entries for human and 0.02%, 116 out of 576,073 entries for mouse) (**Supplementary Table 2**), illustrating how these combined constraints substantially narrow the potential substrate pool. By integrating proteomic evidence with the presence of a compatible N-terminal degron, we were able to substantially narrow the candidate pool and validate three novel proteins, namely QKI, IST1 and HPF1, as FAM72A substrates in various cell lines of both human and mouse origin. IST1 is a member of the ESCRT-III complex, which catalyzes reverse-topology membrane scission. This mechanism underlies various processes within the cell, including mitotic spindle assembly, nuclear envelope remodelling, multivesicular body biogenesis, and the final scission step of cytokinesis [31-33]. Thus, degradation of IST1 by FAM72A-MKLN1-CTLH may represent a mechanism to prevent premature ESCRT-mediated cytokinetic abscission or nuclear remodelling. In line with this proposed mechanism are observations by our group and others that FAM72A expression and activity are cell-cycle regulated [17, 34]. Interestingly, the abundance of several ESCRT proteins such as VPS37A and VPS37B is dysregulated in our whole proteome screens, which is possibly a consequence of impaired ESCRT function. Future functional work is

required to test this hypothesis and identify specific defects in the ESCRT pathway caused by FAM72A deletion or overexpression.

The second identified FAM7A substrate, HPF1, is a co-factor of PARP1 and PARP2, enzymes that detect single- and double-stranded breaks and catalyze the addition of poly ADP-ribose (PAR) chains to chromatin and DNA repair factors, thereby promoting chromatin relaxation and facilitating repair [35, 36]. Cells deficient in HPF1 are hypersensitive to DNA damage induced by alkylating agents and topoisomerase inhibitors [37, 38]. It is tempting to speculate that FAM72A-mediated degradation of HPF1 may serve a complementary, albeit upstream, role to UNG2 degradation in B cells by stalling faithful repair of DNA lesions to further promote a mutagenic environment conducive to antibody diversification. Additionally, FAM72A expression appears to be highest in developing T cells of the thymus and to a lesser extent, early B cells developing in the bone marrow [22]. Both cell types undergo V(D)J recombination, a process involving RAG1/2-mediated double-strand DNA cleavage to initiate antigen receptor gene rearrangement. HPF1 degradation by FAM72A may therefore help regulate the unique repair requirements of RAG1/2-mediated lesions in these cell types. Further work is required to characterize the functional significance of degrading HPF1 in the context of DNA repair and lymphocyte development.

Lastly, the third identified substrate, QKI, is a STAR-family RNA-binding protein containing a KH domain that regulates alternative splicing, mRNA stability and localization, and circular RNA biogenesis[39-41] . Notably, QKI has well-established tumour-suppressive functions, with its loss contributing to tumorigenesis in angiocentric glioma through recurrent MYB–QKI rearrangements [42]. Reduced QKI expression has also been implicated in lung and colorectal cancers, where it is associated with increased tumour cell proliferation and, in lung cancer, with poorer clinical outcomes [43, 44]. Our identification of QKI as a substrate of the CTLH–MKLN1–FAM72A complex therefore raises the possibility that elevated FAM72A expression in tumours could promote oncogenesis through the post-translational depletion of a tumour suppressor, independently of genetic loss or transcriptional silencing.

An additional unresolved question concerns the functional diversification of the human FAM72 paralogues. Whereas most species contain a single ancestral FAM72 gene, the human lineage contains four highly similar paralogues, FAM72A–D, which differ at only a small number of residues (**Supplementary Fig. 4**). Whether FAM72B–D function as CTLH substrate receptors

analogous to FAM72A remains unclear. Our previous pulldown experiments showed that all four paralogues can associate with MKLN1, whereas only FAM72A detectably binds UNG2 [22]. This observation is consistent with the recently reported FAM72A–UNG2–MKLN1 cryo-EM structure, as the FAM72A residues that contact MKLN1 are conserved among FAM72B–D [23]. In contrast, our FAM72A–UNG2 degron structure provides a potential structural explanation for paralogue-specific substrate recognition. One notable difference between FAM72A and FAM72B is the substitution of Leu94 with proline. Leu94 lies within the α1 helix positioned directly above the peptide-binding site. Because proline disfavors α-helical geometry, the L94P substitution could perturb the local structure of this helix and thereby alter the geometry of the degron-binding pocket. FAM72C and FAM72D instead differ from FAM72A at residues that may influence substrate recognition through distinct mechanisms. Gly99 is replaced by arginine in these paralogues. Gly99 lies near the Met1 side chain of the bound degron, and introduction of a substantially larger, charged side chain at this position could sterically interfere with accommodation of the substrate N terminus. Trp125 is also replaced by arginine. A W125R substitution introduced into FAM72A has previously been reported to abolish UNG2 binding [45], although the structural basis for this effect is less obvious because Trp125 lies more than 12 Å from the degron-binding site within a C-terminal region lacking well-defined secondary structure. Its effect may therefore be indirect, potentially altering the conformation or dynamics of the substrate-binding surface rather than contacting the degron directly. While we can rationalize why FAM72B-D cannot bind UNG2, we can't rule out that each can bind to degrons with an altered consensus. Thus, further investigation is warranted.

Together, our findings define the molecular principles underlying substrate recognition by FAM72A and establish a framework for understanding its broader role as a CTLH substrate receptor. FAM72A recognizes a restricted class of acetylated N-terminal degrons through a binding pocket in which backbone interactions anchor the peptide while stringent steric constraints dictate sequence selectivity. Coupling this structural mechanism with proteomic datasets enables the prediction and identification of additional cellular substrates and provides a route toward defining FAM72A functions beyond UNG2 regulation. The expansion of FAM72 into four closely related human paralogues further raises the possibility that this substrate-recognition system has

undergone functional diversification, potentially allowing distinct FAM72 proteins to control overlapping or specialized substrate repertoires in different cellular contexts.

Methods

Bacterial expression construct generation:

WT and mutant N-terminal His-SUMO tagged zebrafish FAM72A (His-SUMO-zFAM72A) were codon optimized for *E. coli* expression and cloned into pET24a(+) bacterial expression vector (GenScript).

Baculovirus expression construct generation:

DNA sequences for WT human N-terminal Strep tagged FAM72A (Strep-hFAM72A), WT and mutant UNG2 residues 1 to 25 fused to GST-His tags ($UNG2^{1-25}$-GST-His) used in pull down interaction studies, and WT N-terminal His tagged FAM72A (6xHis-hFAM72A) used in FP degron binding studies were codon optimized for Sf9 expression and cloned into pLIB baculovirus expression vectors (GenScript). Bacmid was generated via transformation into DH10Bac competent cells (Thermo Fisher Scientific) following supplier protocols. Effectene Transfection Reagent (Qiagen) was used for all Sf9 cell transfections. Baculovirus was amplified for three generations following standard protocols.

Sf9 cell maintenance:

Sf9 cells were maintained in I-Max serum free medium (Wisent inc. 301-045-LL) supplemented with 1x L-glutamine (Wisent inc. 609-065 EL), 1x Penicillin Streptomycin (Wisent inc. 450-201 EL), and 5% FBS (Wisent inc. 090150). Sf9 cells were maintained in suspension culture at 27°C, 110 rpm, in the dark.

Protein interaction studies by pull downs from SF9 cells:

12.5 ml cultures of suspension Sf9 insect cells grown to 2 million per mL density were co-transfected with 1.25 ml each of the indicated P3 baculoviruses for $UNG2^{1-25}$-GST-His and Strep-FAM72A protein expression. Cells were harvested by centrifugation at 1000 xg for 10 minutes and lysed by sonication in 20 mM HEPES pH 7.5, 500 mM NaCl, 0.25 mM TCEP, 20 mM imidazole and 0.05% Brij-35 (Buffer A) supplemented with 1x cOmplete, EDTA-free protease inhibitor cocktail tablet (Sigma-Aldrich) and 1 mM PMSF at 4 ℃. Cleared lysates were split into two aliquots and each aliquot incubated for 30 min with either Ni^{2+}NTA (EMD Millipore) or StrepTactin XT 4 Flow (IBA Lifesciences) resin. The resin was washed 5 times with Buffer A, resuspended in Laemmli buffer and visualized by SDS-PAGE and Coomassie staining.

