nature communications

Article

https://doi.org/10.1038/s41467-024-52009-x

FAM72A degrades UNG2 through the GID/CTLH complex to promote mutagenic repair during antibody maturation

Received: 24 November 2023

Accepted: 23 August 2024

Published online: 30 August 2024

Check for updates

Philip Barbulescu[1,9], Chetan K. Chana[2,3,9], Matthew K. Wong[1], Ines Ben Makhlouf[4], Jeffrey P. Bruce[5], Yuqing Feng[1], Alexander F. A. Keszei[5], Cassandra Wong[2,6], Rukshana Mohamad-Ramshan[6], Laura C. McGary[2,6], Mohammad A. Kashem[1], Derek F. Ceccarelli[3], Stephen Orlicky[3], Yifei Fang[1], Huihui Kuang[7], Mohammad Mazhab-Jafari[5], Rossanna C. Pezo[8], Ashok S. Bhagwat[6], Trevor J. Pugh[5], Anne-Claude Gingras[3,4], Frank Sicheri[2,3,4,10] ✉ & Alberto Martin[1,10] ✉

A diverse antibody repertoire is essential for humoral immunity. Antibody diversification requires the introduction of deoxyuridine (dU) mutations within *immunoglobulin* genes to initiate somatic hypermutation (SHM) and class switch recombination (CSR). dUs are normally recognized and excised by the base excision repair (BER) protein uracil-DNA glycosylase 2 (UNG2). However, FAM72A downregulates UNG2 permitting dUs to persist and trigger SHM and CSR. How FAM72A promotes UNG2 degradation is unknown. Here, we show that FAM72A recruits a C-terminal to LisH (CTLH) E3 ligase complex to target UNG2 for proteasomal degradation. Deficiency in CTLH complex components result in elevated UNG2 and reduced SHM and CSR. Cryo-EM structural analysis reveals FAM72A directly binds to MKLN1 within the CTLH complex to recruit and ubiquitinate UNG2. Our study further suggests that FAM72A hijacks the CTLH complex to promote mutagenesis in cancer. These findings show that FAM72A is an E3 ligase substrate adaptor critical for humoral immunity and cancer development.

Antibody diversification is essential to protect vertebrates from pathogens. Following V(D)J recombination, the antibody repertoire is expanded through secondary antibody diversification events including somatic hypermutation (SHM) and class switch recombination (CSR). Mature B cells undergo SHM to increase antibody affinity and CSR to confer different effector functions to the expressed antibody. Activation-induced cytidine deaminase (AID) initiates these processes by deaminating dC to dU within the *Immunoglobulin (Ig)* genes[1,2]. dU lesions require recognition and processing through base excision repair (BER) or mismatch repair (MMR) pathways that function to expand the mutational signature within the *Ig* genes in B cells[3–8]. How these DNA repair pathways are subverted from faithfully repairing DNA to producing mutations has been the subject of research for decades.

[1]Department of Immunology, University of Toronto, Toronto, ON, Canada. [2]Department of Biochemistry, University of Toronto, Toronto, ON, Canada. [3]Lunenfeld-Tanenbaum Research Institute, Sinai Health System, Toronto, ON, Canada. [4]Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada. [5]Princess Margaret Cancer Centre, University Health Network, University of Toronto, Toronto, ON, Canada. [6]Department of Chemistry, Wayne State University, Detroit, MI, USA. [7]Cryo-Electron Microscopy Core, New York University School of Medicine, New York, NY, USA. [8]Sunnybrook Health Sciences Center, Toronto, ON, Canada. [9]These authors contributed equally: Philip Barbulescu, Chetan K. Chana. [10]These authors jointly supervised this work: Frank Sicheri, Alberto Martin. ✉e-mail: sicheri@lunenfeld.ca; alberto.martin@utoronto.ca

Genomic dU lesions are typically recognized and excised by UNG2 of the BER pathway to facilitate DNA repair. However, in germinal center B cells, FAM72A is expressed which promotes the degradation of UNG2[9,10]. Reduced BER efficiency thereby allows AID-generated dU lesions in *Ig* genes to persist. By doing so, dU•dG mismatches are instead processed through mismatch repair (MMR) towards mutagenic outcomes that contributes to SHM and CSR[9–11]. Although FAM72A has been shown to interact with UNG2[9,10], exactly how FAM72A promotes the degradation of UNG2 remained unknown. As will be discussed below, we found that FAM72A promotes degradation of UNG2 through a GID/CTLH E3 ubiquitin ligase complex (referred to as CTLH complex hereafter).

CTLH complexes are modular multi-subunit E3 ligases implicated in the regulation of diverse eukaryotic biology, including glucose metabolism in yeast[12], AMPK activity[13], cell cycle regulation[14], cell migration[15], Hedgehog signaling[16], and erythropoiesis[17] in higher vertebrates. The best characterized CTLH E3, which is conserved from yeast to human, consists of the scaffolding proteins RANBP9/10, TWA1, WDR26, and ARMC8A, the catalytic subunits RMND5A/B and MAEA, and the substrate receptor GID4 (GID1, GID8, GID7 GID8, GID5, GID2, GID9 and GID4 respectively in yeast). GID4 recruits substrate by recognizing degenerate linear N-terminal sequences (termed N-degrons) commencing with a hydrophobic residue[18–20]. Apart from GID4, no additional CTLH substrate receptors have been identified in humans, yet the expansive repertoire of potential substrates suggests that multiple substrate adapters exist. Relative to yeast, added complexity to human CTLH E3s arises from the presence of subunit paralogs (e.g., RANBP9/10, RMND5A/B, and ARMC8A/B), including the presence of a more distant paralog of WDR26 called MKLN1 with similar domain architecture but different quaternary structure. Interestingly, the degradation of some inferred CTLH E3 substrates is dependent on either MKLN1 or WDR26 but not GID4[21,22], which implies that MKLN1 and WDR26 may themselves play a role in substrate recruitment. However, neither has been shown to engage substrates directly.

Here, we report that FAM72A is a substrate adapter of a CTLH E3 containing MKLN1 together with RANBP9, TWA1, RMND5A, and MAEA (termed CTLH-MKLN1^{FAM72A}). In cultured cells and mouse models, we show that CTLH-MKLN1^{FAM72A} is required for the secondary antibody diversification processes SHM and CSR, through the targeted degradation of UNG2. We reconstitute in vitro ubiquitination of UNG2 using purified CTLH-MKLN1^{FAM72A} and use cryogenic electron microscopy (cryo-EM) to show that CTLH-MKLN1^{FAM72A} forms a ring-shaped structure with MKLN1 in a dimer configuration. Each of the two MKLN1 dimers directly binds and orients two FAM72A molecules towards the center of the ring, analogous to the architecture of CTLH-WDR26^{GID4}. Comparison of the dimeric structure of MKLN1 within CTLH-MKLN1^{FAM72A} and the tetrameric structure of MKLN1 in isolation reveals that tetramerization occludes the binding surface for FAM72A, and thus represents an auto-inhibited state. Lastly, we present data that suggests a broader role for CTLH-MKLN1^{FAM72A} as a promoter of mutagenic DNA repair, which threatens genomic stability. Indeed, our analysis shows a correlation between *FAM72A* overexpression and mutation rates in human cancers.

Results

CTLH-deficiency impairs CSR and induces UNG2 accumulation

Since deficiency in an E3 ubiquitin ligase could inhibit UNG2 proteasomal degradation and thus cause a defect in CSR, we reasoned that gRNAs targeting the E3 ligase responsible for UNG2 degradation might be depleted in the IgA-positive fraction from our CRISPR-Cas9 screen in the murine CH12 B cell line performed previously[9]. This cell line is unique in its ability to undergo IgM to IgA CSR following stimulation with αCD40, IL-4, and TGFβ (CIT) in vitro. We noted that gRNAs targeting five of the ten known members of the CTLH complex ranked within the top 10% of our screen (19193 genes total), including the central catalytic E3 *Maea* ligase, *Mkln1*, *Wdr26*, *Twa1/Gid8*, and *Ube2h* (Fig. 1a, b)[9]. Separately, two CTLH subunits, *Mkln1* and *Twa1/Gid8*, appeared as top 100 hits in an independent CRISPR-Cas9 screen[10].

To confirm that MAEA and a CTLH complex are responsible for FAM72A-mediated degradation of UNG2, we generated *Fam72a*$^{-/-}$, *Maea*$^{-/-}$, and *Fam72a*$^{-/-}$*Maea*$^{-/-}$ CH12 clones by CRISPR/Cas9 (Supplementary Figs. 1a, b; and Supplementary Fig. 2a) and found that all had similar defects in CSR following 3-days of CIT stimulation (Fig. 1c) suggesting that these genes are epistatic. In addition, UNG2 protein levels were increased in all mutants (Fig. 1d). We also observed increased UNG2 protein in human BJAB B cells and Jurkat T cells that were deficient in MAEA (Supplementary Fig. 1c, d; and Supplementary Fig. 2b, c). In parallel, we observed increased protein levels of the CTLH subunit MKLN1 in *Maea*$^{-/-}$ but not *Fam72a*$^{-/-}$ conditions for both human and mouse B cell lines (Supplementary Fig. 1b, c, e). Indeed, previous work showed that MKLN1 is degraded by CTLH complexes[23]. To test whether the catalytic function of CTLH E3s are required for CSR, we re-introduced MAEA and a MAEA^{C314S} mutant with a catalytically inactive RING domain into *Maea*$^{-/-}$ and *Fam72a*$^{-/-}$*Maea*$^{-/-}$ CH12 cells. Re-expression of MAEA in *Maea*$^{-/-}$ CH12 cells returned the protein levels of UNG2 and MKLN1 comparable to *wildtype* cells, rescuing CSR. Similarly, double reconstitution of FAM72A and MAEA in *Fam72a*$^{-/-}$*Maea*$^{-/-}$ CH12 cells produced a modest rescue of CSR compared to FAM72A + MAEA^{C314S} reconstitution. (Fig. 1e; and Supplementary Figs. 1e, f). Consistent with the increased UNG2 levels, *Maea*$^{-/-}$ and *Fam72a*$^{-/-}$ CH12 cells both had reduced genomic dU compared to wildtype CH12 cells in CIT-treated conditions (Supplementary Fig. 1h).

Deletion of MAEA did not affect AID protein levels or germline transcripts in CH12 cells (Supplementary Figs. 3a, b), but it did cause a proliferation defect (Supplementary Fig. 3c) that was due to increased apoptosis rather than an effect on cell cycle progression (Supplementary Fig. 3d–f)[14]. As a proliferation defect was not observed in *Fam72a*$^{-/-}$ CH12 cells[9,10], these results suggest that MAEA may have additional roles in B cell biology separate from UNG2 degradation.

To determine whether other members of the CTLH complex are required for CSR and UNG2 degradation, we generated CH12 knockouts for specific CTLH complex subunits belonging to the catalytic core (*Rmnd5a*$^{-/-}$, *Rmnd5b*$^{-/-}$, *Rmnd5a*$^{-/-}$*Rmnd5b*$^{-/-}$, *Maea*$^{-/-}$)[14,24], the scaffolding subunits (*Wdr26*$^{-/-}$, *Mkln1*$^{-/-}$, *Armc8α/β*$^{-/-}$, *Twa1*$^{-/-}$, *Ranbp9*$^{-/-}$, *Ranbp10*$^{-/-}$, *Ranbp9*$^{-/-}$*Ranbp10*$^{-/-}$)[14], subunits with unknown function (*Ypel5*$^{-/-}$)[14,25,26], the substrate adaptor (*Gid4*$^{-/-}$)[27], and the cognate E2 ubiquitin conjugating enzyme (*Ube2H*$^{-/-}$)[28–30] (Supplementary Fig. 2a). We observed CSR defects, along with UNG2 and MKLN1 accumulation when the catalytic core of CTLH E3s or the UBE2H E2 ubiquitin conjugating enzyme were disrupted in CH12 cells (Fig. 1f, g; and Supplementary Figs. 4a, b). *Rmnd5a*$^{-/-}$ and *Rmnd5b*$^{-/-}$ CH12 cells had no CSR defect, while *Rmnd5a*$^{-/-}$*Rmnd5b*$^{-/-}$ CH12 cells resulted in a defect in CSR along with UNG2 and MKLN1 accumulation (Fig. 1f, g; Supplementary Figs. 4a, e; 5a, b), suggesting redundancy between RMND5A and RMND5B. Likewise, *Ranbp9*$^{-/-}$ or *Ranbp10*$^{-/-}$ CH12 cells had normal CSR, while *Ranbp9*$^{-/-}$*Ranbp10*$^{-/-}$ CH12 cells had defective CSR, UNG2 and MKLN1 accumulation (Fig. 1f, g; Supplementary Figs. 4b, f; 5a, b). *Mkln1*$^{-/-}$ CH12 cells also had reduced CSR and UNG2 accumulation (Fig. 1f, g; Supplementary Figs. 5a, b). Deficiency in other members of the CTLH complex had no effects on CSR or UNG2 accumulation (Fig. 1f, g; Supplementary Figs. 4c-f; 5a, b). Collectively, these results show that a CTLH E3 composed of the catalytic core, MAEA, RMND5A/B, and the scaffolding subunits MKLN1, TWA1 and RANBP9/10 are responsible for FAM72A-mediated degradation of UNG2 and CSR.

MAEA-deficiency phenocopies FAM72A-deficiency for SHM and CSR defects in mice

To examine if deficiencies in CTLH complex activity mirrors the previously observed defects in secondary antibody diversification in

https://doi.org/10.1038/s41467-024-52009-x

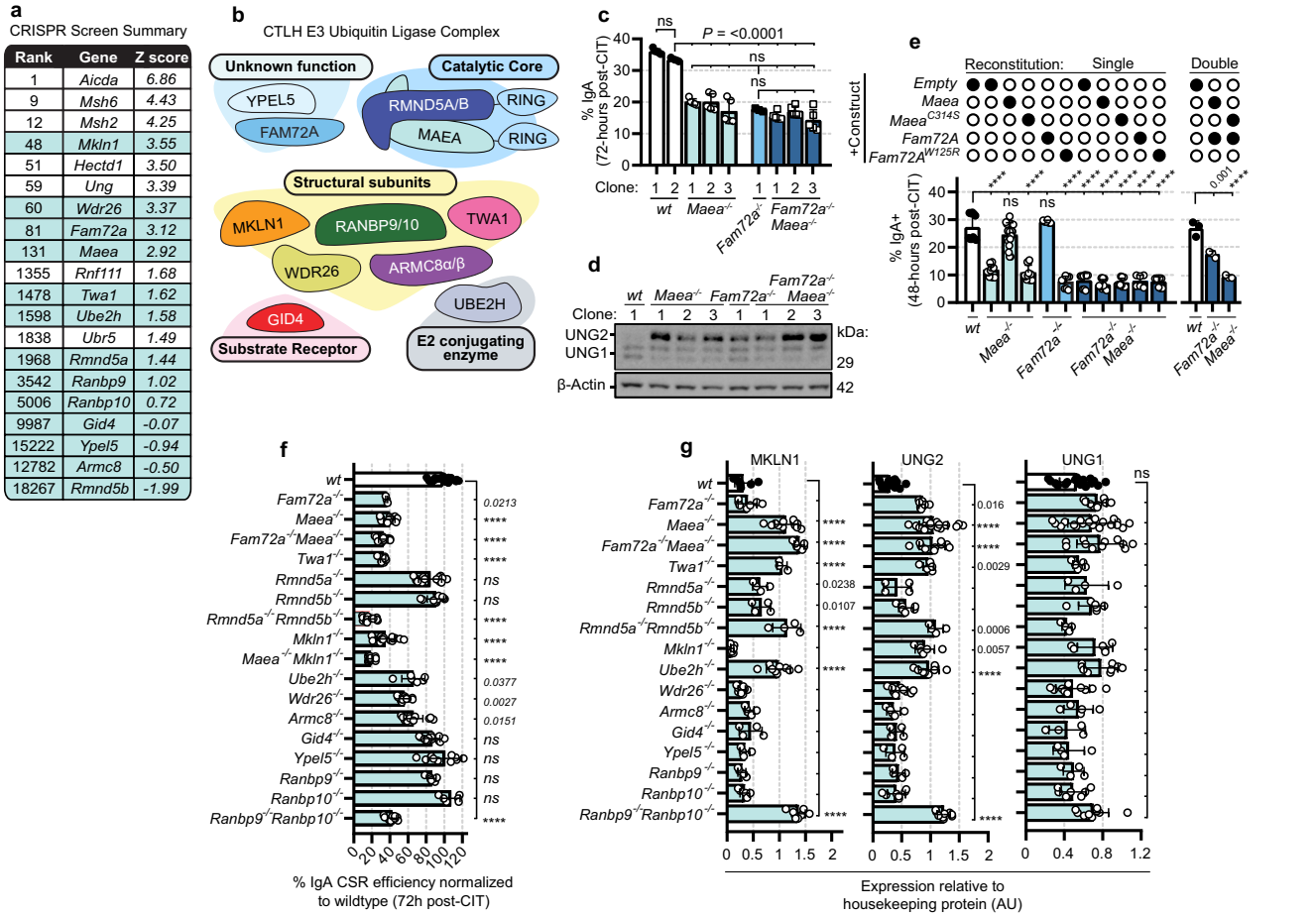
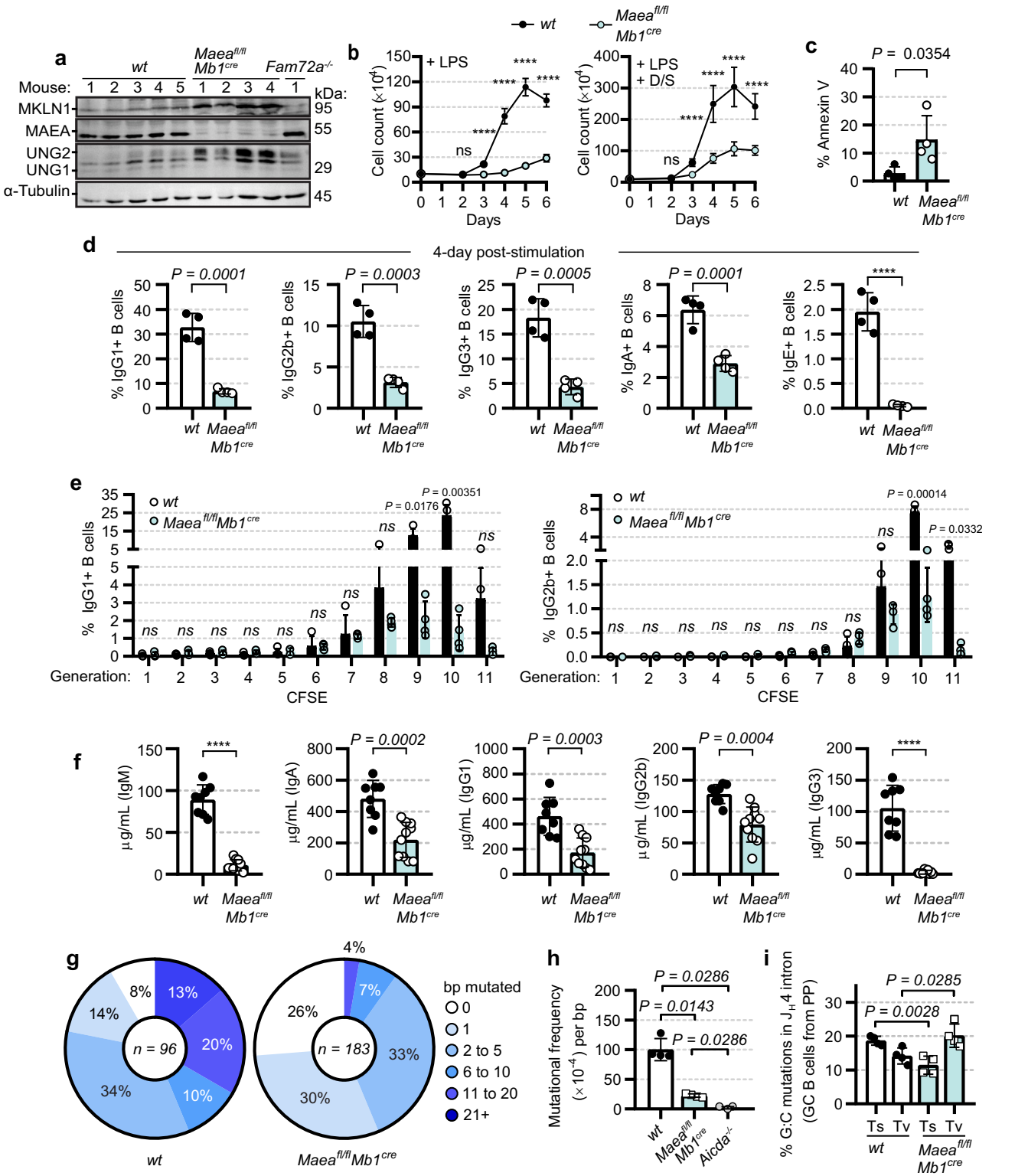