Mapping of post-translational modifications on UNG2 by mass spectrometry:

Arg-C in-gel digestion: An excised SDS-PAGE gel band containing the separated $UNG2^{1-25}$-GST-His protein isolated from a Strep-hFAM72A pulldown on StrepTactin XT resin was first washed in 50 mM Tris-HCl pH 7.4 buffer and then dehydrated with 95% ethanol. Cysteines were then reduced using 5 mM dithiothreitol (DTT) for 30 minutes at 56 °C. Then, the gel band was then dehydrated again and cysteines were capped by reaction with 50 mM iodoacetamide in the dark for 45 minutes at room temperature. The gel band was dehydrated again with ethanol, then treated with 1 µg clostripain (endoproteinase-Arg-C) in 5 mM DTT and 10 mM $CaCl_2$ in 50 mM Tris-HCl pH 7.4 for overnight digestion at at 37 °C. Digested peptides were collected and the gel band was washed with two fractions of 5% formic acid. The wash fractions were collected and added to the initial extracted peptides, then dried by vacuum centrifugation.

Liquid chromatography mass spectrometry: Dried peptides were resuspended in 5% formic acid and analyzed via liquid chromatography mass spectrometry (LC-MS) using a Thermo Fusion Lumos Tribrid mass spectrometer system. One quarter of the sample was injected using a Thermo Vanquish Neo liquid chromatography system and trapped onto an Optimize Technologies C18 in-line trap column then separated across an IonOpticks Aurora 3 Ultimate (25 cm x 75 um) XT column held at 40 °C using an IonOpticks heater. A 0.1% formic acid (A) /0.1% formic acid in acetonitrile (B) gradient was delivered at 400 nl/min from 3.2% B to 16.8% B over 56 minutes, 16.8% to 24.8% B over 21 minutes, and 24.8% to 35.2% B over 13 minutes. This was followed by a column wash of 76% B over 13 minutes. The total gradient time was 105 minutes. Mass spectrometry acquisition was performed in data-dependent acquisition mode using only the Orbitrap mass analyzer. The MS1 scan had an accumulation time of 50 ms within a mass range of 400–1600Da, using Orbitrap resolution of 120000, 30% RF lens, AGC target of 200% and 1800 volts. This was followed by MS/MS scans with a total cycle time of 2 seconds. An accumulation time of 35 ms and 33% HCD collision energy were used for each MS/MS scan. Each candidate ion was required to have a charge state from 2-7 and an AGC target of 400%, isolated using Orbitrap resolution of 15,000. Previously analyzed candidate ions were dynamically excluded for 9 seconds.

MS data search: The RAW files were searched using MSFragger v4.1 [46], with peptide and protein identifications validated using PeptideProphet and ProteinProphet at 1% FDR rate, within the

FragPipe 22 software pipeline. Default parameters for Orbitrap DDA data were used, except that N-terminal methionine excision was disabled and all modifications were enabled for the preliminary search. The data were search against the Spodoptera frugiperda (Sf9) UniProt proteome (UP000829999) supplemented with decoys, common contaminants, and the sequence of human UNG2 appended.

Large Scale Protein Expression and Purification:

All purification steps were completed at 4°C. 6xHis-hFAM72A was expressed in Sf9 insect cells by infecting 4-8 litres of suspension culture grown to 2 million/mL density in 2.8 liter Fernbach flasks with 10% volume of P3 baculovirus. 72 hours following infection, cells were harvested by centrifugation at 1000 xg for 10 minutes, and lysed by sonication in buffer containing 20 mM HEPES pH 7.5, 500 mM NaCl, 0.25 uM TCEP, and 20 mM imidazole supplemented with 1x cOmplete, EDTA-free protease inhibitor cocktail tablet (Sigma-Aldrich), 1 mM PMSF, and benzonase. Lysate cleared by centrifugation at 38,000 xg for 40 minutes were applied to HisTrap FF columns (Cytiva), and the captured protein eluted on an AKTA Fast Protein Liquid Chromatography purification system (FPLC; Cytiva) using a gradient from 20mM to 350 mM imidazole. Fractions containing the FAM72A protein identified by SDS-PAGE were pooled, dialyzed in lysis buffer without imidazole, and incubated overnight with 6×His-tagged TEV protease. The sample was subsequently loaded onto a HisTrap FF column and the flowthrough containing the FAM72A protein was concentrated using an 15 mL Amicon Ultra centrifugal filter device to a final volume of 1 mL. The concentrated FAM72A protein was then further purified through a HiLoad Superdex 75 pg column (Cytiva) equilibrated in buffer containing 20 mM HEPES pH 7.5, 200 mM NaCl, 5% glycerol and 0.25 mM TCEP, then concentrated to ~ 10mg/ml and flash frozen in liquid nitrogen.

WT and mutant His-SUMO-zFAM72A were expressed in T7Shuffle *E. coli* (New England Biolabs) grown in TB medium supplemented with 50 ug/mL kanamycin. 4L cultures were grown at 30°C to an optical density (OD_{600}) of 1.0-1.2, and protein expression was induced with 0.5 mM IPTG overnight at 16°C. For lysis, harvested cells were resuspended in 150 mL buffer containing 50 mM HEPES pH 7.5, 500 mM NaCl, 5% glycerol, 20 mM imidazole and 0.5 mM TCEP supplemented with lysozyme, benzonase, and 1mM PMSF. Cells were disrupted by four passages through an Avisten homogenizer at 1000 bar. Clarified lysate were applied to HisTrap FF columns

(Cytiva), and the captured protein eluted on an AKTA FPLC system using a gradient of 20mM to 250 mM imidazole. Fractions containing the target protein identified by SDS-PAGE were pooled, dialyzed in the lysis buffer without imidazole, and incubated overnight with 6xHis-tagged Ulp1 enzyme. The sample was subsequently loaded onto a HisTrap FF column, and the flowthrough containing zFAM72A protein was concentrated to a final volume of 5 mL. The protein was further purified through a HiLoad Superdex 75 pg column (Cytiva) equilibrated in 20 mM HEPES pH 7.5, 200 mM NaCl, 5% glycerol and 0.5 mM TCEP, then concentrated to 2 to 20 mg/ml and flash frozen in liquid nitrogen.

Fluorescence polarization degron peptide binding assays:

FAM72A-degron peptide binding experiments in 25 ul final volume were performed with 10 nM of the indicated FITC-labelled peptide (Biomatik or GenScript) and increasing concentrations of hFAM72A or zFAM72A in a final solution containing 20 mM HEPES pH 7.5, 200 mM NaCl, 0.25 mM TCEP, 0.01% Brij-35, and 0.1 mg/mL BSA. Competitive displacement assays were performed in the same solution condition using 100 nM final hFAM72A or zFAM72A in complex with 20 μM Ac-UNG2(1-13)-FITC and increasing concentrations of $UNG2^{Nterm}$ peptide (Biomatik or GenScript). Samples were incubated for 1h and fluorescence intensities were measured using a BioTek Synergy Neo plate reader with excitation and absorbance at 485.528 nm, respectively. Fluorescence polarization was calculated with the Gen5 Data Analysis Software. Binding curves and K_D values were generated using the One Site -Total function in GraphPad Prism. Competition curves and IC_{50} values were generated using a four-parameter dose-response function in GraphPad Prism.

FAM72A-degron substrate complex structure determination:

Protein Crystallization: Full length WT zFAM72A (residues 1-147) was pre-incubated with the Ac-MIGQ-NH2 degron peptide at a final concentration of 500 uM and 572 uM, respectively, for 30 min. 0.2 uL protein complex was then mixed 1:1 with crystallization buffer containing 1.5M ammonium sulphate, 0.1M BTP pH 7.0 after in a 96-well sitting drop tray. Protein crystals were observed in one day. Crystals were transferred into a cryoprotection buffer containing 1.5M ammonium sulphate, 0.1M BTP pH 7.0, and 25% ethylene glycol before flash frozen using liquid nitrogen.