Fig. 1 | The CTLH complex promotes CSR in CH12 cells. a Top ranking genes important for IgA CSR in CH12 cells from a CRISPR-Cas9 screen previously published by our group[9]. The diagram outlines rankings of CTLH complex members, additional E3 ligases in the top 2000 hits, and respective drug-z scores (IgA-) of 19193 genes. **b** Schematic displaying subgroupings of the CTLH complex with FAM72A included as a subunit of unknown function. **c** IgA CSR levels of CH12 clones 3 days post-CIT stimulation of indicated genotypes. %IgA represents %IgA+ of stimulated condition minus %IgA+ of unstimulated condition (background switching). **d** Western blot of CH12 clones of indicated genotypes for UNG1/2. **e** Rescue experiment in *Maea*$^{-/-}$, *Fam72a*$^{-/-}$, *and Fam72a*$^{-/-}$*Maea*$^{-/-}$ with indicated constructs, assaying %IgA CSR (IgA-PE) through flow cytometry. Double reconstitutions were performed separately and stained with IgA-biotin/streptavidin-BV421 due to spectral overlap with IgA-PE. **f** IgA CSR efficiency normalized to wildtype clone CSR levels 72 hours post-CIT treatment: summary of 2-4 clonal knockout CH12 cells of indicated genotypes. Data was analyzed using a one-way Kruskal-Wallis test and *post hoc* Dunn's multiple comparisons test. **g** Densitometry for UNG2, UNG1, and MKLN1 of 2-4 clonal CH12 knockouts for indicated genotypes, results were normalized to β-Actin or α-Tubulin protein levels of each sample and values are denoted by arbitrary units (AU). Data was analyzed using a one-way Kruskal-Wallis test and *post hoc* Dunn's multiple comparisons test for UNG1 and UNG2 due to non-normal distribution. Unless otherwise stated, $n = 3$ biological replicates were performed and data was analyzed using one-way ANOVA and *post hoc* Dunnett's test. Data is presented as mean ± s.d., ns, not significant; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$ ****; $P < 0.0001$. Source data are provided as a Source data file.

Fam72a$^{-/-}$ mice, conditional knockout mice were generated by crossing *Maea*$^{fl/fl}$ with *Mb1*^{cre} mice to knock out *Maea*, the catalytic E3 subunit of the CTLH complex, starting from early B cell ontogeny (Supplementary Fig. 6a). A conditional knockout approach was necessary since full-body *Maea* knockout is embryonically lethal[31]. As expected, splenic B cells from *Maea*$^{fl/fl}$*Mb1*^{cre} mice had increased UNG2 and MKLN1 proteins (Fig. 2a; and Supplementary Fig. 6b). To test if *Maea* deficiency impacts B cell development and maturation, immunophenotyping of bone marrow, spleens, and gut lymphoid follicles (Peyer's patches) were performed. *Maea* deficiency resulted in altered B cell subsets in the bone marrow and secondary lymphoid organs that was not observed in *Fam72a*$^{-/-}$ mice (Supplementary Figs. 6c–h; and Supplementary Figs. 7a-f)[9]. These findings suggest that unlike FAM72A, MAEA and the CTLH complex display additional critical roles in B cell development independent of UNG2 degradation.

To assess CSR, splenic B cells were cultured ex vivo for four days in the presence of various stimuli to induce CSR to various isotypes. Consistent with the CH12 cell line data, we observed that primary *Maea*$^{fl/fl}$*Mb1*^{cre} B cells had reduced proliferative capacity compared to controls after 4 days of culturing (Fig. 2b). This defect in cell numbers was due to higher numbers of apoptotic cells in *Maea*$^{fl/fl}$*Mb1*^{cre} B cells compared to controls (Fig. 2c). We observed a CSR defect to all isotypes in *Maea*$^{fl/fl}$*Mb1*^{cre} B cells in ex vivo cultured primary B cells (Fig. 2d). As the altered proliferation in *Maea*$^{fl/fl}$*Mb1*^{cre} B cells may indirectly reduce CSR efficiency[32,33], we pulsed purified B cells with CFSE to track cell division and CSR within each generation simultaneously. CFSE pulsed B cells were then stimulated with cocktails that induced CSR to either IgG1 or IgG2b. CSR frequency per cell division was reduced in *Maea*$^{fl/fl}$*Mb1*^{cre} B cells compared to the wildtype littermates (Fig. 2e; and Supplementary Fig. 7h). Similar to the CSR defect in ex vivo cultured *Maea*$^{fl/fl}$*Mb1*^{cre} B cells, analysis of serum revealed reduced circulating antibody titers in *Maea*$^{fl/fl}$*Mb1*^{cre} mice compared to wildtype littermates (Fig. 2f). Reduction in circulating IgM titers prompted investigation into the plasma cell phenotype in MAEA-deficient B cells. Indeed, reduced absolute cell counts of IgM^{+} and IgM^{-} plasma cells (CD138^{+}) were observed in the *Maea*$^{fl/fl}$*Mb1*^{cre} mice

https://doi.org/10.1038/s41467-024-52009-x



suggesting additional contributions of the CTLH complex in facilitating antibody production (Supplementary Figs. 8a–d). Hence, MAEA-deficiency phenocopies the CSR defect in *Fam72a*$^{-/-}$ mice, while additionally perturbing B cell development.

To test if *Maea* promotes SHM, the V-region of germinal center (GC) B cells from the Peyer's patches were sequenced. Compared to controls, Peyer's patch GC B cells from *Maea*$^{fl/fl}$*Mb1*^{cre} mice had a reduced mutation frequency (Fig. 2g, h; and Supplementary Figs. 9a–e), and increased frequency of G:C transversion mutations and reduction of G:C transition mutations compared to controls (Fig. 2i), similar to *Fam72a*$^{-/-}$ mice[9,10]. To assess the effects of *Maea*-deletion on affinity maturation, we immunized mice with Nitrophenyl Chicken Gamma Globulin (NP-CGG) and measured the α-NP specific antibody response 21 days post-immunization. Using NP_{30}-BSA or NP_4-BSA coating for ELISA assays that measure for low- and high-affinity α-NP response, respectively, *Maea*$^{fl/fl}$*Mb1*^{cre} mice showed both reduced

Fig. 2 | MAEA-deficient mice display defective CSR and skewed SHM spectrums. **a** Western blot of primary B cells cultured ex vivo for 4-days in the presence of LPS and dextran sulfate, $n = 2$ independent experiments were conducted ($n = 1$ *Fam72a*$^{-/-}$ control mouse used[9]). **b** Proliferation of ex vivo cultured splenic B cells in the presence of LPS or LPS + dextran sulfate, $n = 2$. **c** Annexin V apoptosis assay on splenic B cells cultured ex vivo 4-days post-stimulation for IgG2b, $n = 3$. **d** CSR to indicated isotypes 4-days post-stimulation of splenic B cells cultured ex vivo, $n = 3$. **e** CFSE pulsed splenic B cells stimulated for IgG1 or IgG2b CSR for 4-days, $n = 2$. The generation number was determined based on CFSE signal peaks. **f** Circulating serum Ig levels of 6-8 week old littermate-controlled mice, $n = 2$ ($n = 8$ Wt and $n = 10$ *Maea*$^{fl/fl}$*Mb1*^{cre} littermates). **g** *Ig* J_H4 mutational analysis (mutations per J_H4 sequence) of germinal center Peyer's patch B cells from 5-month-old aged mice (**g**–**i:** $n = 4$ mice per Wt and *Maea*$^{fl/fl}$*Mb1*^{cre} groups and $n = 3$ *Aicda*$^{-/-}$ mice). Sequencing read numbers (per mouse) can be found in Supplementary Fig. 9a-e. **h** Mutational frequency per base pair of the J_H4 region germinal center B cells, analyzed using one-tailed unpaired *t*-test. **i** G:C transversion and transition mutation frequency (%) of the J_H4 region. Data points in **g**–**i** represent the mutational frequency per mouse (# of unique sequence mutations / total number of analyzed J_H4 sequencing reads). Unless otherwise stated, $n = 4$ mice per group were used, and data was analyzed using two-tailed unpaired *t*-tests. All primary B cells were cultured in 15% FBS ex vivo. Data is represented as mean ± s.d., *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$ ****; $P < 0.0001$. Source data are provided as a Source data file.

low-affinity and high-affinity α-NP antibody responses compared to controls (Supplementary Figs. 9f, g). Together, these findings demonstrate that *Maea* is essential to SHM and CSR in primary B cells.

FAM72A interacts with the CTLH complex

To determine if FAM72A, UNG2, and the identified CTLH subunits interact in cells, we expressed C-terminal and N-terminal HA and FLAG-tagged constructs linked to *Fam72a* or mutant *Fam72a*^{W125R} that do not bind UNG2[9,10] into *Fam72a*$^{-/-}$ and *Fam72a*$^{-/-}$*Maea*$^{-/-}$ CH12 cells. HA- or FLAG-tagging of FAM72A on the N-terminus, but not the C-terminus, rescued CSR and reduced UNG2 levels in *Fam72a*$^{-/-}$ CH12 cells (Fig. 3a, b), suggesting that C-terminal tagging of FAM72A (FLAG^{G4S}FAM72A and HA^{G4S}FAM72A) disrupts CTLH induced degradation. N-terminal HA-tagged FAM72A (HA$^{(G4S)3}$FAM72A) co-precipitates with MKLN1 and UNG2 in both *Fam72a*$^{-/-}$ and *Fam72a*$^{-/-}$*Maea*$^{-/-}$ CH12 cells (Fig. 3c). In contrast, C-terminal FAM72A^{G4S}HA only bound to UNG2 and the HA$^{(G4S)3}$FAM72A^{W125R} mutant did not bind to either UNG2 or MKLN1 (Fig. 3c).

To assess the FAM72A interactome in B cells, we performed affinity-purification mass spectrometry (AP-MS) on immunoprecipitated N-terminal HA$^{(G4S)3}$FAM72A, HA$^{(G4S)3}$FAM72A^{W125R}, C-terminal FAM72A^{G4S}HA, and untagged FAM72A expressed in *Fam72a*$^{-/-}$ and *Fam72a*$^{-/-}$*Maea*$^{-/-}$ CH12 cells. The analysis confirmed that HA$^{(G4S)3}$FAM72A associates to UNG, CTLH complex members (MKLN1, TWA1, RANBP9, ARMC8), and with other proteins (Fig. 3d). All interactions with CTLH subunits were lost in the HA$^{(G4S)3}$FAM72A^{W125R} mutant (Fig. 3d). We observed UNG, but no CTLH subunit interactions, following AP-MS of HA$^{(G4S)3}$FAM72A from *Fam72a*$^{-/-}$ CH12 cells, likely due to low levels of endogenous MKLN1 when the CTLH complex is catalytically active in the presence of MAEA[23]. Ruling out disruption of overall protein fold in FAM72A^{W125R} as a cause of lost interaction with UNG and CTLH subunits, FAM72A^{W125R} and FAM72A^{WT} proteins elute as monodispersed peaks with similar elution volumes during size exclusion chromatography (Fig. 3e) and exhibit comparable thermostabilities (melting temperature of FAM72A^{W125R} = 44.7 °C and FAM72A^{WT} = 46.9 °C; Fig. 3f).

We hypothesized that FAM72A acts as an adapter to recruit UNG2 to a CTLH complex containing MKLN1, RANBP9/10, TWA1, RMND5A/B, and MAEA and for ubiquitination (termed CTLH-MKLN1). As RANBP9/10, TWA1, RMND5A/B, and MAEA are common subunits of previously characterized CTLH complexes, we reasoned that MKLN1 is the direct interactor of FAM72A. We confirmed this prediction by co-purification of the stable binary MKLN1-FAM72A complex expressed in insect cells (Fig. 4a). Furthermore, deletion analysis revealed the kelch but not discoidin domain of MKLN1 (Fig. 4b) was sufficient for binding FAM72A (Fig. 4c).

AlphaFold2 structure prediction indicated that the N-terminus of FAM72A (residues 1 to 129) adopts a yippee-like fold followed by an unstructured 20-residue C-terminal tail (residues 130 to 149) (Uniprot ID: Q5TYM5, Fig. 4b, Supplementary Fig. 10a). Deletion of the last 11-residues (FAM72A$^{\Delta 11Cter}$) or 19-residues (FAM72A$^{\Delta 19Cter}$) of FAM72A greatly reduced binding to MKLN1 (Fig. 4b, d; Supplementary Fig. 10b). Further analysis revealed that the C-terminal 21 residues of FAM72A (FAM72A^{21Cter}) were sufficient for binding MKLN1 whereas binding to the isolated kelch domain of MKLN1 was greatly attenuated (Supplementary Fig. 10c). As assessed using a fluorescence polarization assay, the C-terminal 19 residues of FAM72A (FAM72A^{19Cter}–FITC) bound MKLN1 with a dissociation constant (K_d) of 826 nM (Fig. 4e). As expected, binding of FAM72A^{19Cter}–FITC to the kelch domain of MKLN1 was greatly attenuated (Fig. 4e). Competition experiments revealed that full-length FAM72A was 130-fold more efficient than FAM72A^{19Cter} at displacing FAM72A^{19Cter}–FITC from binding to MKLN1, indicating that the yippee domain of FAM72A is also involved in binding MKLN1 (Fig. 4f). Together these results support a binding model where both the yippee domain and C-terminal tail of FAM72A contribute to MKLN1 binding.

UNG2 consists of a C-terminal uracil-DNA glycosylase-like domain (residues 68 to 313) preceded by a disordered N-terminal tail (residues 1 to 67) (Fig. 4b). Previous studies demonstrated direct binding of murine FAM72A to both full-length murine UNG2 and to the first 25 residues of murine UNG2[34]. We recapitulated these results with the corresponding human proteins expressed in insect cells (Fig. 5a). Further analysis revealed that while the C-terminal tail of FAM72A is required for binding to MKLN1 (Fig. 4d; Supplementary Fig. 10b), it is dispensable for binding UNG2 (Fig. 5b; and Supplementary Fig. 10d). This result implicates the yippee domain of FAM72A in binding to both UNG2 and MKLN1.

Since FAM72A binds both MKLN1 and UNG2, we tested whether FAM72A can bind both simultaneously. Indeed, the three-protein complex expressed in insect cells was stable to co-purification on a sizing column (Fig. 4a). Consistent with FAM72A serving as a bridge between MKLN1 and UNG2, the latter two proteins co-expressed in insect cells showed no detectable interaction (Fig. 5c). Together these data support a linear binding model where MKLN1 binds FAM72A and FAM72A binds UNG2. Within this arrangement, the yippee domain of FAM72A engages both MKLN1 and the N-terminus of UNG2. In addition, the C-terminus of FAM72A makes contact with unresolved regions of MKLN1.

A CTLH complex containing MKLN1 and FAM72A ubiquitinates UNG2

We next examined if the MKLN1-FAM72A-UNG2 complex could integrate stably into a CTLH complex containing RANBP9, TWA1, RMND5A, and MAEA (referred to as UNG2-bound CTLH-MKLN1^{FAM72A} with the substrate adapter in superscript). Single-step affinity purification of RANBP9 from insect cells expressing all 7 proteins revealed 7 major bands that co-migrate on sizing columns (Supplementary Fig. 10e). The identity of each band was subsequently confirmed as the expected 7 proteins by trypsin digest and mass spectrometry analysis (Supplementary Fig. 10f). To determine whether UNG2 could be ubiquitinated within the purified complex, we conducted in vitro experiments using the CTLH cognate E2 enzyme, UBE2H, followed by visualization of UNG2 by western blot using an UNG2-specific antibody. We observed a laddering of UNG2 that was dependent on

https://doi.org/10.1038/s41467-024-52009-x

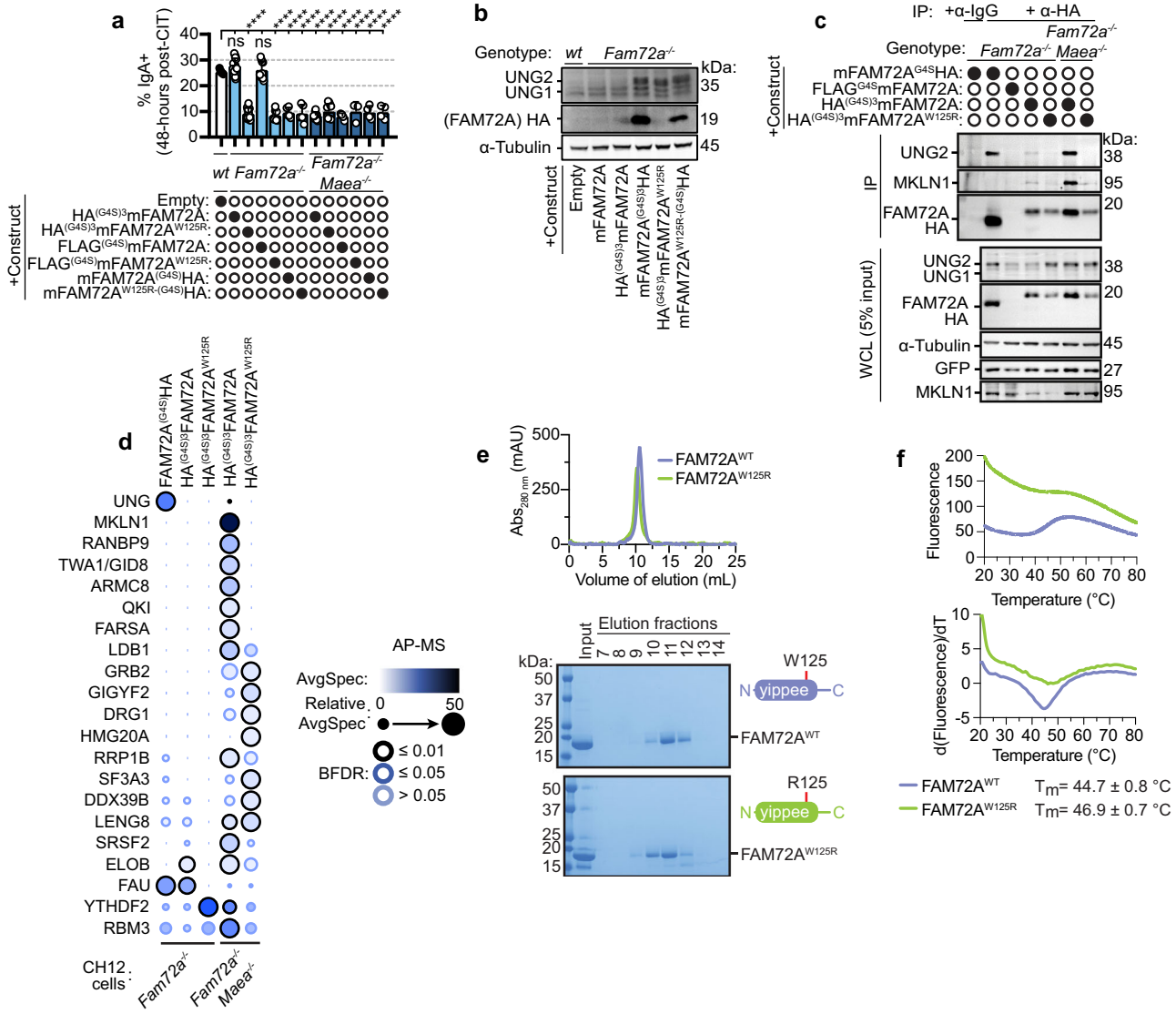


Fig. 3 | FAM72A binds to MKLN1 and UNG2 in cells. a *Fam72a*$^{-/-}$ and *Fam72a*$^{-/-}$*Maea*$^{-/-}$ CH12 cells were retrovirally transduced with constructs containing C-terminal and N-terminal tagged mouse HA/FLAG-tagged FAM72A, $n = 3$. Data was analyzed using one-way ANOVA and *post hoc* Dunnett's test. Stable transduced cells were stimulated with CIT to induce CSR to IgA, 48-hours post-stimulation. Data is presented as mean ± s.d., ns, not significant; ****, $P < 0.0001$. **b** Western blotting for HA (FAM72A) and UNG in a whole cell lysate using α-Tubulin as a control of CH12 cells transduced for the indicated construct, $n = 3$. **c** Immunoprecipitation of tagged FAM72A constructs expressed in *Fam72a*$^{-/-}$ and *Fam72a*$^{-/-}$*Maea*$^{-/-}$ CH12 cells. Cells were treated with MG132 prior to harvest. GFP blotting was performed to confirm retroviral transduction efficiency of the pMX-PIE-GFP vector harboring FAM72A constructs, $n = 2$. **d** For AP-MS, two biological replicates of CH12 cells were collected following treatment with MG132 (3 hours, 20 μM) for each condition of specified genotype + construct: *Fam72a*$^{-/-}$ + FAM72A (no tag control), *Fam72a*$^{-/-}$ + FAM72A^{G4S}HA, *Fam72a*$^{-/-}$ + HA$^{(G4S)3}$FAM72A, *Fam72a*$^{-/-}$ + HA$^{(G4S)3}$FAM72A^{W125R}, *Fam72a*$^{-/-}$*Maea*$^{-/-}$ + HA$^{(G4S)3}$FAM72A, and *Fam72a*$^{-/-}$*Maea*$^{-/-}$ + HA$^{(G4S)3}$FAM72A^{W125R}. Visualized using the Pro-Hits viz program. Fill color and size of dots represent AvgSpec and Relative AvgSpec, respectively. Border color represents the BFDR (false discovery rate) and indicates significance compared to control samples (*Fam72a*$^{-/-}$ + untagged FAM72A). An explanation of the analysis is detailed in the methods section. **e** (top) Size exclusion elution profiles of Strep-Tag II FAM72A wild-type (FAM72A^{WT}) and W125R mutant (FAM72A^{W125R}) proteins. (bottom) Coomassie-stained SDS-PAGE of the indicated 1-mL elution fractions. **f** Thermal shift analysis of FAM72A^{WT} and FAM72A^{W125R} proteins. Representative plots of an experiment showing (top) raw fluorescence and (bottom) first derivative of fluorescence versus temperature. Melting temperature (T_m) values represent the mean ± sd, $n = 3$. Source data are provided as a Source data file.

ubiquitin, ATP, E1, UBE2H and time (Fig. 5d). Reassuringly, exogenous UNG2 was only ubiquitinated by CTLH-MKLN1 (purified without FAM72A), when exogenous FAM72A was added to the reaction (Supplementary Fig. 10g). Together, these data support the notion that RANBP9, TWA1, RMND5A, MAEA and MKLN1 form a productive E3 ligase (denoted CTLH-MKLN1^{FAM72A} by E3 nomenclature convention) with FAM72A acting as the substrate adapter for the recruitment and subsequent ubiquitination of UNG2.

Cryo-EM structural analysis of FAM72A and UNG2 recruitment by CTLH-MKLN1

To uncover the molecular interactions governing the binding of MKLN1 and FAM72A, we conducted single-particle cryo-EM analysis on the minimal binary complex. Following two-step co-purification, we routinely observed a large ratio of MKLN1 to FAM72A protein by SDS-PAGE (Fig. 4a). Perhaps unsurprisingly, we could not visualize the presence of FAM72A, but could resolve MKLN1 in isolation as a

https://doi.org/10.1038/s41467-024-52009-x

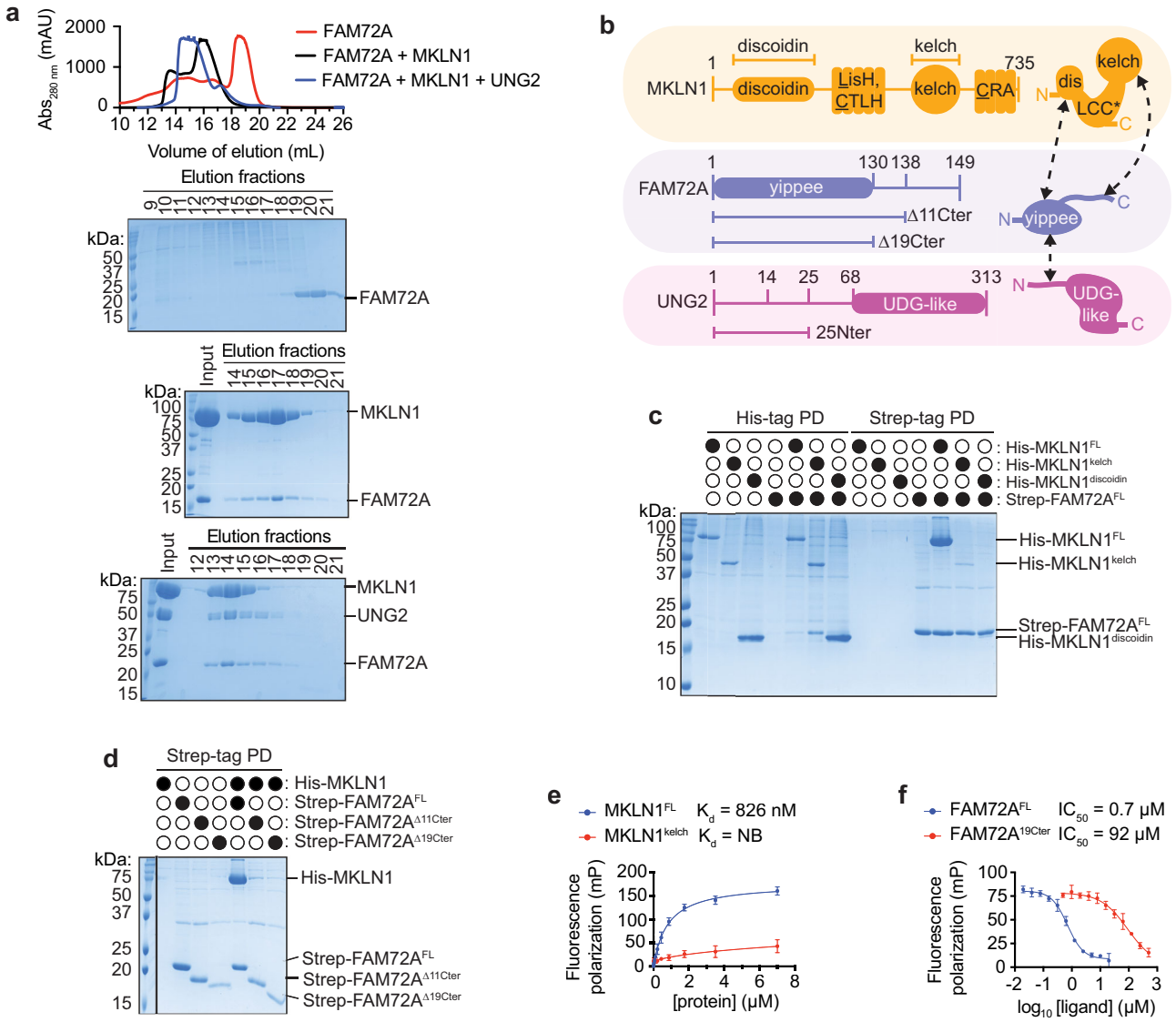


Fig. 4 | FAM72A binds to MKLN1 in vitro. a (top) Size exclusion elution profiles of the indicated proteins and protein complexes. (bottom) Coomassie-stained SDS-PAGE of the indicated 1-mL elution fractions, $n = 1$. **b** Domain organization of MKLN1, FAM72A and UNG2 and constructs used in this study indicated. Dashed arrows illustrate regions of interaction between the three proteins. LCC*= LisH, CTLH and CRA domains. **c**, **d** His_{6x}-tag and Strep-Tag II pull-downs from Sf9 lysates expressing the indicated proteins, followed by visualization by Coomassie-stained SDS-PAGE. The unclipped gel for **d** is shown in Supplementary Fig. 10b., $n = 2$ independent experiments. **e** Florescence polarization binding analysis of the indicated MKLN1 proteins (full-length (FL) or isolated kelch domain) to a FITC-labeled peptide corresponding to the 19 C-terminal residues of FAM72A ($FAM72A^{19Cter}$–FITC). Points represent the mean ± s.d., $n = 3$ independent experiments. **f** Competitive displacement of $FAM72A^{19Cter}$–FITC binding to MKLN1 by full-length FAM72A or unlabeled $FAM72A^{19Cter}$. Points represent the mean ± sd, $n = 3$ independent experiments. Source data are provided as a Source data file.

tetramer to 3.6 Å (Fig. 6a, Supplementary Fig. 11). The structure was consistent with the previously determined 10.1 Å structure of MKLN1 in a CTLH subcomplex containing RANBP9, TWA1, ARMC8A, and GID4[29]. The MKLN1 tetramer consists of two copies of a symmetric dimer, denoted outer dimers, that associate to produce a distinct inner dimer (Fig. 6a). The outer MKLN1 dimer is formed by interaction of the LisH, CTLH and CRA domains, which leaves the kelch and discoidin domains accessible to solvent. In forming a tetramer, two of four kelch domains and all four discoidin domains become occluded. Of the four MKLN1 protomers in the tetramer, the inner two were more rigid allowing us to resolve structural details to 3.3 Å following local refinement (Fig. 6b; Supplementary Fig. 11). Notably, the MKLN1 tetramer is a closed multimer since further polymerization by repetition of the inner and outer modes of dimerization is sterically prohibited (Supplementary Fig. 12a). The functional significance of the MKLN1 tetramer and our inability to visualize FAM72A was only apparent after solving the structure of MKLN1 within the CTLH-$MKLN1^{FAM72A}$ complex.