X-ray Data Collection, Processing and Structure Refinement: Diffraction data for the zFAM72A + Ac-MIGQ-NH2 degron peptide complex (PDB ID: 39LZ) was collected on a MicroMax 007HF (Rigaku Americas Corp) with an Rigaku HyPix-6000HE detector (Rigaku Americas Corp) and processed with Crysallis Pro. Five percent of the processed data was set aside for the R-free calculation using Reflection Editor. Molecular replacement was performed using Phaser [47] with an AlphaFold model for zFAM72A [48] used as a search model. Crystallographic refinement was performed using PHENIX [47] and model building was performed using Coot [49]. Software used in this project was installed and configured by SBGrid [50]. X-ray data collection and refinement statistics are summarized in **Table 1**.

Analytical Size Exclusion Chromatography:

200 ug of the purified WT and mutant zFAM72A proteins diluted to a final volume of 200 uL were injected onto a Superdex 75 Increase 13/300 GL column (Cytiva) attached to an AKTA FPLC equilibrated in buffer containing 20 mM HEPES pH 7.5, 200 mM NaCl, 5% glycerol and 0.5 mM TCEP. Absorbance was monitored at 280 and 260 nm wavelength.

Protein stability reporter assay:

The pCS2+-degron-GFP-IRES-mCherry reporter constructs were generated as in [51]. All pCS2-degron-GFP-IRES-mCherry constructs are listed in **Supplementary Table 3**. Wild-type HEK293T or FAM72A KO cells were seeded in 6-well plates at a density of 250,000 cells. The next day, 40 ng of reporter plasmid and empty vector up to 400 ng total were transfected using PEI and harvested for flow cytometry after 48h. When FAM72A wild-type or mutant proteins were over-expressed, 100ng of FAM72A plasmids were co-transfected. Cells were analyzed on an Attune NxT flow-cytometer and the GFP/mCherry ratio was analyzed using FlowJo.

Mammalian cell lines and culture conditions:

Jurkat and CH12F3-2 (CH12) cells were cultured in RPMI 1640 medium containing L-glutamine supplemented with 10% fetal bovine serum (FBS), penicillin-streptomycin, and 50 μM β-mercaptoethanol. HEK293T and BOSC23 cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and penicillin-streptomycin. Cells were cultured in a 37°C, 5% CO_2 direct heat incubator.

Plasmids for mammalian expression studies and oligonucleotides:

The pMX-PIE expression vector was a gift from B.R. San-Martin. All plasmids were sequence-validated by Sanger or Nanopore sequencing. Oligonucleotides for cloning of sgRNAs into px459, cDNA cloning into pMX-PIE, and site-directed mutagenesis or tagging of pMX-PIE were purchased from Invitrogen.

CRISPR/Cas9 gene disruption in Jurkat, CH12, and HEK293T cell lines:

For the generation of *FAM72A*$^{-/-}$ and *MAEA*$^{-/-}$ Jurkat cells, 4 x 10^6 cells were seeded 24 hours prior to transfection with the px459 V2.0 vector (Addgene, #62988) containing sgRNAs targeting exons of human *FAM72A-D* and *MAEA*. Cells were resuspended in serum-free OptiMEM medium (Thermo Fisher) and 20 µg of px459 before being transferred into a 4 mm electroporation cuvette. Cells were electroporated (BioRad Gene Pulser XCell) under the following conditions: 400 V, 300 µF, ∞ Ω). Electroporated cells were incubated on ice for 5 minutes before being transferred into pre-warmed complete RPMI. One day after electroporation, cells were selected with complete RPMI containing 0.5 µg/ml puromycin for 48 hours and allowed to recover prior to subcloning by limiting-dilution in 96-well plates. Knockout clones were identified by Sanger sequencing of PCR amplicons of the targeted exon and selected based on homozygous out-of-frame indels. Clones were further validated by Western blotting. *Fam72a*$^{-/-}$, *Mkln1*$^{-/-}$, and *Maea*$^{-/-}$ CH12 cells were previously generated [17]. *FAM72*$^{-/-}$ HEK293T and HCT116 cells were previously generated [22].

Retroviral transduction:

Retrovirus was generated as previously described [17]. In brief, 4 x 10^5 Bosc 23 packaging cells were seeded 24 hours prior to non-liposomal transfection with Opti-MEM medium, Fugene 6 (Fugent), pMX-PIE constructs containing the desired cDNA, and pCL-Amphotropic (Novus Biologicals, #NBP2-29541). The transfection mixture was incubated at room temperature for 15 minutes and added to the packaging cells, dropwise. Two days later, viral supernatants were filtered through a 0.45 um syringe filter prior to the addition of 10 µg/ml of polybrene (Sigma) and 20 mM HEPES (Fisher Scientific). 1 x 10^6 CH12 or LLC1 cells were resuspended with the viral supernatant and subjected to a room temperature spinfection for 1.5 hours at 1150 x *g*, without brakes. Cells were incubated at 37°C overnight and the media was replaced with complete RPMI the following day. Cells were selected with puromycin for 72 hours two days following infection.

Western blotting of cell lysates:

Western blotting was performed using standard protocols. To prepare lysates, cells were washed twice in PBS, resuspended in radioimmunoprecipitation assay (RIPA) buffer, and left on ice to lyse for 10 minutes. 5% of lysate was used to quantify total protein by BCA assay (Thermo Fisher), while 6x Laemmli buffer was added to the remaining lysate prior to boiling at 95°C for 10 minutes. Nitrocellulose membranes were probed with the following antibodies: anti-HA (Cell Signaling, 3724), anti-FLAG (Thermo Fisher, MA1-91878 and Sigma, F3165), anti-UNG (Abcam, 245630), anti-IST1 (Proteintech, 51002-1-AP), anti-HPF1 (Cell Signaling, 90876), anti-MKLN1 (Thermo Fisher, PA5-88356), anti-GAPDH (Cell Signaling, 2118), anti-α-tubulin (Abcam, ab4074), and anti-HSP90beta (Cell Signaling, D3F2, cat# 7411S). ImageJ software (NIH) was used for densitometry analysis.

Cycloheximide chase assay:

WT and *FAM72A*$^{-/-}$ Jurkat cells were seeded at a density of 6 x 10^5 cells /ml and treated with 10 µg/ml cycloheximide (CHX) for 0 (harvested immediately), 30, 60, 90, 120, and 180 minutes. At each timepoint, cells were washed twice with ice-cold PBS and pellets were snap-frozen in a dry ice slurry. Samples from each timepoint were lysed at the same time and total protein was quantified by BCA assay prior to SDS-PAGE and Western blotting.

Co-immunoprecipitation:

HCT116 cells were seeded at a density of 5.0 x 10^5 cells/mL and co-transfected 24 hours later with constructs containing wildtype or mutant HA-FAM72A and FLAG-UNG2 cDNA using the Lipofectamine 3000 reagent (Thermo Fisher Scientific). Four hours before harvesting, cells were treated with 10 uM MG132 (Sigma) in complete DMEM. Cells were washed three times in ice-cold PBS, mechanically detached, and lysed in non-ionic IP lysis buffer (40 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM EDTA pH 8.0, 10 mM NaF, 1% NP-40 substitute, 50 µM NEM, cOmplete mini protease inhibitor) for 1 hour rolling at 4°C.

CH12 cells retrovirally transduced with pMX-PIE constructs were seeded at a density of 1.0 x 10^6 cells/mL and treated with 20 uM of MG132 for 3 hours before harvesting. Cells were washed three times with PBS and lysed in non-ionic IP lysis buffer for 1 hour rolling at 4°C.

For input collection, 10% of each lysate was mixed with 6X Laemmli buffer, boiled, and stored at -80°C. In parallel, Protein G Sepharose 4 Fast Flow resin (Cytiva, 17061801) was washed twice

with IP wash buffer (40 mM Tris-HCL pH 7.5, 150 mM NaCl, 1 mM EDTA pH 8.0, 10 mM NaF, 0.5% NP-40 substitute, 5% glycerol), resuspended in IP lysis buffer, and incubated with HA (Cell Signaling, 3724) or FLAG antibody (Sigma, F3165) for 3 hours at 4°C. Antibody-resin complexes were added to pre-cleared lysates and incubated overnight at 4°C. The following day, the resin was washed 5 times in IP wash buffer, resuspended in 1X Laemmli buffer, boiled, and briefly pelleted before loading for SDS-PAGE and Western blotting.