CTLH E3s containing GID4 and WDR26 (denoted CTLH-$WDR26^{GID4}$) form ring-shaped structures that orient the substrate receptor GID4 and catalytic module composed of MAEA and RMND5A towards the center of the ring where substrates are captured and efficiently ubiquitinated[29]. We predicted that MKLN1 might substitute for WDR26 within the ring structure as both proteins bind RANBP9 through a common mechanism involving their LisH, CTLH, and CRA domains[29]. However, since MKLN1 adopts a tetramer configuration while WDR26 adopts a dimer configuration[29], it was not clear how tetrameric MKLN1 could be integrated into a functional CTLH complex. Somewhat surprisingly, single particle cryo-EM analysis of CTLH-$MKLN1^{FAM72A}$ co-purified with UNG2 revealed an overall ring-like structure to 12 Å resolution (Fig. 7a; Supplementary Fig. 13). Using

https://doi.org/10.1038/s41467-024-52009-x

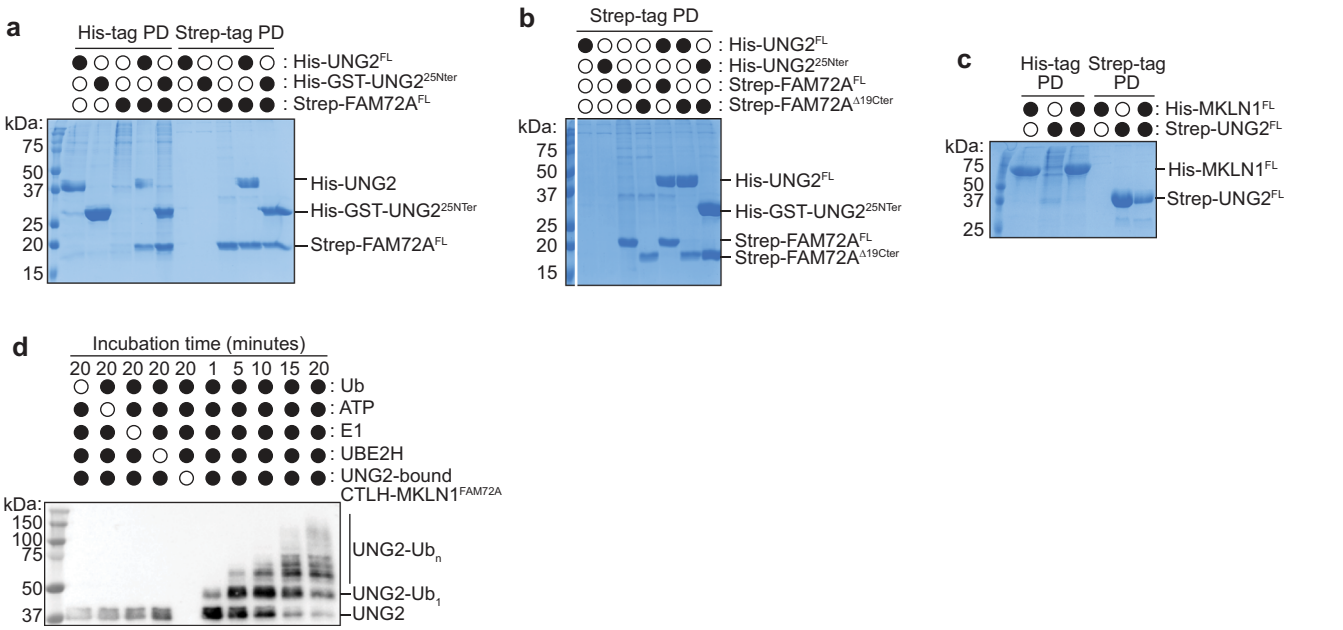


Fig. 5 | UNG2 binds to FAM72A and is ubiquitinated in vitro. a–**c** His_{6x}-tag and Strep-Tag II pull-downs from Sf9 lysates expressing the indicated proteins, followed by visualization by Coomassie-stained SDS-PAGE. The unclipped gel for **b** is shown in Supplementary Fig. 10d, $n = 2$. **d** Western blot analysis of an in vitro ubiquitination reaction reconstituted with the indicated components (ubiquitin (Ub), ATP, E1, UBE2H, and CTLH-MKLN1^{FAM72A} in complex with UNG2). The blot was probed using an anti-UNG2 antibody. $n = 3$ independent experiments.

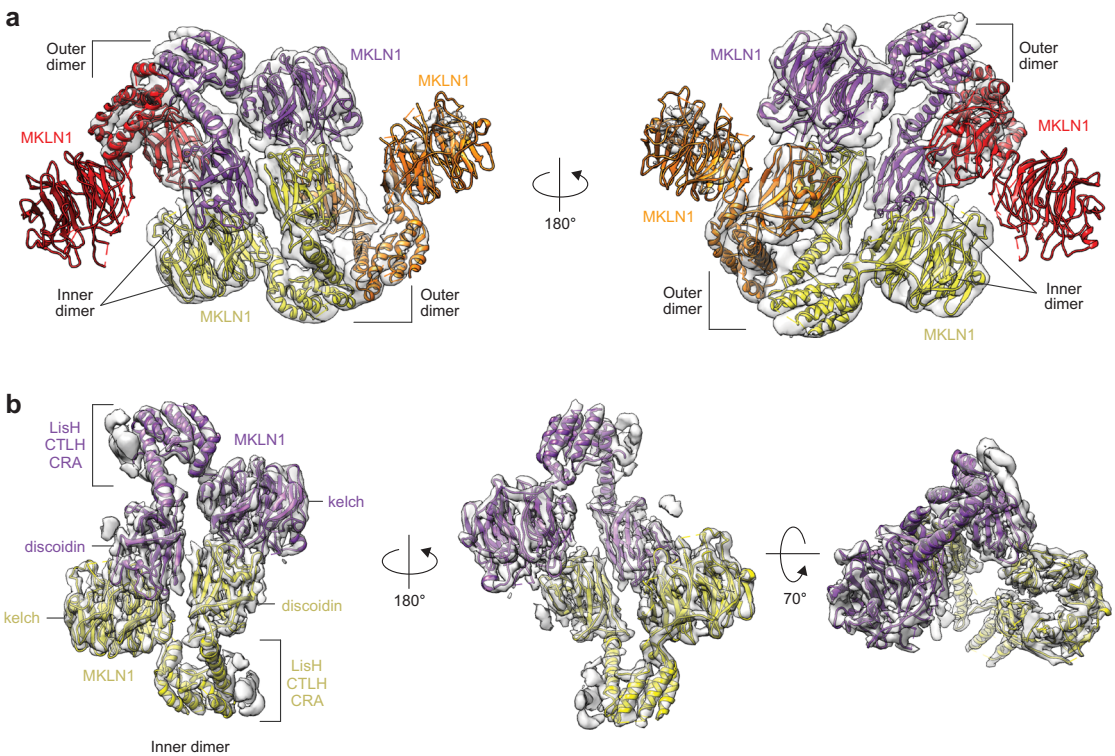


Fig. 6 | Cryo-EM structures of MKLN1 in isolation. a Cryo-EM density map (3.6 Å) and model of an MKLN1 tetramer with protomers colored orange, yellow, red, and purple. Cryo-EM density for the inner dimer protomers (yellow and purple) is superior to that for the two outer dimer protomers (red and orange), particularly for the kelch and CTLH-LisH-CRA domains. The model was generated first by rigid docking of the inner MKLN1 dimer structure (from this study; PDB: 8TTQ) into the map. Then, a copy of each protomer from the inner dimer was docked into the remaining outer protomer densities, for which the discoidin domains were well-defined. **b** Cryo-EM density map (3.3 Å) and fitted model (PDB: 8TTQ) of the inner MKLN1 dimer from focused refinements of the MKLN1 tetramer (shown in **a**). Domains in both MKLN1 protomers are labeled.

https://doi.org/10.1038/s41467-024-52009-x

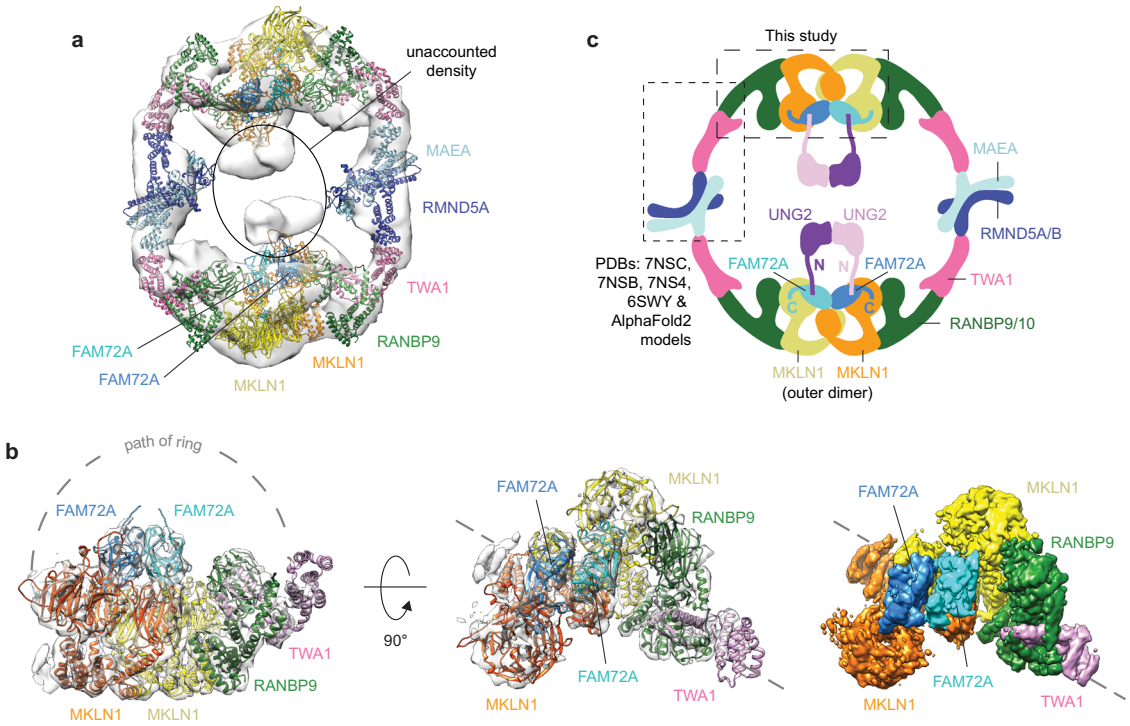


Fig. 7 | Cryo-EM structures of CTLH-MKLN1^{FAM72A} in complex with UNG2 and of MKLN1 bound to FAM72A. a Cryo-EM density map (12 Å) of an UNG2-bound CTLH-MKLN1^{FAM72A} ring assembly with docked models of FAM72A-MKLN1-RANBP9-TWA1 (from this study; see panel **b**) and AlphaFold2-predicted models of RMND5A and MAEA (see methods for details). **b** Cryo-EM density map (5.7 Å) of FAM72A-MKLN1-RANBP9-TWA1 from focused refinements of CTLH-MKLN1^{FAM72A} co-purified with UNG2. (left) Map is docked with AlphaFold2-predicted models of TWA1, RANBP9, two copies of MKLN1 and two copies of FAM72A. Unstructured N-termini of MKLN1 and FAM72A protomers are hidden for clarity. (right) Surface representation of map colored according to docked protein. Dashed lines denote the general path pertaining to the remaining subunits of the UNG2-bound CTLH-MKLN1^{FAM72A} ring assembly. **c** Cartoon representation of a composite model of CTLH-MKLN1^{FAM72A} in complex with UNG2. The experimental density of UNG2 within the complex is ambiguous (shown in panel **a**) and is illustrated speculatively with a binding stoichiometry of 2 UNG2 protomers per FAM72A dimer (see Supplementary Fig. 12d). Dashed boxes denote the structures from which the composite model was generated.

focused refinements, we resolved a 5.7 Å resolution subcomplex that unambiguously contains only the outer dimer of the MKLN1 tetramer (Supplementary Fig. 12b), with each protomer of MKLN1 engaging a single protomer of RANBP9 using the expected binding mode[29]. Upon docking of RANBP9, TWA1, and two MKLN1 protomers into the 5.7 Å focused map, unaccounted-for symmetrical density was apparent at the mouth of the barrel-shaped discoidin domains of MKLN1, which we inferred to be two copies of FAM72A (Supplementary Fig. 12c). AlphaFold2 prediction of a 2:2 MKLN1:FAM72A complex converged on a model that fits well within the density map (Fig. 7b), with the yippee domain of FAM72A engaging the discoidin domain of MKLN1 and consistent with our deletion analysis (Fig. 4d and Supplementary Fig. 10b), the C-terminus of FAM72A engaging the kelch domain of MKLN1. The resultant model of FAM72A-MKLN1-RANBP9-TWA1 reveals additional extensive contacts between the two FAM72A protomers, and potentially between FAM72A and RANBP9.

Next, we generated a composite model of the UNG2-bound CTLH-MKLN1^{FAM72A} complex by combining the model of FAM72A-MKLN1-RANBP9-TWA1 generated here, together with AlphaFold2-predicted models of RANBP9, TWA1, RMND5A and MAEA superimposed on previously solved structures of yeast and human CTLH subcomplexes (PDBs: 7NSC[29], 7NSB[29], 7NS4[29] and 6SWY[35] (Fig. 7a, c and Supplementary Fig. 12d). The composite model fits reasonably into the 12 Å map of UNG2-bound CTLH-MKLN1^{FAM72A}. Notably, all four MKLN1 subunits orient their bound FAM72A subunits towards the inferred active site in the center of the ring and thus are well-poised to recruit and orient UNG2 for ubiquitination. Indeed, our map shows two contiguous unaccounted-for densities in the center of the ring (Fig. 7a), juxtaposed from each FAM72A dimer, that we speculate are copies of two UNG2 dimers (Supplementary Fig. 12d). Consistent with our modeling, molecular weight analysis by size exclusion chromatography coupled to multi angle-light scattering (SEC-MALS) of a co-purified complex of FAM72A with UNG2 revealed a monodispersed peak with molecular weight consistent with 2:2 binding stoichiometry (Observed MW = 109.4 kDa, Theoretical MW = 104.7 kDa; Supplementary Fig. 12e).

Comparison of the isolated MKLN1 tetramer structure with the MKLN1 dimer structure within CTLH-MKLN1^{FAM72A} provides insight into the functional significance of the tetramer arrangement. As revealed by the almost complete steric clash between FAM72A and MKLN1 in its tetramer but not dimer configuration, MKLN1 tetramerization completely buries the FAM72A binding surface (Supplementary Fig. 14a). This observation indicates that MKLN1 binding to FAM72A and MKLN1 tetramerization are mutually exclusive events and that the tetramer structure of MKLN1 is reflective of an auto-inhibitory state. This could account for the highly skewed binding stoichiometry of our MKLN1-FAM72A complex co-purified from insect cells (Fig. 4a).

FAM72A imparts mutagenic properties to the CTLH complex

While members of CTLH-MKLN1 are ubiquitously expressed (Fig. 8a, Supplementary Fig. 15a; Single cell type—MAEA—The Human Protein Atlas), the expression of *FAM72A* is much more restricted (Fig. 8a, Supplementary Fig. 15a; Single cell type—FAM72A—The Human Protein Atlas). Hence, FAM72A might impart a previously unappreciated function to CTLH complexes and promote mutagenesis by its expression in certain cell types. To test this notion, we assessed whether mis-expression of *FAM72A* causes mutations. A previous report suggested that *FAM72A* is overexpressed in human cancers[36]. Hence, we examined RNA-seq datasets from The Cancer Genome Atlas (TCGA)

https://doi.org/10.1038/s41467-024-52009-x

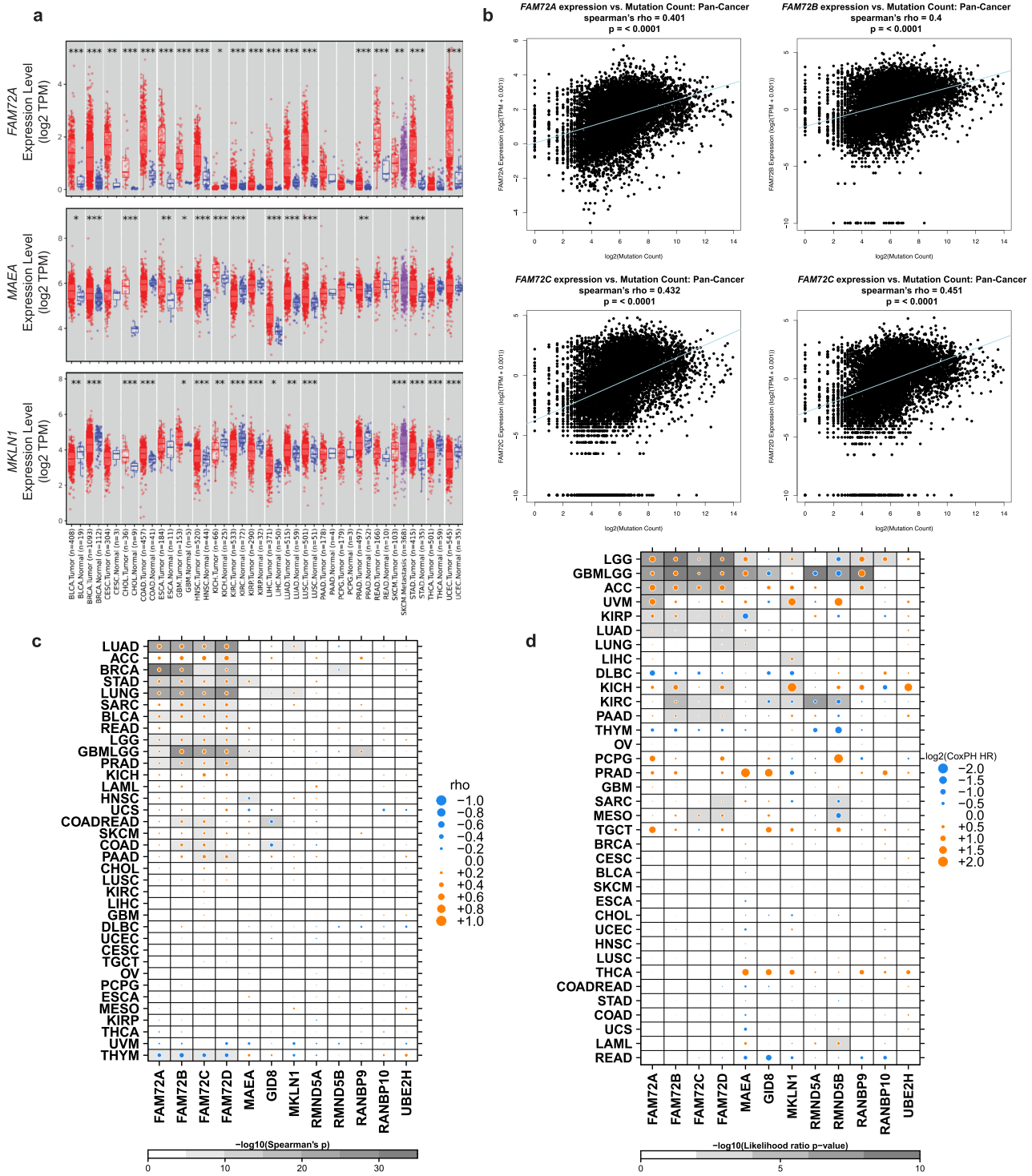


Fig. 8 | *FAM72A* expression is elevated in various human cancers and correlates with mutation count. a Expression of indicated genes in normal flanking tissue (blue) and tumor (red) across all publicly available human RNA-seq datasets available at The Cancer Genome Atlas (TCAG) generated by the TIMER2 algorithm and Gene_DE module[74,76]. The quantity of *n*'s of normal flanking tissue and tumor is indicated on the x-axis. TPM indicates transcripts per million. Data was analyzed using a two-sided Wilcoxon test: *, $P < 0.05$, **; $P < 0.01$; ***, $P < 0.001$. **b** Correlation plot of log2(TPM + 0.001) for *FAM72A-D* versus log(mutation count) for all cancers listed in **c**. **c** UCSC Xena batch effect-corrected TCGA gene expression data across various cancer types with corresponding mutation data obtained from cBioPortal. Dotmap plot Spearman correlation rho and p-values between expression of the indicated genes and mutation count in various cancer types. The size of the dot indicates Spearman rho, while the color indicates negative (blue) or positive (orange) values. Shading indicates -log10(p-values). **d** UCSC Xena batch effect-corrected TCGA gene expression data across different cancer types and corresponding survival data were used for analysis. Dotmap plot showing survival differences in various cancer types (rows) between those with high versus low gene expression of the indicated genes (columns). Shading indicates -log10(p-value). The size of the dot indicates log2(HR), with the color indicating negative (blue) and positive (orange) HRs. TCGA Mutation count data for 32 Pan-Cancer Atlas studies were downloaded from www.cbioportal.org. Spearman correlation analysis was performed using the cor.test base function in R. CoxPH analysis using the 'coxph' function in the 'survival' package. Dotmap plots were generated using the create.dotmap function in the BoutrosLab.plotting.general package. A full list of abbreviations can be found in the methods.

of human *FAM72A* along with highly homologous paralogs, *FAM72B-D*, as well as members of CTLH-MKLN1. Almost all *FAM72* paralogs were over-expressed in an outstanding range of tumor types relative to flanking normal tissue while members of CTLH-MKLN1, including *MAEA*, were generally not. (Fig. 8a, Supplementary Fig. 15a). Examination of the TCGA Pancan data showed a striking correlation between tumor mutation count and *FAM72* expression but not with expression of CTLH-MKLN1 members (Fig. 8b, c). Since genetic mutations are associated with cancer development and progression, we tested whether *FAM72* expression is associated with disease outcomes. Indeed, we found that high *FAM72A-D* expression was associated with poorer survival in some cancers (Fig. 8d, Supplementary Fig. 15b). Together, these findings reveal that *FAM72* expression correlates with tumor mutational load in most human cancers and is not limited to B cell malignancies. As most of these cancers do not express AID, these data show that FAM72 can promote mutations from other sources of dU.

Discussion

Here we have uncovered how a CTLH E3 ligase complex cooperates with FAM72A to promote mutagenesis by counteracting the BER pathway to favor antibody diversification. We describe FAM72A as a CTLH substrate adapter that recruits UNG2 as a substrate for ubiquitination and subsequent proteasomal degradation. FAM72A bridges an interaction between UNG2 and the MKLN1 subunit of the CTLH E3.

The binding of FAM72A to UNG2 substrate is mediated by the globular domain of FAM72A and a degron motif within the unstructured N-terminus[34,36] of UNG2, while the binding of FAM72A to MKLN1 is mediated by multi-partite interactions involving structured domains and the unstructured C-terminus of FAM72A. Structural analysis revealed that MKLN1 forms a tetramer in isolation which reflects an autoinhibited state. Within CTLH-MKLN1^{FAM72A} ring assemblies obtained by co-purification with UNG2, MKLN1 resides as a dimer that is competent for binding FAM72A, which in turn is poised to recruit UNG2 substrate at the center of the ring, in a manner analogous to recruitment of substrates by the orthologous CTLH-WDR26^{GID4} complex in yeast (Supplementary Fig. 14b)[21,29]. We further show that the mutation W125R within the yippee domain of FAM72A abrogates interaction with UNG2, as previously described[9,10], and also with MKLN1. Our cryo-EM structural analysis does not predict a direct interaction of the W125 residue of FAM72A with MKLN1, leading us to speculate that the W125R mutation may induce a long-range conformational change in FAM72A that impairs binding to MKLN1. Notably, humans contain four highly similar FAM72 paralogs characterized by almost identical yippee domains, of which two paralogs (FAM72C and FAM72D) contain an arginine residue in place of tryptophan at position 125. Interestingly, WDR26 (a structural ortholog of MKLN1) also binds a yippee-domain containing protein of unknown function called YPEL5[29]. We speculate that YPEL5 may function as a substrate adapter for CLTH complexes containing WDR26.

Similar to the Cullin-RING ligase (CRL) family of multi-subunit E3s[37], the interchangeability of subunits in CTLH E3s affords great opportunity to regulate diverse biological processes[38] including erythropoiesis[17], embryonic development[39] and now shown here, humoral immunity. The versatility of CTLH E3s is further enhanced by the ability of multiple scaffolding subunits to recruit substrate receptors. For example, while only the Skp1 scaffolding subunit of Cul1 CRLs can engage substrate receptor subunits (i.e., F-box proteins), in CTLH complexes, minimally both MKLN1 and ARMC8A can engage substrate receptors, FAM72A and GID4, respectively. We speculate that this added adaptation could allow for the accommodation of substrates with broader physical properties. The multi-subunit architecture of both E3 families also affords ample opportunities for regulation. For example, CRL assembly and catalytic activity are regulated by Cullin neddylation[40,41]. In the case of CTLH E3s comprised with MKLN1, the transition of MKLN1 from a tetramer to a dimer could serve a similar purpose by regulating the binding of the substrate adapter FAM72A and by the coordinated formation of a productive oligomeric enzyme. Both functions are prohibited by the tetrameric state of MKLN1.

FAM72A promotes mutagenesis of AID-induced dU in B cells[9,10]. However, we find that human *FAM72* expression is correlated with mutation count in a wide range of human cancers (Fig. 8b, c), most of which do not express AID[2]. Incorporation of dU into DNA can arise through spontaneous cytidine deamination, or enzymatically via three cytidine deaminases: AID, APOBEC3A, or APOBEC3B, which are implicated in cancer[42,43]. Hence, FAM72 may promote mutagenesis from many sources of dU. As overexpression of FAM72A[36,44,45] and the dysregulation of CTLH complex function[46] have been implicated in numerous human cancers, the modulation of interactions involving CTLH-MKLN1^{FAM72A} and UNG2 may afford potential avenues for therapeutic intervention.

Methods

Cell lines and culturing

CH12F3-2, BJAB, and Jurkat 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 DMEM supplemented with 10% fetal bovine serum and penicillin-streptomycin. Cells were cultured in a 5% CO_2, 37 °C incubator. Sf9 insect cells were cultured in a 27 °C incubator in I-MAX medium supplemented with L-glutamine, 10% FBS, and penicillin-streptomycin. Cells were routinely checked for mycoplasma.

Plasmids and oligonucleotides

The pMX-PIE expression vector, pBLUEscript-II, pAC8 expression vector, pLIB, were gifts from B.R. San-Martin, G.R.A. Ehrhardt, and N.H. Thomae, respectively. The pBIG1 and pBIG2 expression vectors were purchased from Addgene (Kit #1000000088). Plasmids were sequence verified through sanger or nanopore sequencing. All oligonucleotides and primers are listed in Supplementary Data and were purchased from Invitrogen or Millipore Sigma. Tagged and mutated cDNA in pMX-PIE, site-directed mutagenesis, and insertion were achieved through PCR to generate constructs outlined in Supplementary Data.

CRISPR/Cas9 mutagenesis on CH12F3-2, BJAB, and Jurkat cell lines

For CH12F3-2, *Maea*$^{-/-}$, *Fam72a*$^{-/-}$*Maea*$^{-/-}$, *Mkln1*$^{-/-}$, *and Maea*$^{-/-}$*Mkln1*$^{-/-}$ cells were generated following electroporation with the pX330 vector (Addgene, #42230) in ZAP buffer (25 mM HEPES, 0.75 mM Na_2HPO_4, 140 mM KCL, 5 mM NaCl, 2 mM $MgCl_2$, 0.5% w/v Ficoll) under the following conditions: 325 V, 975 µF, ∞ Ω, and a 4-mm gap using the Gene Pulser Xcell (Bio-Rad) apparatus. Twa1$^{-/-}$, *Rmnd5a*$^{-/-}$, *Rmnd5b*$^{-/-}$, *Rmnd5a*$^{-/-}$*Rmnd5b*$^{-/-}$, *Wdr26*$^{-/-}$, *Armc8α/β*$^{-/-}$, *Ranbp9*$^{-/-}$, *Ranbp10*$^{-/-}$, *Ranbp9*$^{-/-}$*RanBP10*$^{-/-}$, *Ypel5*$^{-/-}$, *Gid4*$^{-/-}$, and *Ube2h*$^{-/-}$ cells were generated following electroporation with the pX459 vector (Addgene, #48139). Cells electroporated with pX330 were subcloned by limiting dilution 72-h post-electroporation into 96-well plates. pX459 electroporated cells were selected in 0.5 µg/mL puromycin for 2-days, then moved to puromycin-free media to recover before proceeding with subcloning. Knockout clones were validated for biallelic, out-of-frame mutations through sequencing, RT-qPCR, or Western blotting. Loss of MAEA and MKLN1 protein were confirmed through Western blotting in the respective *Maea*$^{-/-}$, *Mkln1*$^{-/-}$, and double knockouts. Generation of *Maea*$^{-/-}$ BJAB cells were generated and validated in a similar method to CH12, using the following electroporation conditions: 350 V, 975 µF, ∞ Ω, and a 4-mm gap. *Maea*$^{-/-}$ Jurkat cells were generated through

lentiviral transduction of lentiCRISPR-V2 harboring sgRNA for human *Maea*. All sgRNA and vectors can be found in Supplementary Data.

Mice

Maea$^{tm1A(EUCOMM)Wtsi}$ (EPD0664_2_F07) sperm was obtained from the European Mouse Mutant Archive (EMMA). The in vitro fertilization of C57BL/6 N female mice (Charles River) with *Maea*^{tm1A} mutant sperm and generation of *Maea*^{tm1C} mice were performed at The Centre for Phenogenomics (Toronto, Canada). *Maea*$^{tm1A(EUCOMM)Wtsi}$ sperm contains an insertion of the L1L2_Bact_P cassette at position Chr5:33522583 upstream of critical exon 5 of the *Maea* gene (Chr5:33492916-33530638, + strand, 17.55 cM, cytoband B1, GRCm39) where the critical exon is flanked by intronic loxP sites. *Maea*^{tm1A} mice containing FRT recombination sites were first crossed with mice expressing FLP recombinase to generate *Maea*^{tm1C} mice which were delivered to our Animal Vivarium facility at the University of Toronto. *Maea*^{tm1C} alleles are conditional-ready floxed alleles corresponding to *Maea*^{flox} for simplified nomenclature. *Maea*$^{flox/flox}$ or *Maea*$^{flox/+}$ mice were bred with *Maea*$^{flox/+}$*Mb1*^{cre} mice to obtain the following genotypes: Knockout −*Maea*$^{flox/flox}$*Mb1*^{cre}; Heterozygous−*Maea*$^{flox/+}$*Mb1*^{cre}; Wildtype−*Maea*$^{flox/flox}$, *Maea*$^{flox/+}$, and *Mb1*^{cre}. Both male and female mice were used for experiments, including sex-matched littermate controls using available wildtype genotypes. Experimental procedures were approved by the Animal Care Committee at the University of Toronto (Division of Comparative Medicine/CCBR; protocol number 20011472). Mice were housed on a 14-hour light, 10-hour dark cycle, with 45% humidity and an approximate ambient temperature of 22 °C.

Ex vivo class switch recombination of primary B cells

Splenic B cells were purified using the EasySep Mouse B cell Isolation Kit (StemCell Technologies) with magnetic separation and negative selection. Purified B cells were resuspended in complete RPMI. Following B cell isolation, purified splenic B cells were seeded at $2–4 \times 10^5$ cells/well in complete RPMI for primary cells, 50 µg/mL LPS (Sigma-Aldrich), and the appropriate stimulation necessary to induce ex vivo CSR to IgG1, IgG2b, IgG3, IgA, or IgE, as previously described[47]. Cells were harvested 4-days post-stimulation and stained with FVD eFluor780 or eFluor504 (ThermoFisher) along with mouse IgG1-PE (BD Pharmigen, 550083; 1:150), IgG2b-PE (SouthernBiotech, 1090-09S; 1:150), IgG3-FITC (BD Pharmigen, 553403; 1:100), IgE-FITC (BD Pharmigen, 553415; 1:100) or IgA-PE (SouthernBiotech, 1040-09; 1:150) in the presence of Fc block (2.4G2, 1:100) to assess CSR.

CFSE assay

Purified splenic B cells were pulsed with 5 µM CFSE (Celltrace, Life Technology) for 45 minutes at room temperature, residual CFSE was quenched by adding equal volume of 100% fetal bovine serum and incubating cells at 37 °C for 30 min. Pulsed cells were cultured in complete RPMI containing 50 µg/mL LPS containing 10 ng/µL IL-4 for IgG1 or 30 µg/µL Dextran Sulfate for IgG2b CSR stimulation. Cells were harvested 4-days post CFSE incorporation and IgG1 or IgG2 stained with fixable viability dye eFluor780 or eFluor506 (eBioscience, 65-0865-14/65-0866-14), followed by staining with anti-mouse IgG1-PE (1:150) or IgG2b-PE (1:150) depending on the stimulation condition. Cells were analyzed through flow cytometry on a BD LSRFortessa X-20 instrument.

Primary B cell profiling

Bone marrow and spleen were harvested as previously described[47] from *Maea*$^{flox/flox}$*Mb1*^{cre} knockout, heterozygous *Maea*$^{flox/+}$*Mb1*^{cre}, and wildtype littermates (6–8 week-old). Peyer's patches were similarly collected from sex-matched littermates (11–14 week-old). Single cell suspensions were stained with the following antibody panels in the presence of α-Mouse CD16/32 Fc block (93, eBioscience, 1:100): Spleen −CD23-APC (B3B4, eBioscience, 1:200), B220-FITC (RA3–6B2, SouthernBiotech, 1:400), IgM-eFluor450 (II/41, eBioscience, 1:100), CD93-PE (AA4.1, eBioscience, 1:100), and FVD eF780 (65-0865-14, ThermoFisher, 1:1000); Bone marrow−CD24-APC (30-F1, eBioscience, 1:200), B220-FITC (1:200), IgM-eFluor450 (1:100), BP-1-PE (6C3; BioLegend; 1:100), CD43-BV711 (S7; BD Pharmingen; 1:200), CD138-BV711 (281-2; BD, 1:100), CD38-PE (AB90; BD, 1:100), CD27-APC (LG.3A10; BD, 1:100) and FVD eF780 (1:1000); and Peyer's patches−B220-FITC (1:400), CD3-PerCP-Cy5.5 (GK1.5, eBioscience, 1:200) or (CD3-BV650 (17A2, Biolegend, 1:200), Fas-BV421 (Jo2, BD Horizons, 1:200), GL7-eF660 (RA3–6B2; SouthernBiotech; 1:100), and FVD eF506 (65-0866-14, ThermoFisher, 1:1000).

Mutational analysis of J_H4 regions

Peyer's patches were harvested from 5-month old aged mice and processed for cell sorting and genomic DNA collection. Single-cell suspensions were stained with the Peyer's patch antibody staining panel in the presence of Fc block as described above in preparation for FACS. Approximately 2000–25,000 germinal center B cells ($FVD^-B220^+GL\text{-}7^+Fas^+$) were obtained post-FACS, digested with Proteinase K (Invitrogen) at 55 °C overnight, boiled at 95 °C for 15 min, and then subject to phenol-chloroform genomic DNA purification. A region of the J_H4 intron was PCR amplified using Q5 high-fidelity DNA polymerase with the primers listed in Supplementary Data. J_H4 PCR products (~1.2 kB) were gel extracted with the PureLink Quick Gel Extraction Kit (ThermoFisher). The purified PCR product was then cloned into pCR4Blunt-TOPO vectors, transformed into competent TOP-10 bacteria, and individual clones were sequenced. Analysis of sequencing was analyzed through SHMtool[48].