Whole proteome analysis by liquid chromatography-mass spectrometry:

Whole protein was extracted from CH12 cells (WT, *Fam72a*$^{-/-}$, *Mkln1*$^{-/-}$) or LLC1 cells (pMX^{empty}, pMX^{Fam72a}) in triplicate and solubilized in SDS. Denatured proteins were then reduced, alkylated, and digested with trypsin using ProtiFi S-Trap micro spin columns.

For data-independent acquisition (DIA) LC-MS/MS of peptides from CH12 and LLC1 cells, 250ng of digested peptides were analyzed using a nano-HPLC (high-performance liquid chromatography) coupled to MS. 250 ng of the sample was loaded onto Evotip Pure per manufacturer instructions. Peptides were eluted from the column (cat#: EV-1137 for CH12; EV-1109 for LLC1, 15cm x 150µm with 1.5 µm beads for CH12; 8cm x 150µm with 1.5 µm beads for LLC1) with the 30SPD (CH12) or 60SPD (LLC1) pre-formed acetonitrile gradient generated by an Evosep One system, and analyzed on a timsTOF Pro 2. The total DIA protocol was 44 minutes (CH12) and 22 minutes (LLC1). The MS1 scan had a mass range of 100–1700Da in dia-PASEF. TIMS settings were accumulation and ramp time of 100ms, and within the mobility range (1/K0) of 0.6 to 1.6V·s/cm^2 (CH12) and 0.85 to 1.3V·s/cm^2 (LLC1). This was at a duty cycle of 100% and a ramp rate of 9.52Hz. 1+ ions are excluded from fragmentation using a polygonal filter. The auto-calibration was off.

Mass spectrometry data analysis:

Wildtype versus *Fam72a*$^{-/-}$ and *Mkln1*$^{-/-}$ CH12 proteomes, and pMX^{empty} versus pMX^{Fam72a} LLC1 proteomes were analyzed using the Spectronaut software (Biognosys) (**Supplementary data files 1, 2 and 3**). Spectronaut v18.7 directDIA+ workflow was used to search the data with the Spectronaut-generated mouse spectral library (Mouse_PDB_2023). The FASTA database used is the Uniprot UP000000589 reviewed mouse database with contaminants [52]. Parameters for the search were default. Run-wise imputation was turned on with global normalization. The differential abundance testing used was an unpaired t-test.

742 Data from the AP-MS of HA-FAM72A from CH12 cells were searched and analyzed using the
743 ProHits laboratory information management system (LIMS) platform in a previous publication [19].
744

References

1. Komander, D. & Rape, M. The ubiquitin code. *Annu Rev Biochem* **81**, 203-229 (2012).
2. Swatek, K.N. & Komander, D. Ubiquitin modifications. *Cell Res* **26**, 399-422 (2016).
3. Hershko, A. & Ciechanover, A. The ubiquitin system. *Annu Rev Biochem* **67**, 425-479 (1998).
4. Pickart, C.M. Mechanisms underlying ubiquitination. *Annu Rev Biochem* **70**, 503-533 (2001).
5. Rape, M. Ubiquitylation at the crossroads of development and disease. *Nat Rev Mol Cell Biol* **19**, 59-70 (2018).
6. Alpi, A.F., Chrustowicz, J., Sherpa, D. & Schulman, B.A. The many faces of the GID/CTLH E3 ligase complex. *Biochem Soc Trans* **53**, 1351-1364 (2025).
7. Sherpa, D. *et al.* GID E3 ligase supramolecular chelate assembly configures multipronged ubiquitin targeting of an oligomeric metabolic enzyme. *Mol Cell* **81**, 2445-2459 e2413 (2021).
8. Qiao, S. *et al.* Interconversion between Anticipatory and Active GID E3 Ubiquitin Ligase Conformations via Metabolically Driven Substrate Receptor Assembly. *Mol Cell* **77**, 150-163 e159 (2020).
9. Chen, S.J., Wu, X., Wadas, B., Oh, J.H. & Varshavsky, A. An N-end rule pathway that recognizes proline and destroys gluconeogenic enzymes. *Science* **355** (2017).
10. Dong, C. *et al.* Recognition of nonproline N-terminal residues by the Pro/N-degron pathway. *Proc Natl Acad Sci U S A* **117**, 14158-14167 (2020).
11. Dong, C. *et al.* Molecular basis of GID4-mediated recognition of degrons for the Pro/N-end rule pathway. *Nat Chem Biol* **14**, 466-473 (2018).
12. Owens, D.D.G. *et al.* A chemical probe to modulate human GID4 Pro/N-degron interactions. *Nat Chem Biol* **20**, 1164-1175 (2024).
13. Yazdi, A.K. *et al.* Chemical tools for the Gid4 subunit of the human E3 ligase C-terminal to LisH (CTLH) degradation complex. *RSC Med Chem* **15**, 1066-1071 (2024).
14. Li, Y. *et al.* Design of PROTACs utilizing the E3 ligase GID4 for targeted protein degradation. *Nat Struct Mol Biol* **32**, 1825-1837 (2025).
15. Chana, C.K. *et al.* The molecular glue CLEO4-88 inhibits the ACAA1 thiolase by induced binding to GID4. *Nat Chem Biol* **22**, 1589-1598 (2026).
16. Chana, C.K. *et al.* Discovery and Structural Characterization of Small Molecule Binders of the Human CTLH E3 Ligase Subunit GID4. *J Med Chem* **65**, 12725-12746 (2022).
17. Feng, Y. *et al.* FAM72A antagonizes UNG2 to promote mutagenic repair during antibody maturation. *Nature* **600**, 324-328 (2021).
18. Rogier, M. *et al.* Fam72a enforces error-prone DNA repair during antibody diversification. *Nature* **600**, 329-333 (2021).
19. Barbulescu, P. *et al.* FAM72A degrades UNG2 through the GID/CTLH complex to promote mutagenic repair during antibody maturation. *Nat Commun* **15**, 7541 (2024).

20. Stewart, J.A. & Bhagwat, A.S. A redox-sensitive iron-sulfur cluster in murine FAM72A controls its ability to degrade the nuclear form of uracil-DNA glycosylase. *DNA Repair (Amst)* **118**, 103381 (2022).
21. Krokan, H.E., Nilsen, H., Skorpen, F., Otterlei, M. & Slupphaug, G. Base excision repair of DNA in mammalian cells. *FEBS Lett* **476**, 73-77 (2000).
22. Feng, Y. *et al.* FAM72A promotes UNG2 degradation and mutagenesis in human cancer cells. *Sci Rep* **15**, 23467 (2025).
23. Sepic, S. *et al.* Multimodal substrate recruitment enables CTLH(MKLN1) E3 ligase to target acetylated N-, C-, and internal degrons. *Nat Commun* **17** (2026).
24. Xiao, Q., Zhang, F., Nacev, B.A., Liu, J.O. & Pei, D. Protein N-terminal processing: substrate specificity of Escherichia coli and human methionine aminopeptidases. *Biochemistry* **49**, 5588-5599 (2010).
25. Yi, S.A., Sepic, S., Schulman, B.A., Ordureau, A. & An, H. mTORC1-CTLH E3 ligase regulates the degradation of HMG-CoA synthase 1 through the Pro/N-degron pathway. *Mol Cell* **84**, 2166-2184 e2169 (2024).
26. Jordan, V.N., Ordureau, A. & An, H. Identifying E3 Ligase Substrates With Quantitative Degradation Proteomics. *Chembiochem* **24**, e202300108 (2023).
27. Agromayor, M. *et al.* Essential role of hIST1 in cytokinesis. *Mol Biol Cell* **20**, 1374-1387 (2009).
28. Rudolph, J. & Luger, K. Histone PARylation factor 1: a review of its role in the DNA damage response. *Nucleic Acids Res* **53** (2025).
29. Van Damme, P., Hole, K., Gevaert, K. & Arnesen, T. N-terminal acetylome analysis reveals the specificity of Naa50 (Nat5) and suggests a kinetic competition between N-terminal acetyltransferases and methionine aminopeptidases. *Proteomics* **15**, 2436-2446 (2015).
30. Gamerdinger, M. & Deuerling, E. Cotranslational sorting and processing of newly synthesized proteins in eukaryotes. *Trends Biochem Sci* **49**, 105-118 (2024).
31. Vietri, M. *et al.* Spastin and ESCRT-III coordinate mitotic spindle disassembly and nuclear envelope sealing. *Nature* **522**, 231-235 (2015).
32. Frankel, E.B. *et al.* Ist1 regulates ESCRT-III assembly and function during multivesicular endosome biogenesis in Caenorhabditis elegans embryos. *Nat Commun* **8**, 1439 (2017).
33. Bajorek, M. *et al.* Biochemical analyses of human IST1 and its function in cytokinesis. *Mol Biol Cell* **20**, 1360-1373 (2009).
34. Fu, Y. *et al.* Fam72a functions as a cell-cycle-controlled gene during proliferation and antagonizes apoptosis through reprogramming PP2A substrates. *Dev Cell* **58**, 398-415 e397 (2023).
35. Liu, C., Vyas, A., Kassab, M.A., Singh, A.K. & Yu, X. The role of poly ADP-ribosylation in the first wave of DNA damage response. *Nucleic Acids Res* **45**, 8129-8141 (2017).
36. Martin-Hernandez, K., Rodriguez-Vargas, J.M., Schreiber, V. & Dantzer, F. Expanding functions of ADP-ribosylation in the maintenance of genome integrity. *Semin Cell Dev Biol* **63**, 92-101 (2017).