Total Ig ELISA

Sera of 7-9 week old *Maea*$^{fl/fl}$*Mb1*^{cre} and wildtype littermate mice were collected and used in Ig isotype-specific ELISA assays described previously[49]. Maxisorp 96-well ELISA plates (NUNC) were coated with goat antibodies (SouthernBiotech) against mouse IgM (1021–01, 1:1000 dilution), IgA (1040-01, 1:500 dilution), IgG1 (1071–01, 1:1000 dilution), IgG2b (1090-01, 1:1000 dilution), or IgG3 (1101-01, 1:500 dilution), or overnight at 4 °C in PBS. The following day, plates were blocked with 5% BSA-PBS, and serum was diluted to 1:50 in 2.5% BSA-PBS and added to wells, followed by a serial 1:4 dilution for a 1:50–1:3200 range. Mouse Ig standards were prepared in 2.5% BSA-PBS starting at 800 µg/mL and following a 1:2 dilution series: IgM (11E10, 0101-01), IgA (S107, 0106-01), IgG1 (15H6, 0102-01), IgG2b (A-1, 0104-01), IgG3 (B10, 0105-01). Bound serum Ig and standards were detected by isotype-specific goat anti-mouse Ig conjugated with alkaline phosphatase (AP) in 2.5% BSA-PBS (all from SouthernBiotech): IgM-AP (1021-04, 1:2000 dilution); IgA-AP (1040-04, 1:1000 dilution); IgG1-AP (1070-04, 1:2000 dilution); IgG2b-AP (1090-04, 1:2000 dilution); and IgG3-AP (1100-04, 1:2000 dilution). SIGMAFAST p-Nitrophenyl phosphate substrate was added, allowing for the development of the colorimetric reaction and to measure absorbance (OD_{405}).

NP-Specific ELISA

Maea$^{fl/fl}$*Mb1*^{cre} and wildtype littermate mice of 7-9 weeks were immunized intraperitoneally with 100 ug of $NP_{24\text{-}28}$-CGG (Biosearch Technologies) with equal volume of Imject Alum (Invitrogen) adjuvant or mock PBS controls[49]. Mice were harvested for serum 21-days post-immunization. Sera of mock-immunized and NP-immunized mice was diluted 1:50 in 2.5% BSA-PBS and added to ELISA plates coated with 0.2 µg/mL NP_4-BSA and NP_{30}-BSA (Biosearch Technologies) coated ELISA plates to test for high and low affinity anti-NP antibodies, respectively. Use of AP conjugated antibodies as described above, including IgG2a-AP (1081-04, 1:2000 dilution) were used to assay anti-NP isotypes. SIGMAFAST p-Nitrophenyl phosphate substrate was added allowing for the development of the colorimetric reaction and to measure absorbance (OD_{405}).

Class Switch Recombination Assay on CH12F3-2 Cells

CH12F3-2 cells were seeded at 0.5×10^5 cells/mL in complete RPMI and stimulated with CIT cocktail (2 μg/mL of α-CD40, 10 ng/mL of IL-4, and 1 ng/mL of TGF-β). Unstimulated conditions were run in parallel for each sample. Cells were harvested 3-days post-stimulation, stained with anti-IgA-PE (1:150, 1040-09, SouthernBiotech), and analyzed by flow cytometry. Data represents % IgA^+ of the stimulated conditions minus the unstimulated % IgA^+ background. Anti-IgA-biotin (1:100, 407003, BioLegend) and streptavidin-BV421 (1:200, 405226, BioLegend) were used to measure IgA CSR in retrovirally transduced $GFP^+mCherry^+$ CH12 cells.

Counting assay

CH12 cells were grown to log phase (~3×10^5 cells/mL), collected, resuspended at a concentration of 1×10^5 cells/mL, and seeded at 0.1 mL per well in 96-well tissue culture plates. Primary B cells were harvested and seeded in a similar way. Cell counts were performed in duplicate after seeding, using a hemocytometer, trypan blue staining, and light microscopy.

Annexin V apoptosis assay

The protocol from the APC-conjugated Annexin V Apoptosis Detection Kit (88-8007-72, eBioscience) was followed as indicated by the manufacturer. CH12F3-2 cells were seeded at 2×10^5 cells/mL in complete RMPI under the presence or absence of CIT stimulation. Cells were collected at 48 hours and stained for Annexin V-APC and Propidium Iodide (PI). Primary B cells seeded at 2×10^5 cells/mL in complete RPMI along with 50 μg/mL LPS and 30 μg/mL Dextran Sulfate (Fina BioSolutions LLC) were collected 4-days post-stimulation, processed and stained with Annexin V-APC. Cells were analyzed through flow cytometry on a BD LSRFortessa X-20 instrument and FlowJo V10 software.

Cell cycle analysis

Cells were seeded at 2×10^5 cells/mL in complete RMPI under ± CIT stimulation. Following 48 hours, cells were pulsed with 10 μM of thymidine analog, 5-ethynyl-2′-deoxyuridine (EdU) for 1.5 hours at 37 °C from the Click-iT EdU Alexa Fluor 647 imaging kit (C10340, ThermoFisher). Cells were fixed in 4% formaldehyde and permeabilized in saponin buffer. The manufacturer's protocol was followed for Alexa Fluor 647-Azide and Propidium Iodide (PI) staining, where 1.5 mg of PI was added for every 1×10^5 cells in permeabilization buffer and presence of 100 μg/mL RNAse A. Cell in the different phases of the cell cycle ($G_{0/1}$, S, and G_2/M-phase) were analyzed through flow cytometry and FlowJo software.

Cloning

Site-directed mutagenesis of *Maea* cDNA to generate the $Maea^{C314S}$ mutation was performed using an inverse PCR approach in the pMX-PIE vector. HA and FLAG-tagged mUNG2, mFam72A, and $mFam72A^{W125R}$ containing GGGGS linkers (×1–3) in pMX-PIE were generated through inverse PCR or overhang-PCR followed by restriction cloning. Sf9 insect cell expression vectors were generated by PCR and restriction cloning in the pAC8 vector or by Gibson assembly in the pLIB, pBIG1a, and pBIG2abc vectors[50]. PCR reactions were performed with Q5 high fidelity polymerase (NEB) or KOD polymerase (Sigma-Aldrich) and confirmed by sequencing. All primer information used can be found in Supplementary Data.

Retrovirus transduction

Retrovirus generation was performed as previously described[9]. BOSC23 cells seeded at 4×10^5 cells a day prior were transfected the following day (60–70% confluency) with a mix in Opti-MEM medium (ThermoFisher) containing Fugene 6 (Promega), pMX-PIE expressing cDNA constructs outlined in Supplementary Data, and pCL-Amphotropic. The mix was incubated at room temperature for 15 minutes prior to addition to the BOSC23 cells in culture. After 48 hours, virus-containing supernatant was harvested, filtered through a 0.22 μm filter prior to the addition of 10 μg/mL polybrene and 20 mM HEPES. Supernatant was added to 1×10^6 CH12 cells and a spinfection was performed at $800 \times g$ for 1.5 h at room temperature. Following 24 h, CH12 cells were transferred into complete RPMI containing 1 μg/mL puromycin to allow for the selection of transfected cells and maintained expression of the construct.

Immunoprecipitation and AP-MS of HA-tagged Fam72A from whole cell lysate

CH12 cells were expanded to obtain 1.5×10^7 cells, collected, and resuspended at 1×10^6 cells/mL of complete RMPI in the presence of 20 μM of MG132 for 3 h. Cells were pelleted, washed twice in cold PBS, weighed, and snap-frozen at -80 °C. On day one of immunoprecipitation, samples were thawed on ice, resuspended 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), and lysed for 1 hour at 4 °C. Rabbit IgG (Sigma-Aldrich, 12-370, 1:100) or Rabbit HA antibodies (Cell Signaling, C29F4, 1:100) were added to washed protein G beads and rolled for 3 hours at 4 °C. Beads were 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), added to precleared cell lysate, and rolled overnight. Whole cell lysates from samples were collected following preclearing. On day two, beads were washed 4× in IP wash buffer, resuspended in 6x Laemmli buffer, boiled, and used for Western blotting with 0.2 μM nitrocellulose membranes (BioRad). Membranes were probed for HA, UNG, and MKLN1 using the antibodies previously mentioned. VeriBlot-HRP for IP Detection Reagent (Abcam, ab131366, 1:200) was used as the secondary antibody, the reaction was developed with Pierce ECL substrate (ThermoFisher), and membranes were visualized through the iBright Imaging System (ThermoFisher).

For AP-MS, two biological replicates were prepared for each condition of specified genotype + construct: $Fam72a^{-/-}$ + FAM72A (no tag control), $Fam72a^{-/-}$ + $FAM72A^{G4S}HA$, $Fam72a^{-/-}$ + $HA^{(G4S)3}FAM72A$, $Fam72a^{-/-}$ + $HA^{(G4S)3}FAM72A^{W125R}$, $Fam72a^{-/-}Maea^{-/-}$ + $HA^{(G4S)3}FAM72A$, and $Fam72a^{-/-}Maea^{-/-}$ + $HA^{(G4S)3}FAM72A^{W125R}$. A previously described protocol[51] was followed with minor modifications outlined here. Immunoprecipitation was performed as described above, however on day two, beads were washed 4× in IP wash buffer with gentle centrifugation, washed 2x in 1 mL RIPA buffer followed by 3× washes with 1 mL of 50 mM ammonium bicarbonate (ABC) pH 8.0, and finally resuspended in 200 μL of 50 mM ABC containing 1 μg trypsin. Samples were incubated at 37 °C overnight with agitation. An additional 0.5 μg of trypsin was added and incubated for a further 3 h. The supernatant was transferred to a fresh tube. Beads were washed twice with 100 μL of mass spectrometry grade water (pelleting beads in between) and combined with the supernatant from the previous step. The samples were centrifuged (16,100 × *g*, 10 min), transferred to a new tube, and vacuum centrifugation without heat was used to lyophilize samples which were then stored as dried peptides at -80 °C until mass spectrometry analysis. Mass spectrometry analysis was performed on the TripleTOF™ 6600 instrument. For LC-MS/MS, digested peptides were analyzed using a nano-HPLC coupled to MS. One-quarter of the sample was used for anti-HA AP-MS.

AP-MS data analysis

Mass spectrometry data generated were stored, searched, and analyzed using ProHits laboratory information management system (LIMS) platform[52]. Within ProHits, WIFF files were converted to an MGF format using the WIFF2MGF converter and to an mzML format using ProteoWizard (V3.0.10702) and the AB SCIEX MS Data Converter (V1.3 beta). The data was then searched using Mascot[53] (V2.3.02) and Comet[54] (V2018.01 rev.4). The spectra were searched with the mouse

sequences in the RefSeq database (version 86, February 8, 2018) acquired from NCBI, supplemented with common contaminants from the Max Planck Institute (http://www.coxdocs.org/doku.php?id=maxquant:start_downloads.htm) and the Global Proteome Machine (GPM; ftp://ftp.thegpm.org/fasta/cRAP), forward and reverse sequences (labeled gi|9999 or DECOY), sequence tags, streptavidin, and LYSC_PSEAE for a total of 154,621 entries. Database parameters were set to search for tryptic cleavages, allowing up to 2 missed cleavage sites per peptide with a mass tolerance of 35 ppm for precursors with charges of 2+ to 4+ and a tolerance of 0.15 amu for fragment ions. Variable modifications were selected for deamidated asparagine and glutamine and oxidized methionine. Carbamidomethylation on cysteines was set as a fixed modification. Results from each search engine were analyzed through TPP (the Trans-Proteomic Pipeline, v.4.7 POLAR VORTEX rev 1) via the iProphet pipeline[55]. Proteins were filtered based on iProphet > = 0.95 and unique peptides >= 2. The SAINT analysis tool[56] was used to identify high-confidence protein interactors versus control samples as described previously[51]. Experimental groups were compared against the *Fam72a*$^{-/-}$ + FAM72A (no tag condition) control condition. Two biological replicates for all conditions were analyzed. All proteins with an iProphet probability of ≥95% and 2 unique peptides were used for analysis. Proteins with a BFDR (Bayesian False Discovery Rate) ≤ 0.01 were considered high-confidence protein interactors. The dot plot (Fig. 3d) was generated in ProHits-viz[57].

Quantification of genomic uracil

An identical protocol was followed as previously described[9].

Western blotting of cell lysate preparation

Western blotting was performed using standard protocols. Harvested cells were resuspended in RIPA lysis buffer, followed by the addition of 6x Laemmli buffer. Lysates were boiled at 95 °C for 10 min and centrifuged at max speed before use. For experiments requiring MG132 (Sigma-Aldrich, 474790) treatment, CH12 cells were treated with MG132 (3 h, 20 μM) prior to harvesting the pellet. The following antibodies were used for this study: anti-UNG (Abcam, 245630, for mouse; Origene Technologies, OTI2C12, for human), anti-AID (Invitrogen; 39-2500), anti-β-actin (Sigma-Aldrich, A2066), anti-α-Tubulin (Abcam, ab4074), anti-MAEA/EMP (ThermoFisher, PA5-47914), anti-MKLN1 (ThermoFisher, PA5-88356), anti-HA (NEB, C29F4). ImageJ software (NIH) was used to perform densitometry analysis.

Reverse Transcription of mRNA and Quantitative PCR

TRIzol (Life Technologies)-chloroform extractions of total RNA from cell pellets (> 5×10^5 cells) were performed under the manufacturer's protocol. RNA was resuspended in UltraPure DNase/RNase-Free Distilled Water (10977015, ThermoFisher), Maxima H Minus reverse transcriptase (ThermoFisher) was used to generate cDNA. Amplification of cDNA was achieved using PowerTrack SYBR Green Master Mix (Applied Biosystems), and gene-specific primers listed in Supplementary Data were used for qPCR with the CFX96 Touch Real-Time PCR Detection System and CFX Manager 3.0 (Bio-Rad).

Flow Cytometry and FACS

If possible, biological replicates were performed on the same flow cytometer instruments. This study used the BD FACSCanto II, BD LSRII, BD LSR Fortessa X20, BD LSR Fortessa, BD FACS Aria III sorter, and BD Influx sorter.

Sf9 transfection and Baculovirus generation

For baculovirus generation from pLIB, pBIG1a, and pBIG2abc expression vectors (Supplementary Data), bacmid was generated via transformation of DH10Bac competent cells (Thermo Fisher Scientific) following supplier protocols, followed by transfection of cells. For baculovirus generation from pAC8 expression vectors, co-transfection was performed with Bac10:KO$_{1629}$ linearized viral DNA[58]. Effectene Transfection Reagent (Qiagen) was used for all Sf9 cell transfections. Baculovirus was amplified for three generations following standard protocols.

Pull-downs from Sf9 lysates

Sf9 insect cells were co-transfected with baculovirus for expression of the indicated proteins. Cells were harvested and lysed by sonication in 20 mM HEPES pH 7.5, 500 mM NaCl, 250 μM TCEP, 20 mM imidazole (Buffer A) and 0.05% Brij-25 supplemented with 1x cOmplete, EDTA-free protease inhibitor cocktail tablet (Sigma-Aldrich) and 1 mM PMSF at 4 °C. Cleared lysates were split into two aliquots and each aliquot incubated for 30 min with either Ni^{2+}–NTA (EMD Millipore), Strep-Tactin XT 4 Flow (IBA Lifesciences) or Glutathione sepharose 4B (GE Healthcare Bio-sciences) resin. The resin was washed 5 times with Buffer A, resuspended in Laemmli buffer and visualized by SDS-PAGE and Coomassie staining.

Protein purification

All purification steps were completed at 4 °C. Sf9 insect cells were lysed by sonication in Buffer A supplemented with 1x EDTA-free cOmplete protease inhibitor tablet, 1 mM PMSF, and benzonase. For purifications of His$_{6x}$-MKLN1^{FL}, His$_{6x}$-MKLN1^{kelch}, and His$_{6x}$-UNG2^{FL}, cleared lysates were applied to HisTrap FF columns (Cytiva), and the captured protein eluted on an AKTA Fast Protein Liquid Chromatograph purification system (FPLC; Cytiva) using a gradient to 350 mM imidazole. Fractions containing the protein of interest were pooled and incubated with His$_{6x}$-tagged TEV protease overnight. MKLN1^{FL} was diluted in 20 mM HEPES pH 7.5, 500 mM NaCl, 250 μM TCEP (Buffer B) to a final concentration of 40 mM imidazole and passed through a HisTrap column. The flowthrough was concentrated using an Amicon Ultra centrifugal filter device (Sigma-Aldrich) and injected onto a superposed 6 increase 30/100 GL column (Cytiva) in Buffer B to obtain MKLN1^{FL}. For His$_{6x}$-MKLN1^{kelch} and His$_{6x}$-UNG2^{FL}, during incubation with TEV protease, the sample was simultaneously dialyzed against 20 mM HEPES pH 7.5, 500 mM NaCl, 250 μM TCEP, 1 mM EDTA overnight. The cleaved and dialyzed samples were passed through a HisTrap column. The flowthrough was concentrated and further purified through a superdex 200 30/100 GL (Cytiva) in 20 mM HEPES pH 7.5, 300 mM NaCl, 250 μM TCEP (Buffer C) to obtain MKLN1^{kelch} or His$_{6x}$-UNG2^{FL}.

Lysate co-expressing His$_{6x}$-MKLN1 and Strep-Tag II FAM72A, lysate co-expressing His$_{6x}$-MKLN1, Strep-Tag II FAM72A and UNG2, and lysate co-expressing His$_{6x}$-TWA1, Strep-Tag II RANBP9, His$_{6x}$-MKLN1 and Strep-Tag II FAM72A, and were purified by two-step affinity purifications. His$_{6x}$-tagged proteins were pulled down and eluted as described above. Pooled fractions were immediately applied to StrepTactin resin to pull down Strep-Tag II tagged proteins and washed with Buffer C. MKLN1-FAM72A and MKLN1-FAM72A-UNG2 complexes were eluted from resin by 1× BXT Buffer. FAM72A-MKLN1-RANBP9-TWA1 complex was incubated with His$_{6x}$-TEV protease overnight at 4 °C to elute proteins. Elution fractions containing either the MKLN1-FAM72A complex or the MKLN1-FAM72A-UNG2 complex were each pooled, concentrated, and further purified on a superposed 6 increase 30/100 GL column.

Lysate expressing Strep-Tag II FAM72A, lysate expressing Strep-Tag II FAM72A^{W125R}, lysate expressing UNG2 in complex with CTLH-MKLN1^{FAM72A}, and lysate expressing CTLH-MKLN1 were applied to StrepTactin resin. Strep-Tag II FAM72A and FAM72A^{W125R} was eluted from the resin by overnight cleavage with either TEV protease or 1x BXT Buffer. UNG2-bound CTLH-MKLN1^{FAM72A} and CTLH-MKLN1 were eluted from StrepTactin resin by 1× BXT Buffer. The eluted proteins were concentrated and further purified on a superdex 75 30/100 GL (Cytiva) in 20 mM HEPES pH 7.5, 200 mM NaCl, 250 μM TCEP (Buffer D) to obtain FAM72A, Strep-Tag II FAM72A and Strep-Tag II FAM72A^{W125R}

or a superose 6 increase 30/100 GL column in Buffer C to obtain UNG2-bound CTLH-MKLN1^{FAM72A}.

His_{6x}-UBE2H was expressed in *E. coli* BL21(DE3) Codon+ cells (Agilent Technologies) and induced with 0.25 mM IPTG for 18 hours at 18 °C. Harvested cells were resuspended in 20 mM HEPES pH 7.5, 400 mM NaCl, and 5 mM imidazole and lysed by homogenization (Avestin Inc.). Cleared lysate was applied to a HisTrap HP column (Cytiva) and the captured protein eluted as described above. Pooled fractions were dialyzed into 20 mM HEPES pH 7.5, 50 mM NaCl and 200 μM TCEP, loaded onto a HiTrap Q HP column (Cytiva) and eluted with a gradient to 500 mM NaCl. Fractions containing His_{6x}-Ube2H were dialyzed into 20 mM HEPES pH 7.5, 200 mM NaCl, and 200 μM TCEP storage buffer. UBA1 (E1) and ubiquitin were purified as previously described[59,60].

Analytical size exclusion chromatography

Affinity-purified protein were injected onto a superose 6 increase 30/100 GL column or superdex 75 30/100 GL column attached to an AKTA FPLC. Absorbance at a wavelength of 280 nm was monitored as 1-mL fractions of the elution were collected.

Thermal shift assays

A real-time PCR instrument (LightCycler 480 II, Roche) was used to monitor protein unfolding via SYPRO orange fluorescent dye (Invitrogen, Carlsbad, CA). Experiments were performed with 75 μM of Strep II-tagged FAM72A wild-type and W125R mutant proteins in 20 mM HEPES, 200 mM NaCl, 0.25 μM TCEP pH 7.5, 5× SYPRO orange, and 0.5% DMSO in a final volume of 20 μL. Samples were subjected to a temperature range of 20 to 85 °C with a ramp rate of 0.01 °C/s. Melting temperature values were deduced by identifying the minima of the first derivative of each curve in GraphPad Prism, followed by calculation of the mean and standard deviation from three replicates.

In-gel digest and mass spectrometry analysis

Elution fractions of CTLH-MKLN1^{FAM72A} bound to UNG2 after purification through a superose 6 increase 30/100 GL column were resolved by SDS-PAGE and stained with Coomassie. Gel bands were excised (Supplementary Fig. 10f) and dehydrated with 95% ethanol in 100 mM ammonium bicarbonate (ABC). Samples were reduced with 5 mM DTT at 56 °C for 30 min and alkylated with 50 mM iodoacetamide (IAA) for 45 min and digested with 0.42 μg trypsin overnight at 37 °C. Peptides were extracted from the gel bands with 5% formic acid and dried down. Mass spectrometry analysis in Data Dependent Acquisition (DDA) mode was performed as previously described[61] with separation on a 45 min gradient. Samples were analyzed with MSFragger v3.7[62] within the ProHits LIMS interface[63] using default settings unless specified. Precursor and fragment mass tolerance were set to 25 ppm. Peptideprophet was turned on with the following command line: --minprob 0.05 --decoyprobs --ppm --accmass --decoy DECOY --nonparam –expectscore. iProphet Cmd line: --nonsp. PTMProphet Command line: KEEPOLD STATIC EM = 1 NIONS=b MINPROB = 0.5. Proteinprophet Command line was as follows: --maxppmdiff 30. FDR filter and Report Command line: --sequential --razor --prot 0.01 --picked –mapmods.

Fluorescence polarization binding assays

Binding experiments were performed with 20 nM FAM72A^{19Cter}–FITC (Biomatik, Ontario, Canada) and increasing concentrations of MKLN1^{FL} or MKLN1^{kelch} proteins in a buffer containing 20 mM HEPES pH 7.5, 200 mM NaCl, 1 mM DTT, 0.01% Brij-35, and 0.1 mg/mL BSA. Competitive displacement assays were performed in the same buffer using 1 μM MKLN1^{FL} in complex with 20 nM FAM72A^{19Cter}–FITC and increasing concentrations of FAM72A^{FL} protein or FAM72A^{19Cter} peptide (Biomatik, Ontario, Canada). Samples were incubated for 1 h 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. Graphs and IC_{50} values were generated using a four-parameter dose–response function in GraphPad Prism.

In vitro ubiquitination assays

To re-constitute ubiquitination in vitro by UNG2-bound CTLH-MKLN1^{FAM72A}, 20 μM ubiquitin, 0.25 μM E1, 12.5 μM UBE2H, 0.3 μM CTLH-MKLN1^{FAM72A} co-purified with UNG2, 1.25 mM ATP and 2.5 mM $MgCl_2$ in 20 mM HEPES pH 7.5, 200 mM NaCl and 0.5 μM DTT were combined. For reactions using CTLH-MKLN1, the same components were used except with 1 μM E1, 0.3 μM CTLH-MKLN1, 1.4 μM FAM72A and 0.62 μM UNG2. All reactions were incubated at room temperature and quenched by addition of Laemmli buffer at the indicated time points. Western blot analysis was performed following standard protocols using an anti-UNG2 antibody (MA5-25686, Invitrogen, 1:500).

Cryo-EM sample preparation and imaging

The map of the MKLN1 tetramer and the map of the inner MKLN1 dimer originated from a single sample of affinity-purified FAM72A-MKLN1-RANBP9-TWA1 with Strep-Tag II and His_{6x}- tags on RANBP9 and TWA1, respectively. The map of UNG2-bound CTLH-MKLN1^{FAM72A} and the map of FAM72A-MKLN1-RANBP9-TWA1 originated from a single sample of affinity-purified UNG2-bound CTLH-MKLN1^{FAM72A} with Strep-Tag II and His_{6x}- tags on RANBP9 and UNG2, respectively. 4 μL of 0.3 mg/mL of each sample were applied to glow-discharged (10 s, 15 mA) holey carbon grids (CF312; Electron Microscopy Sciences). Grids were blotted with cellulose filter paper (Ted Pella, Inc.) for 3 s with a blot force of either -5 or -10, and vitrified by plunging into liquid ethane using a Vitrobot Mark IV cryo-plunger (Thermo Fisher Scientific) at 4 °C and 100% humidity.

Two datasets were collected from two grids that were combined to produce maps of MKLN1 tetramer and inner MKLN1 dimer. The grids were imaged on a Titan Krios microscope (Thermo Fisher Scientific) operated at 300 kV equipped with a GatanK3 imaging system at a 105,000x nominal magnification corresponding to a pixel size of 0.4125 Å. Two datasets of 3313 and 2209 movies were collected using Leginon (v3.6)[64] at a dose rate of 28.54 $e^-/Å^2/s$ with a total exposure of 2.00 s at a nominal defocus range of 0.9–2.4 μm. Intermediate frames were recorded every 0.040 s for a total of 50 frames per micrograph (see Supplementary Fig. 11 for cryo-EM workflow).

Three datasets were collected from a single grid that produced maps of UNG2-bound CTLH-MKLN1^{FAM72A} and FAM72A-MKLN1-RANBP9-TWA1. One dataset of 9836 movies was imaged on a Falcon4i detector at a 75,000× nominal magnification corresponding to a pixel size of 1.03 Å at a nominal defocus of 1.0–2.0 μm, ultimately producing the map of UNG2-bound CTLH-MKLN1^{FAM72A}. Movies were collected using EPU (v3.3) at a dose rate of 5.45 $e^-/Å^2/s$ with a total exposure of 7.15 s. Intermediate frames were recorded every 0.178 s for a total of 40 frames per micrograph. Two additional datasets were imaged from the same grid on a Titan Krios microscope operated at 300 kV equipped with a GatanK3 imaging system at a 105,000x nominal magnification corresponding to a pixel size of 0.4125 Å, to ultimately produce the focused map of FAM72A-MKLN1-RANBP9-TWA1. One dataset of 7886 movies (untilted) and one dataset of 4184 movies with a 40° grid tilt were collected using Leginon (v3.6)[64] at a dose rate of 26.99 $e^-/Å^2/s$ with a total exposure of 1.80 s. The untilted dataset was collected at a nominal defocus range of 0.8–3.2 μm and intermediate frames were recorded every 0.045 s for a total of 40 frames per micrograph. The tilted dataset was collected at a nominal defocus range of -2.4–5.0 μm and intermediate frames were recorded every 0.030 s for a total of 60 frames per micrograph (see Supplementary Fig. 13 for cryo-EM workflow).

Cryo-EM image analysis

Movies were motion corrected, and contrast transfer function (CTF) parameters were estimated with CTFFIND (v4.1.14) using either cryoSPARC Live (v4.1.2) or RELION (v4.0.1) (see Supplementary Fig. 11 for cryo-EM workflow). Micrographs outside an optimal defocus range or with poor CTF fits were removed. For the sample of FAM72A-MKLN1-RANBP9-TWA1 that was used to generate the map of the MKLN1 tetramer and the map of the inner MKLN1 dimer, particles were picked by crYOLO (v1.8.2)[65] using a general model. Picked particles were imported into cryoSPARC[66] and subject to multiple rounds of 2D classification followed by ab initio reconstruction to generate initial volumes. Particles were further classified over multiple rounds of heterogeneous refinement and finally homogenous refinement to yield a map of MKLN1 tetramer in isolation at 3.6 Å resolution (FSC threshold = 0.143). Focused refinement of the MKLN1 tetramer yielded a map of the inner MKLN1 dimer at 3.3 Å resolution (FSC threshold = 0.143; Supplementary Table 1) (see Supplementary Fig. 11 for cryo-EM workflow).

For the sample of UNG2-bound CTLH-MKLN1^{FAM72A} that was used to generate the map of UNG2-bound CTLH-MKLN1^{FAM72A}, particles were initially picked by crYOLO using a general model and the best picks used to train a model by TOPAZ[67] in cryoSPARC. Intact CTLH complexes were selected from 2D classifications and aligned to EMD_12542 as an initial model[29]. Iterative rounds of heterogeneous and homogenous refinements yielded a map of an UNG2-bound CTLH-MKLN1^{FAM72A} ring assembly at 12.0 Å resolution (FSC threshold = 0.143) (see Supplementary Fig. 13 for cryo-EM workflow).

For the sample of UNG2-bound CTLH-MKLN1^{FAM72A} that was used to generate the focused map of FAM72A-MKLN1-RANBP9-TWA1, initial particles were manually picked in crYOLO and used to train a custom model that was used for auto-picking. Particles were imported into RELION and filtered using multiple rounds of 2D classification and 3D classification. Selected particles were symmetry expanded along the C2 axis, signal subtracted, and passed through one round of masked 3D classification, focusing on the expected MKLN1-bound region. Particles from the best classes were subject to 3D refinement using a mask around the improved MKLN1 region. Local 3D auto-refinements were performed with adjusted masks to improve surrounding densities to a final resolution of 5.7 Å (FSC threshold = 0.143) (see Supplementary Fig. 13 for cryo-EM workflow).

Model building and refinement

Half maps from the focused refinement of the MKLN1 inner dimer were sharpened in DeepEMhancer using the highRes model[68]. Two copies of an AlphaFold2-predicted[69] model of MKLN1 were used as initial models for model building. The kelch domains, discoidin domains, and individual helices from the CTLH-LisH-CRA domains were fit into the map as rigid bodies in UCSF Chimera (v1.14)[70]. Multiple iterations of real space refinements by Phenix (v1.20.1-4487)[71] and ISOLDE[72] and manual adjustments in Coot[73] were used to generate the final model of the MKLN1 inner dimer.

For modeling of the MKLN1 tetramer, the model of MKLN1 inner dimer generated as described above, was first docked into the map. Density for the discoidin domains of the outermost MKLN1 protomers was well-resolved, whereas map density for the kelch and LisH-CTLH-CRA domains was diffuse, likely reflecting flexibility. Therefore, atomic models for the outer MKLN1 protomers were generated by positioning copies of the inner MKLN1 protomers using discoidin domains as anchors, followed by rigid body fitting.

To generate the composite model of the FAM72A-MKLN1-RANBP9-TWA1 complex, AlphaFold2 multimer prediction was used to generate five similar models of a 2:2 complex of MKLN1 and FAM72A (default settings, number of multimer predictions per model = 1) that were first rigid-body docked into the map. The model with the best fit was retained, and then each chain was individually fit into the map by Chimera. The AlphaFold2-predicted model of RANBP9 (UniProt ID: Q96S59) was docked and superimposed with a previously solved structure of TWA1-RANBP9 (PDB: 7NSC) to guide docking of an AlphaFold2-predicted model of TWA1 (UniProt ID: Q9NWU2).

To generate the composite model of the CTLH-MKLN1^{FAM72A} ring assembly, structures of yeast and human CTLH subcomplexes were superimposed (PDBs: 7NSC, 7NSB, 7NS4, and 6SWY). Yeast and/or human copies of WDR26, ARMC8, and GID4 were removed, while single copies of yeast and/or human RMND5A, MAEA, TWA1 and RANBP9 were retained. AlphaFold2-predicted models of full-length RMND5A (UniProt ID: Q9H871), MAEA (UniProt ID: Q7L5Y9), TWA1 (UniProt ID: Q9NWU2) and RANBP9 (UniProt ID: Q96S59) were then superimposed to generate a composite model of a subcomplex bound in the following sequence: RMND5A-MAEA-TWA1-RANBP9. Since there were no structures of RMND5A bound to TWA1 at the time of publication, the binding mode of MAEA for TWA1 was recapitulated for the RMND5A-TWA1 interaction. This resulted in a composite subcomplex of AlphaFold2-predicted protein chains, based on available structures in the following sequence: RANBP9-TWA1-RMND5A-MAEA-TWA1-RANBP9. Two copies of the aforementioned subcomplex were rigid-body docked into the ring. The remaining density was fit with a 2:2 complex of MKLN1-FAM72A generated from our composite model of FAM72A-MKLN1-RANBP9-TWA1 described above.

SEC-MALS

An Advancebio SEC 300 A 2.7 μm 4.6 × 300 mm column was equilibrated at 4 ˚C in 20 mM HEPES, 200 mM NaCl, 250 μM TCEP pH 7.5 at a flow rate of 0.2 mL/min with a 1260 Infinity II bio-inert HPLC (Agilent Technologies). A MiniDAWN TREOS Multi-Angle Light Scattering (MALS) and Quasi-Elastic Light Scattering (QELS) detection instrument (Wyatt Technology) and OptiLab T-rEX Refractive Index (RI) detection instrument (Wyatt Technology) signals were collected along with the UV_{280} chromatogram. Proteins were centrifuged at 13.3 × g at 4 °C for 10 min prior to injection. A preliminary 20 μL injection of 2 mg/mL BSA solution standard was performed to calibrate protein peak and retention time characteristics for each of the flow cells required for subsequent sample analysis. 30 μg injections of FAM72A in complex with UNG2 were conducted with a UV_{280} extinction coefficient of 2.382 mL/(mg cm).