37. Gibbs-Seymour, I., Fontana, P., Rack, J.G.M. & Ahel, I. HPF1/C4orf27 Is a PARP-1-Interacting Protein that Regulates PARP-1 ADP-Ribosylation Activity. *Mol Cell* **62**, 432-442 (2016).
38. Smith, R. *et al.* HPF1-dependent histone ADP-ribosylation triggers chromatin relaxation to promote the recruitment of repair factors at sites of DNA damage. *Nat Struct Mol Biol* **30**, 678-691 (2023).
39. Larocque, D. *et al.* Nuclear retention of MBP mRNAs in the quaking viable mice. *Neuron* **36**, 815-829 (2002).
40. Conn, S.J. *et al.* The RNA binding protein quaking regulates formation of circRNAs. *Cell* **160**, 1125-1134 (2015).
41. Neumann, D.P., Goodall, G.J. & Gregory, P.A. The Quaking RNA-binding proteins as regulators of cell differentiation. *Wiley Interdiscip Rev RNA* **13**, e1724 (2022).
42. Bandopadhayay, P. *et al.* MYB-QKI rearrangements in angiocentric glioma drive tumorigenicity through a tripartite mechanism. *Nat Genet* **48**, 273-282 (2016).
43. Zong, F.Y. *et al.* The RNA-binding protein QKI suppresses cancer-associated aberrant splicing. *PLoS Genet* **10**, e1004289 (2014).
44. Yang, G. *et al.* RNA-binding protein quaking, a critical regulator of colon epithelial differentiation and a suppressor of colon cancer. *Gastroenterology* **138**, 231-240 e231-235 (2010).
45. Guo, C. *et al.* Ugene, a newly identified protein that is commonly overexpressed in cancer and binds uracil DNA glycosylase. *Cancer Res* **68**, 6118-6126 (2008).
46. Kong, A.T., Leprevost, F.V., Avtonomov, D.M., Mellacheruvu, D. & Nesvizhskii, A.I. MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nat Methods* **14**, 513-520 (2017).
47. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol* **75**, 861-877 (2019).
48. Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493-500 (2024).
49. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501 (2010).
50. Morin, A. *et al.* Collaboration gets the most out of software. *Elife* **2**, e01456 (2013).
51. Haakonsen, D.L. *et al.* Stress response silencing by an E3 ligase mutated in neurodegeneration. *Nature* **626**, 874-880 (2024).
52. Frankenfield, A.M., Ni, J., Ahmed, M. & Hao, L. Protein Contaminants Matter: Building Universal Protein Contaminant Libraries for DDA and DIA Proteomics. *J Proteome Res* **21**, 2104-2113 (2022).

Acknowledgements

We thank the Sicheri and Martin labs for helpful discussion. The authors wish to thank Laura McGary and Bhavisha Rathod of the Network Biology Collaborative Centre Proteomics Facility (RRID: SCR 025375) at the Lunenfeld-Tanenbaum Research Institute for Mass Spectrometry Analysis. The facility is supported by the Canada Foundation for Innovation and the Ontario Government. This work was supported by grants from the Canadian Institutes of Health Research grants PJT-180269 to AM and PJT-178026, PJT-180338 and PJT- 186218 to FS.

Author Contributions

Conceptualization: AY, LB, CC, DLH, AM, FS

Investigation: AY, LB, SM, DM, CC

Supervision: AM, FS, DLH

Writing – original draft: AY, LB, FS, AM, DLH,

Competing Interests

FS is a founder and consultant for Induxion Therapeutics.

Data for this manuscript will be provided upon request.

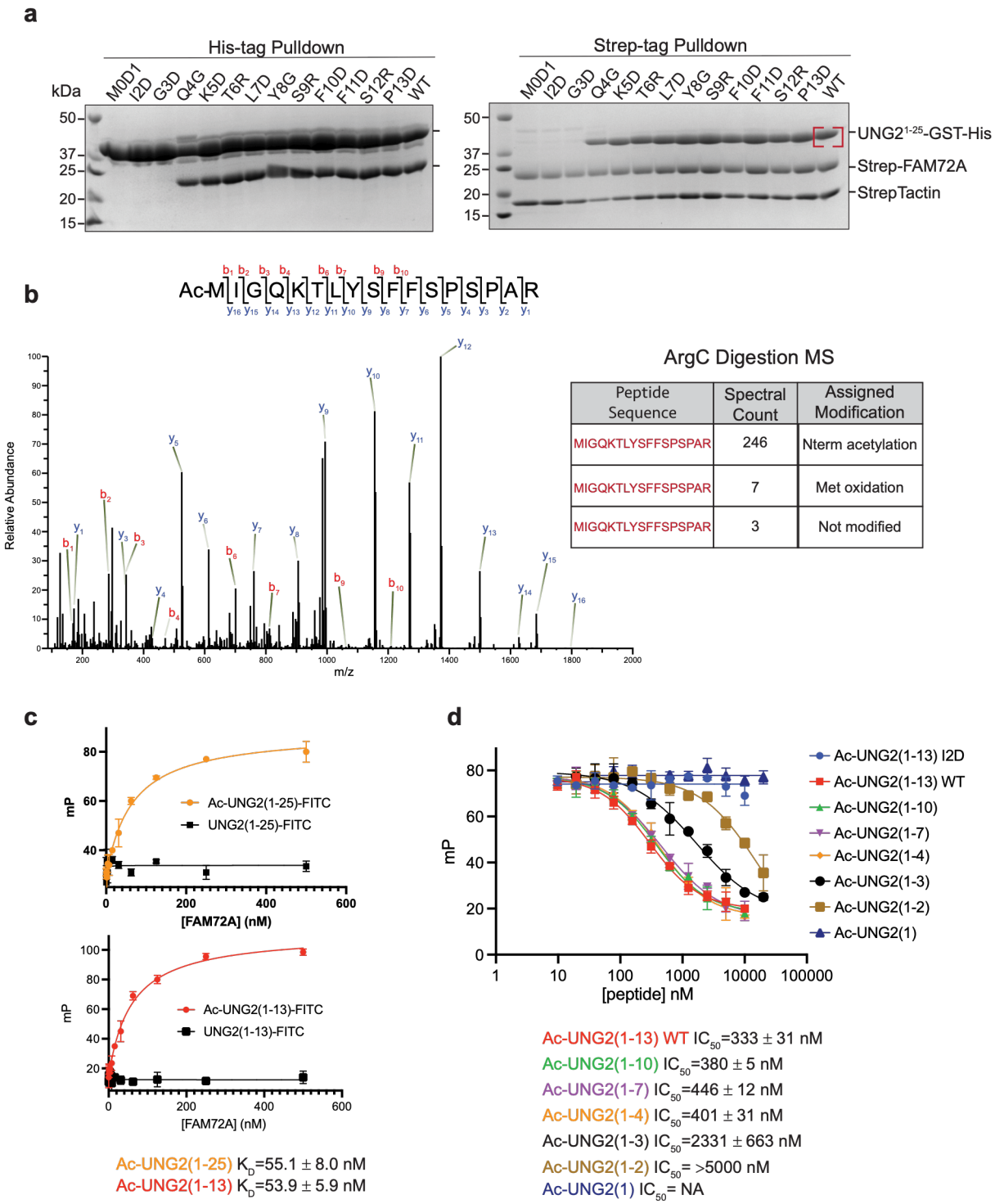
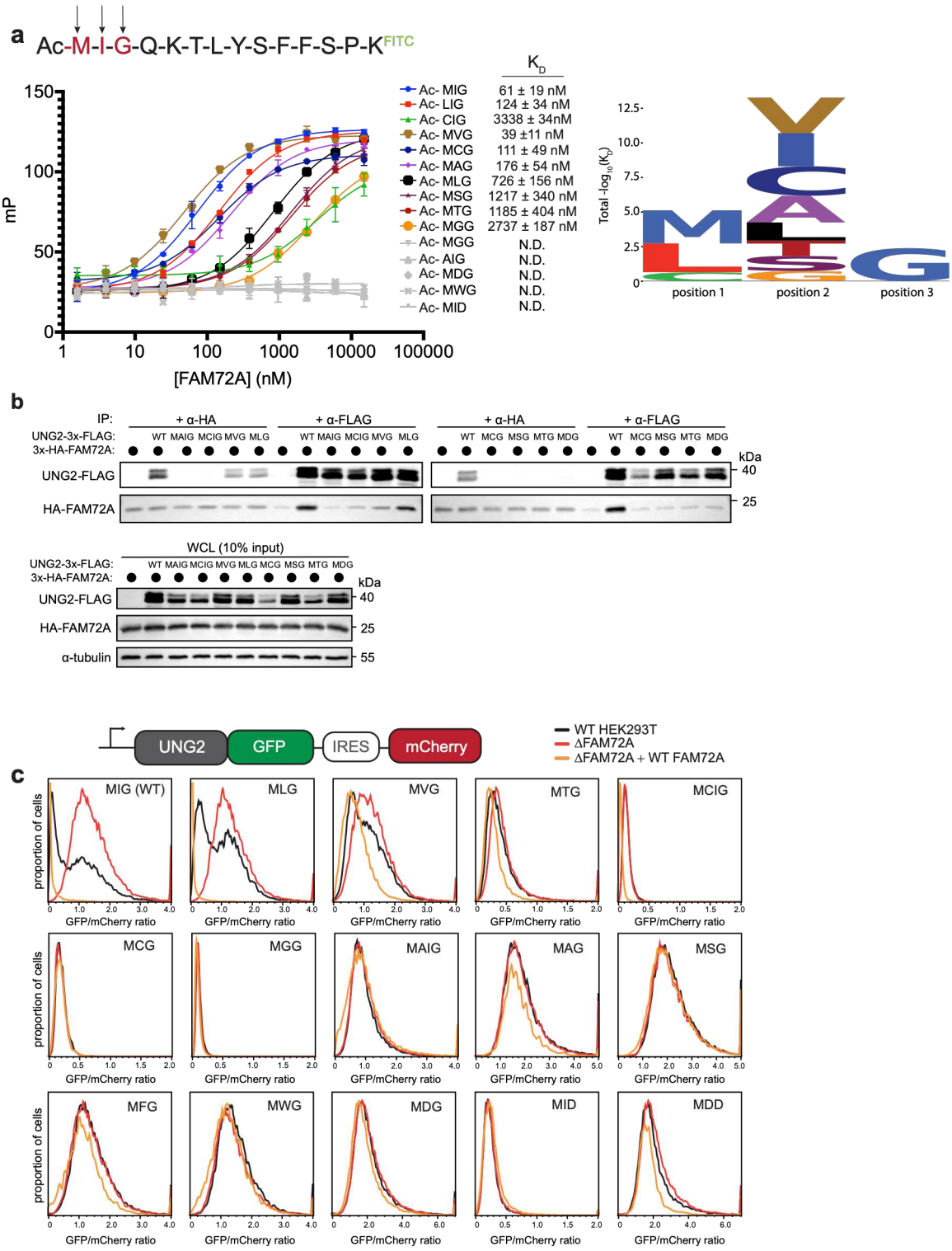


Fig. 1: FAM72A recognizes an N-terminal, acetylated degron on UNG2.

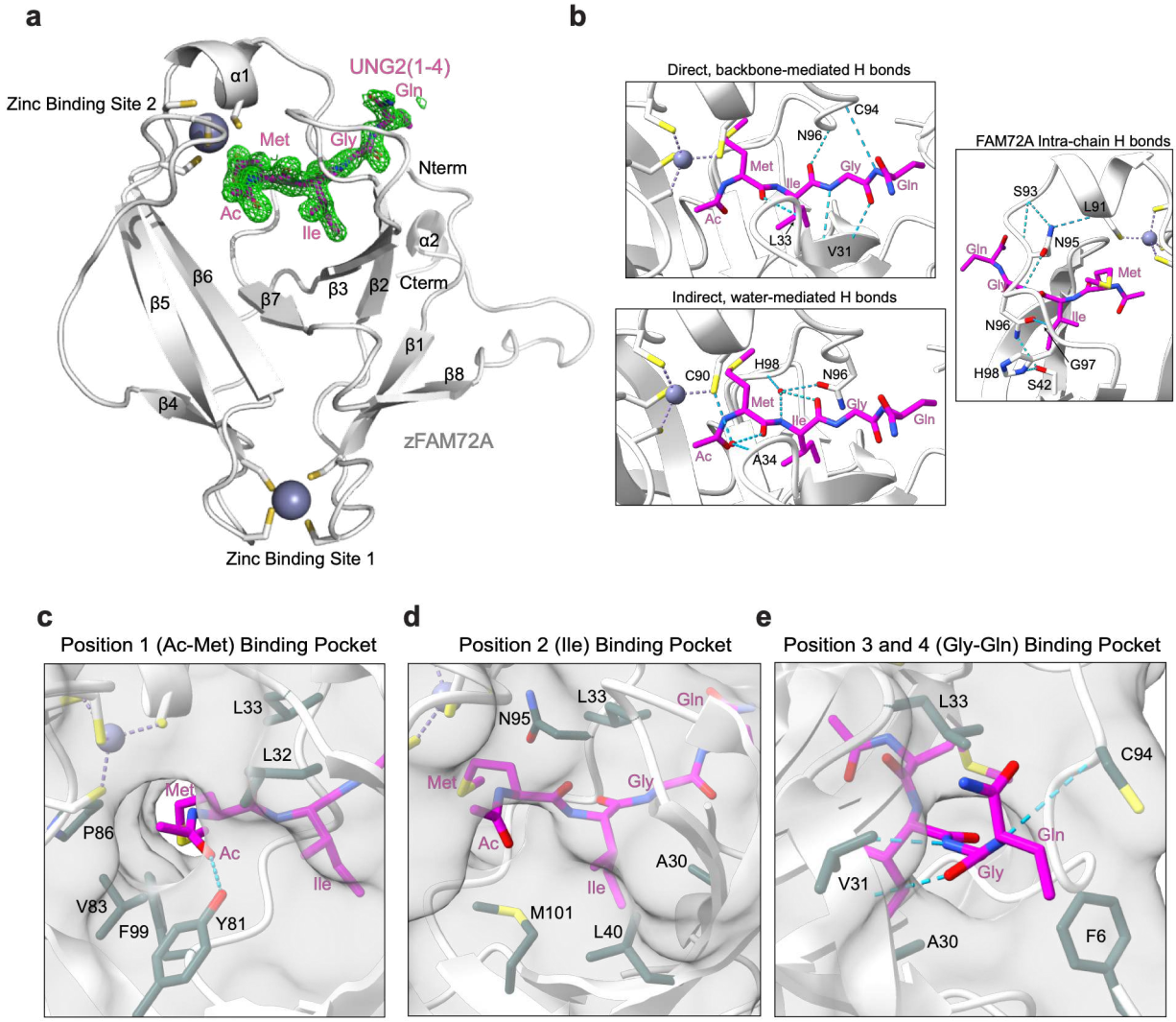
a Reciprocal His_{6x}-tag and Strep-Tag II pull-downs from Sf9 lysate expressing the indicted proteins, followed by visualization by Coomassie-stained SDS-PAGE. n=1 independent experiment. **b** ArgC-digestion mass spectrometry of $UNG2^{1-25}$-GST-His excised from the band boxed in red in **a**. n=1 independent experiment. See **Supplementary Table 1** for associated fragment intensities. **c** Fluorescence polarization binding analysis of purified hFAM72A to the indicated FITC-labelled peptide corresponding to the 25 and 13 N-terminal residues of UNG2. Points represent the mean ± s.d., n=3 independent experiments. **d** Competitive displacement of Ac-UNG2(1-13)-FITC binding to hFAM72A by the indicated unlabelled UNG2 peptide. Points represent the mean ± s.d., n=3 independent experiments.



907

908 **Fig. 2: FAM72A recognizes a 3 amino acid degron consensus.**

a (left) Fluorescence polarization binding analysis of purified hFAM72A to the indicated N-terminally acetylated, FITC-labelled peptides. Select non-binders are shown as light grey traces. Schematic of the peptide series shown at top. K_d values for the strongest binders shown at far right, all other K_d >>> 15μM. Points represent the mean ± s.d., n=2 independent experiments. See **Supplementary Fig. 1a** for profiles of other non-binders. **(right)** Sequence logo representation of the biochemical degron binding consensus of FAM72A. **b** Reciprocal co-immunoprecipitation of HA-tagged FAM72A and FLAG-tagged wildtype or mutant UNG2 constructs from lysates of *FAM72A*$^{-/-}$ HCT116 cells transiently transfected with the indicated constructs. Cells were treated with 10 μM MG132 4 hours prior to harvesting (n=1 independent experiments). **c** Wildtype and mutant UNG2 stability was monitored by flow cytometry in wild-type HEK293T, FAM72A KO HEK293T, and FAM72A KO HEK293T cells overexpressing WT hFAM72A. Similar results in n=2 independent experiments.



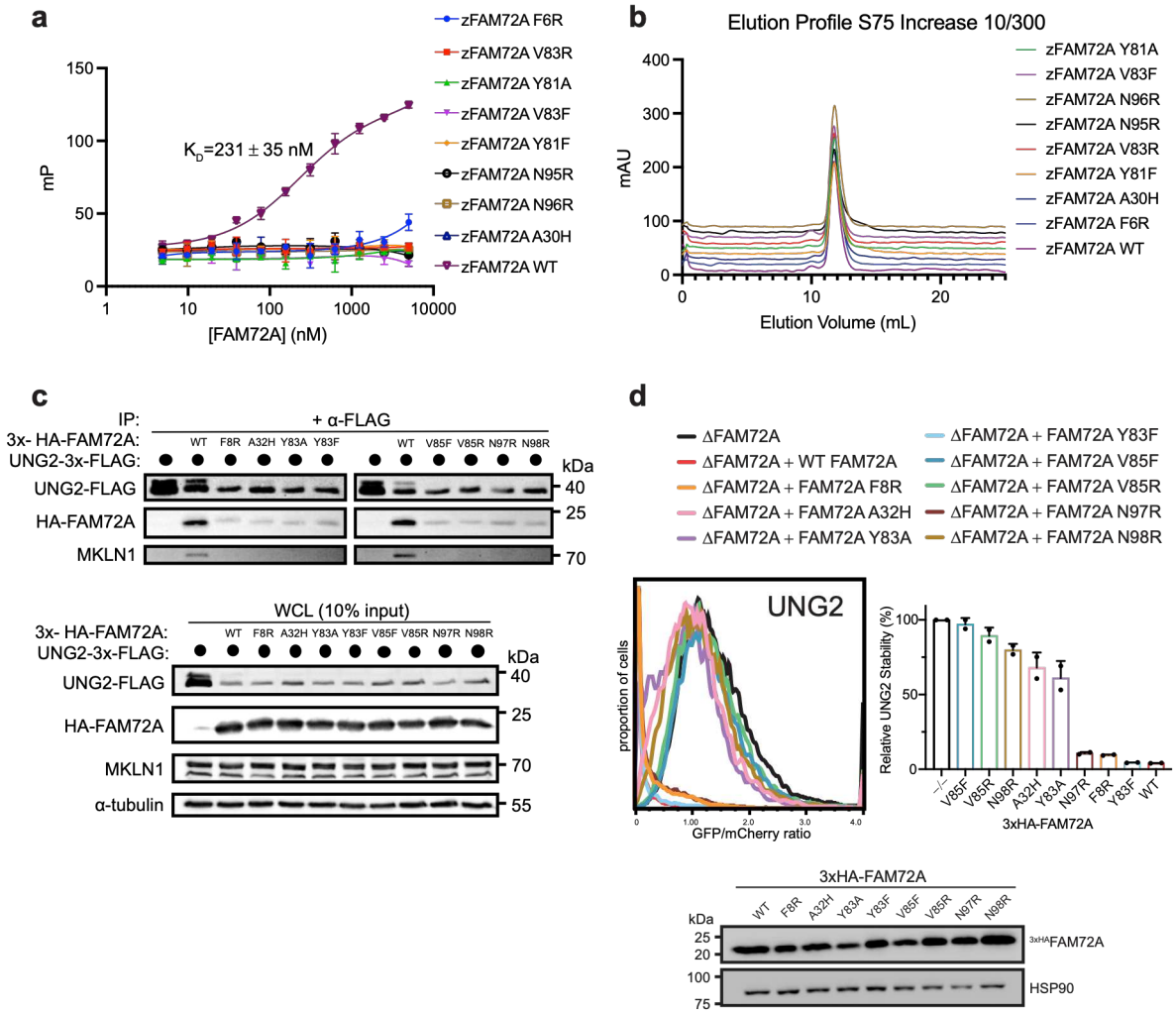
923

924 **Fig. 3: X-ray crystal structure of zebrafish FAM72A in complex with an UNG2(1-4) degron.**

925 **a** 1.8 Å x-ray crystal structure of zFAM72A (grey) in complex with the UNG2(1-4) degron peptide
926 (magenta). Electron density for the peptide (Fo-Fc map) was contoured at ±3σ and shown in green.
927 Zn^{2+} ions are show as gray spheres. **b** Direct and water bridged hydrogen bonds that stabilize
928 degron binding and the degron binding pocket. Hydrogen bonds represented as blue dashed lines.
929 **c, d, e** Detailed views of the molecular interactions involving degron position 1 (Ac-Met), position
930 2 (Ile) and positions 3 and 4 (Gly-Gln), respectively.

931

932



933

934 **Fig. 4: Mutational analysis of the FAM72A degron binding pocket.**

935 **a** Fluorescence polarization binding analysis of the indicated purified wildtype and mutant
936 zFAM72A proteins to the FITC-labelled Ac-UNG2(1-13) peptide. Points represent the mean ± s.d.,
937 n=3 independent experiments. **b** Size exclusion elution profile of the indicated wild type and
938 mutant zFAM72A proteins. **c** Co-immunoprecipitation of FLAG-tagged UNG2 from lysates of
939 *FAM72A*$^{-/-}$ HCT116 cells transiently transfected with the indicated HA-tagged FAM72A mutant
940 constructs and wildtype FLAG-UNG2. Cells were treated with 10 μM MG132 4 hours prior to
941 harvesting (n=1 independent experiments). **d** Wildtype UNG2 stability was monitored by flow
942 cytometry in FAM72 KO HEK293T cells overexpressing the indicated hFAM72A protein.

Relative UNG2 stability was quantified from the median GFP/mCherry ratio for each distribution. Data are presented as mean ± sd from two independent experiments (n=2). Western blot analysis of FAM72A protein expression levels shown at bottom.

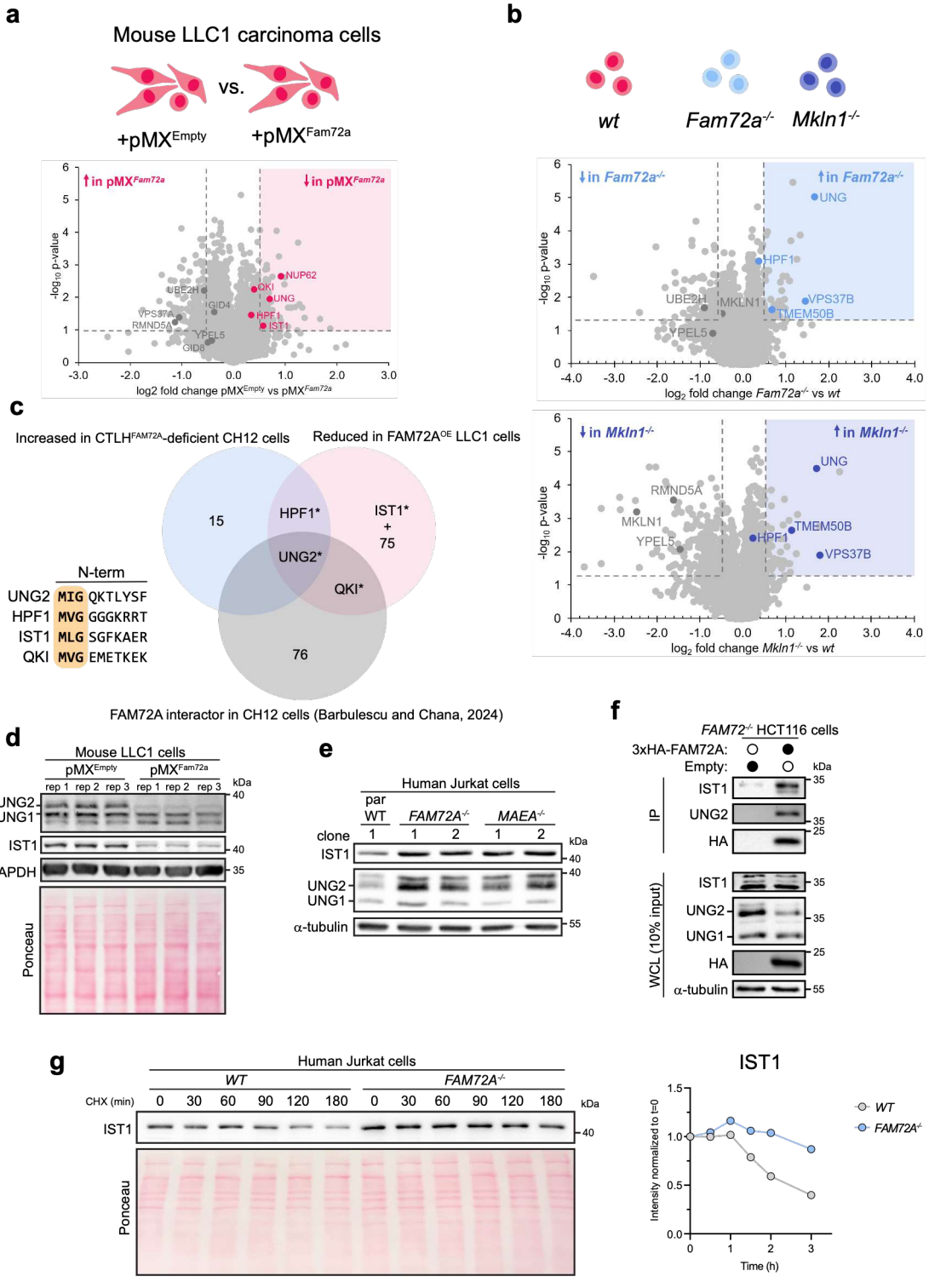


Figure 5. Label-free proteomics identifies novel substrates of FAM72A-MKLN1-CTLH.

a. Whole proteome mass spectrometry of mouse LLC1 cells retrovirally transduced to over-express FAM72A (pMX^{Fam72a}) or transduced with an empty vector control (pMX^{Empty}) in triplicate. Proteins containing the consensus FAM72A degron that were reduced in the pMXFam72a condition and subsequently validated are shaded in red. Proteins were selected for further analysis using a fold-change cutoff of ≥1.4 and a nominal p-value cutoff of <0.08. **b.** Whole proteome mass spectrometry of wildtype versus *Fam72a*$^{-/-}$ mouse CH12 cells (upper) and wildtype versus *Mkln1*$^{-/-}$ CH12 cells (lower) in triplicate. Proteins containing the consensus FAM72A degron that were increased in *Fam72a*$^{-/-}$ or *Mkln1*$^{-/-}$ cells and subsequently validated are shaded in light blue or dark blue, respectively. Proteins were considered significantly altered using a fold-change cutoff of ≥1.5 and a p-value cutoff of <0.05. **c.** Venn diagram (left) summarizing candidate FAM72A substrates identified in a and b, with corresponding mouse N-terminal sequences shown on the left. Asterisks (*) denote proteins containing the consensus degron. **d.** Western blotting of UNG and IST1 from pMX^{Empty} and pMX^{Fam72a} LLC1 whole cell lysates using GAPDH as control. **e.** Western blotting of IST1 and UNG from lysates of WT parental (par WT), *FAM72A*$^{-/-}$, and *MAEA*$^{-/-}$ human Jurkat T cells **f.** Co-immunoprecipitation of HA-tagged FAM72A from lysates of *FAM72*$^{-/-}$ HCT116 cells transiently transfected with 3x-HA-FAM72A or an empty vector control. Cells were treated with 10 μM MG132 4 hours prior to harvesting. **g.** WT and *FAM72A*$^{-/-}$ Jurkat cells were treated with cycloheximide (CHX) for the indicated time (minutes) and lysed for Western blotting of IST1 protein (left). Densitometry analysis of IST1 band intensity over time (right), normalized to the t = 0 min time point for each genotype.

Data Collection	
Wavelength	1.54184 Å
Resolution range	22.8 - 1.8 (1.91 - 1.8)
Space group	P 65 2 2
Unit cell	76.729 76.729 108.795 90 90 120
Total reflections	302620 (42790)
Unique reflections	18163 (2945)
Multiplicity	16.7 (14.5)
Completeness	99.90 (99.97)
Mean I/sigma(I)	26.14 (8.49)
Wilson B-factor	12.36
R-merge	0.09036 (0.2803)
R-meas	0.09322 (0.2905)
R-pim	0.02255 (0.07555)
CC1/2	0.999 (0.975)
CC*	1 (0.994)
Refinement	
Reflections used in refinement	18148 (2945)
Reflections used for R-free	906 (147)
R-work	0.1981 (0.2105)
R-free	0.2263(0.2349)
Number of non-hydrogen atoms	1053
macromolecules	955
ligands	11
solvent	87
Protein residues	125
RMS(bonds)	0.009
RMS(angles)	1.10
Ramachandran favored (%)	97.52
Ramachandran allowed (%)	2.48
Ramachandran outliers (%)	0.00
Rotamer outliers (%)	0.00
Clashscore	1.05
Average B-factor	15.72
macromolecules	14.87
ligands	32.09
solvent	22.96

973

974 **Table 1. X-ray crystallographic data collection and refinement statistics**

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTable2.xlsx
- SupplementaryTable3V3.pdf
- ReportingSummary.pdf
- SupplementaryDataFile2CH12MKLN1KOproteomics.xlsx
- CombinedSupplementaryFiguresV2.pdf
- SupplementaryDataFile1CH12FAM72AKOproteomics.xlsx
- SupplementaryDataFile3LLC1FAM72Aoverexpressionproteomics.xlsx
- 39LZPDBValidationReport.pdf