TIMER2.0 gene_DE analysis

The TIMER2.0 algorithm and the Gene_DE Module (Available at the webserver: http://timer.cistrome.org/) provides visualization of differential gene expression between tumor and matched normal flanking tissue across all cancer datasets available at The Cancer Genome Atlas (TCGA)[74,75]. Unpaired tumor groups lacking data lacking normal flanking counterpart for comparison were excluded (Fig. 8, Supplementary Fig. 15). Abbreviations: ACC, Adenoid cystic carcinoma; LUAD, Lung Adenocarcinoma; BRCA, Breast Invasive Carcinoma; STAD, Stomach Adenocarcinoma; LUNG, Lung Cancer; SARC, Sarcoma; BLCA, Bladder Urothelial Carcinoma; READ, Rectum Adenocarcinoma; LGG, Low-Grade Glioma; GBMLGG, Glioblastoma and Low-Grade Glioma; PRAD, Prostate Adenocarcinoma; KICH, Kidney Chromophobe; LAML, Acute Myeloid Leukemia; HNSC, Head and Neck Squamous Cell Carcinoma; HNSC-HPV, Head and Neck Squamous Cell Carcinoma - Human Papillomavirus; UCS, Uterine Carcinosarcoma; COADREAD, Colon and Rectum Adenocarcinoma; SKCM, Skin Cutaneous Melanoma; COAD, Colon Adenocarcinoma; PAAD, Pancreatic Adenocarcinoma; CHOL, Cholangiocarcinoma; LUSC, Lung Squamous Cell Carcinoma; KIRC, Kidney Renal Clear Cell Carcinoma; LIHC, Liver Hepatocellular Carcinoma; GBM, Glioblastoma Multiforme; DLBC, Diffuse Large B-Cell Lymphoma; UCEC, Uterine Corpus Endometrial Carcinoma; CESC, Cervical Squamous Cell Carcinoma and

Endocervical Adenocarcinoma; TGCT, Testicular Germ Cell Tumors; OV, Ovarian Serous Cystadenocarcinoma; PCPG, Pheochromocytoma and Paraganglioma; ESCA, Esophageal Carcinoma; MESO, Mesothelioma; KIRP, Kidney Renal Papillary Cell Carcinoma; THCA, Thyroid Carcinoma; UVM, Uveal Melanoma; THYM, Thymoma.

Sequencing
Sanger sequencing was performed at The Center for Applied Genomics (TCAG) in Toronto, Ontario, Canada while nanopore sequencing was performed at Plasmidsaurus in Eugene, Oregon, USA.

Bioinformatics analysis
TCGA Expression and clinical data were downloaded using the UCSCXenaTools package in R. TCGA Mutation count data for 32 Pan-Cancer Atlas studies were downloaded from www.cbioportal.org. Spearman correlation analysis was performed using the cor.test base function in R. CoxPH analysis was performed using the 'coxph' function in the 'survival' package. Dotmap plots were generated using the create.dotmap function in the BoutrosLab.plotting.general package.

Statistics
All statistical comparisons were performed on GraphPad Prism software version 9.5. The Shapiro–Wilk normality test was performed on datasets to validate normal distribution. P values of ≤ 0.05 were considered significant: *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; and ****, $P < 0.0001$.

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

Data availability
The mass spectrometry proteomics data has been deposited to the ProteomeXchange Consortium via the massIVE partner repository with the dataset identifie PXD054121. The mass spectrometry data from in-gel digests have been deposited to the ProteomeXchange Consortium via the massIVE partner repository with the dataset identifier PXD044515. Cryo-EM density maps have been deposited into the Electron Microscopy Data Bank (EMDB) under accession codes EMDB-45088 (Cryo-EM structure of an autoinhibited MKLN1 tetramer), EMD-41612 (Cryo-EM structure of the inner MKLN1 dimer from an autoinhibited MKLN1 tetramer), EMDB-45138 (Cryo-EM structure of CTLH-MKLN1^{FAM72A} in complex with UNG2) and EMDB-45186 (Cryo-EM structure of a FAM72A-MKLN1-RANBP9-TWA1 complex). Model coordinates have been deposited into the Protein Data Bank (PDB) under accession code 8TTQ (Cryo-EM structure of the inner MKLN1 dimer from an autoinhibited MKLN1 tetramer). Other structures used in this study were obtained from the PDB using accession codes 7NSC, 7NSB, 7NS4 and 6SWY and from the AlphaFold2 database using accession codes Q9UL63 (MKLN1), Q96S59 (RANBP9), Q9NWU2 (TWA1), Q7L5Y9 (MAEA) and Q9H871 (RMND5A). Source data are provided in this paper.

References

1. Muramatsu, M. et al. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* **102**, 553–563 (2000).
2. Feng, Y., Seija, N., JM, D. I. N. & Martin, A. AID in antibody diversification: there and back again. *Trends Immunol.* **41**, 586–600 (2020).
3. Cascalho, M., Wong, J., Steinberg, C. & Wabl, M. Mismatch repair co-opted by hypermutation. *Science* **279**, 1207–1210 (1998).
4. Michael, N. et al. Effects of sequence and structure on the hypermutability of immunoglobulin genes. *Immunity* **16**, 123–134 (2002).
5. Phung, Q. H. et al. Increased hypermutation at G and C nucleotides in immunoglobulin variable genes from mice deficient in the MSH2 mismatch repair protein. *J. Exp. Med.* **187**, 1745–1751 (1998).
6. Di Noia, J. & Neuberger, M. S. Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase. *Nature* **419**, 43–48 (2002).
7. Wiesendanger, M., Kneitz, B., Edelmann, W. & Scharff, M. D. Somatic hypermutation in MutS homologue (MSH)3-, MSH6-, and MSH3/MSH6- deficient mice reveals a role for the MSH2-MSH6 heterodimer in modulating the base substitution pattern. *J. Exp. Med.* **191**, 579–584 (2000).
8. Rada, C., Di Noia, J. M. & Neuberger, M. S. Mismatch recognition and uracil excision provide complementary paths to both Ig switching and the A/T-focused phase of somatic mutation. *Mol. Cell* **16**, 163–171 (2004).
9. Feng, Y. et al. FAM72A antagonizes UNG2 to promote mutagenic repair during antibody maturation. *Nature* **600**, 324–328 (2021).
10. Rogier, M. et al. Fam72a enforces error-prone DNA repair during antibody diversification. *Nature* **600**, 329–333 (2021).
11. Hayran, A. B. et al. RPA guides UNG to uracil in ssDNA to facilitate antibody class switching and repair of mutagenic uracil at the replication fork. *Nucleic Acids Res.* **52**, 784–800 (2024).
12. Santt, O. et al. The yeast GID complex, a novel ubiquitin ligase (E3) involved in the regulation of carbohydrate metabolism. *Mol. Biol. Cell* **19**, 3323–3333 (2008).
13. Liu, H. et al. The GID ubiquitin ligase complex is a regulator of AMPK activity and organismal lifespan. *Autophagy* **16**, 1618–1634 (2020).
14. Lampert, F. et al. The multi-subunit GID/CTLH E3 ubiquitin ligase promotes cell proliferation and targets the transcription factor Hbp1 for degradation. *Elife* **7**, https://doi.org/10.7554/eLife.35528 (2018).
15. Bagci, H. et al. The hGID^{GID4} E3 ubiquitin ligase complex targets ARHGAP11A to regulate cell migration. *bioRxiv*, 2023.2007.2020.549906, https://doi.org/10.1101/2023.07.20.549906 (2023).
16. Hantel, F. et al. Cilia-localized GID/CTLH ubiquitin ligase complex regulates protein homeostasis of sonic hedgehog signaling components. *J. Cell Sci.* **135**, https://doi.org/10.1242/jcs.259209 (2022).
17. Sherpa, D. et al. Modular UBE2H-CTLH E2-E3 complexes regulate erythroid maturation. *Elife* **11**, https://doi.org/10.7554/eLife.77937 (2022).
18. 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).
19. 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**, https://doi.org/10.1126/science.aal3655 (2017).
20. Dong, C. et al. Recognition of nonproline N-terminal residues by the Pro/N-degron pathway. *Proc. Natl Acad. Sci. USA* **117**, 14158–14167 (2020).
21. Mohamed, W. I. et al. The human GID complex engages two independent modules for substrate recruitment. *EMBO Rep.* **22**, e52981 (2021).
22. Zavortink, M. et al. The E2 Marie Kondo and the CTLH E3 ligase clear deposited RNA binding proteins during the maternal-to-zygotic transition. *Elife* **9**, https://doi.org/10.7554/eLife.53889 (2020).
23. Maitland, M. E. R. et al. The mammalian CTLH complex is an E3 ubiquitin ligase that targets its subunit muskelin for degradation. *Sci. Rep.* **9**, https://doi.org/10.1038/s41598-019-46279-5 (2019).
24. Braun, B. et al. Gid9, a second RING finger protein contributes to the ubiquitin ligase activity of the Gid complex required for catabolite degradation. *FEBS Lett.* **585**, 3856–3861 (2011).

25. Delto et al. The LisH Motif of Muskelin Is Crucial for Oligomerization and Governs Intracellular Localization. *Structure* **23**, 364–373 (2015).
26. Hosono, K. et al. YPEL5 protein of the YPEL gene family is involved in the cell cycle progression by interacting with two distinct proteins RanBPM and RanBP10. *Genomics* **96**, 102–111 (2010).
27. Kobayashi, N. et al. RanBPM, Muskelin, p48EMLP, p44CTLH, and the armadillo-repeat proteins ARMC8alpha and ARMC8beta are components of the CTLH complex. *Gene* **396**, 236–247 (2007).
28. Qiao, S. et al. Cryo-EM structures of Gid12-bound GID E3 reveal steric blockade as a mechanism inhibiting substrate ubiquitylation. *Nat. Commun.* **13**, 3041 (2022).
29. 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).
30. Francis, O., Han, F. & Adams, J. C. Molecular Phylogeny of a RING E3 Ubiquitin Ligase, Conserved in Eukaryotic Cells and Dominated by Homologous Components, the Muskelin/RanBPM/CTLH Complex. *PLoS ONE* **8**, e75217 (2013).
31. Soni, S. et al. Absence of erythroblast macrophage protein (Emp) leads to failure of erythroblast nuclear extrusion. *J. Biol. Chem.* **281**, 20181–20189 (2006).
32. Hodgkin, P. D., Lee, J. H. & Lyons, A. B. B cell differentiation and isotype switching is related to division cycle number. *J. Exp. Med.* **184**, 277–281 (1996).
33. Rush, J. S., Liu, M., Odegard, V. H., Unniraman, S. & Schatz, D. G. Expression of activation-induced cytidine deaminase is regulated by cell division, providing a mechanistic basis for division-linked class switch recombination. *Proc. Natl Acad. Sci. USA* **102**, 13242–13247 (2005).
34. 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).
35. 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).
36. 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).
37. Petroski, M. D. & Deshaies, R. J. Function and regulation of cullin-RING ubiquitin ligases. *Nat. Rev. Mol. Cell Biol.* **6**, 9–20 (2005).
38. Maitland, M. E. R., Lajoie, G. A., Shaw, G. S. & Schild-Poulter, C. Structural and Functional Insights into GID/CTLH E3 Ligase Complexes. *Int. J. Mol. Sci.* **23**, https://doi.org/10.3390/ijms23115863 (2022).
39. Cao, W. X. et al. Precise temporal regulation of post-transcriptional repressors is required for an orderly Drosophila maternal-to-zygotic transition. *Cell Rep.* **31**, 107783 (2020).
40. Pierce, N. W. et al. Cand1 promotes assembly of new SCF complexes through dynamic exchange of F box proteins. *Cell* **153**, 206–215 (2013).
41. Wu, K., Chen, A. & Pan, Z. Q. Conjugation of Nedd8 to CUL1 enhances the ability of the ROC1-CUL1 complex to promote ubiquitin polymerization. *J. Biol. Chem.* **275**, 32317–32324 (2000).
42. Alexandrov, L. B. et al. Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
43. Ramiro, A. R. et al. AID is required for c-myc/IgH chromosome translocations in vivo. *Cell* **118**, 431–438 (2004).
44. Yu, Y., Wang, Z., Zheng, Q. & Li, J. FAM72 serves as a biomarker of poor prognosis in human lung adenocarcinoma. *Aging (Albany NY)* **13**, 8155–8176 (2021).
45. Bai, Y., Cao, K., Zhang, P., Ma, J. & Zhu, J. Prognostic and Immunological Implications of FAM72A in Pan-Cancer and Functional Validations. *Int. J. Mol. Sci.* **24**, https://doi.org/10.3390/ijms24010375 (2022).
46. Huffman, N., Palmieri, D. & Coppola, V. The CTLH Complex in Cancer Cell Plasticity. *J. Oncol.* **2019**, 4216750 (2019).
47. Li, C. et al. The H2B deubiquitinase Usp22 promotes antibody class switch recombination by facilitating non-homologous end joining. *Nature Communications* **9**, https://doi.org/10.1038/s41467-018-03455-x (2018).
48. Maccarthy, T., Roa, S., Scharff, M. D. & Bergman, A. SHMTool: a webserver for comparative analysis of somatic hypermutation datasets. *DNA Repair (Amst.)* **8**, 137–141 (2009).
49. Boulianne, B. et al. AID and caspase 8 shape the germinal center response through apoptosis. *J. Immunol.* **191**, 5840–5847 (2013).
50. Weissmann, F. et al. biGBac enables rapid gene assembly for the expression of large multisubunit protein complexes. *Proc. Natl Acad. Sci. USA* **113**, E2564–E2569 (2016).
51. Aprosoff, C. M. et al. Comprehensive interactome mapping of the DNA repair scaffold SLX4 using proximity labeling and affinity purification. *J. Proteome Res.* **22**, 1660–1681 (2023).
52. Liu, G. et al. Data Independent Acquisition analysis in ProHits 4.0. *J. Proteom.* **149**, 64–68 (2016).
53. Perkins, D. N., Pappin, D. J., Creasy, D. M. & Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551–3567 (1999).
54. Eng, J. K., Jahan, T. A. & Hoopmann, M. R. Comet: an open-source MS/MS sequence database search tool. *Proteomics* **13**, 22–24 (2013).
55. Shteynberg, D. et al. iProphet: multi-level integrative analysis of shotgun proteomic data improves peptide and protein identification rates and error estimates. *Mol. Cell Proteom.* **10**, M111 007690 (2011).
56. Teo, G. et al. SAINTexpress: improvements and additional features in Significance Analysis of INTeractome software. *J. Proteom.* **100**, 37–43 (2014).
57. Knight, J. D. et al. A web-tool for visualizing quantitative protein-protein interaction data. *Proteomics* **15**, 1432–1436 (2015).
58. Zhao, Y., Chapman, D. A. & Jones, I. M. Improving baculovirus recombination. *Nucleic Acids Res.* **31**, E6–E6 (2003).
59. Ceccarelli, D. F. et al. An allosteric inhibitor of the human Cdc34 ubiquitin-conjugating enzyme. *Cell* **145**, 1075–1087 (2011).
60. Pickart, C. M. & Raasi, S. Controlled synthesis of polyubiquitin chains. *Methods Enzymol.* **399**, 21–36 (2005).
61. Hesketh, G. G. et al. The GATOR-Rag GTPase pathway inhibits mTORC1 activation by lysosome-derived amino acids. *Science* **370**, 351–356 (2020).
62. 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).
63. Liu, G. et al. ProHits: integrated software for mass spectrometry-based interaction proteomics. *Nat. Biotechnol.* **28**, 1015–1017 (2010).
64. Suloway, C. et al. Automated molecular microscopy: the new Leginon system. *J. Struct. Biol.* **151**, 41–60 (2005).
65. Wagner, T. et al. SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Commun. Biol.* **2**, 218 (2019).
66. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
67. Bepler, T. et al. Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nat. Methods* **16**, 1153–1160 (2019).
68. Sanchez-Garcia, R. et al. DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. *Commun. Biol.* **4**, 874 (2021).
69. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

70. Pettersen, E. F. et al. UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput Chem.* **25**, 1605–1612 (2004).
71. 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).
72. Croll, T. I. ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr D. Struct. Biol.* **74**, 519–530 (2018).
73. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of coot. *Acta Crystallogr D. Biol. Crystallogr* **66**, 486–501 (2010).
74. Li, T. et al. TIMER: a web server for comprehensive analysis of tumor-infiltrating immune cells. *Cancer Res* **77**, e108–e110 (2017).
75. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
76. Li, B. et al. Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. *Genome Biol.* **17**, 174 (2016).

Acknowledgements

This work was supported by grants from the Canadian Institutes of Health Research grants PJT-153307 and PJT-156330 to AM and PJT-178026, PJT-180338 and PJT- 186218 to FS. The work in the laboratory of AB was supported by National Institutes of Health grant 1R21CA252858-01A1. PB is supported by The Queen Elizabeth II Graduate Scholarship in Science and Technology. We thank Dr. Zhijie Li and Dr. Samir Benlekbir (The Hospital for Sick Children, Toronto) for help with Cryo-EM data collection and processing. Some Cryo-EM data was collected at Toronto High Resolution High-Throughput Cryo-EM facility, supported by the Canada Foundation for Innovation and Ontario Research Fund. Some Cryo-EM data was collected at NYU Langone Health's Cryo-Electron Microscopy Laboratory (RRID: SCR_019202), which is partially supported by the Laura and Isaac Perlmutter Cancer Center Support Grant NIH/NCI P30CA016087. SEC-MALS data was collected at the Structural and Biophysical Core Facility (The Hospital for Sick Children, Toronto) with help from Gregory A. Wasney and James M. Jorgensen. We thank Dr. Nathalie Simard, Vincent Cheng, and Parva Thakker at the Temerty Faculty of Medicine Flow Cytometry Facility (University of Toronto; Toronto, Ontario, Canada) for operating the cell sorters.

Author contributions

P.B., C.K.C., F.S. and A.M. designed the project. P.B., C.K.C. M.K.W., I.B.M., J.P.B., Y. Feng, A.F.A.K., C.W., R.M.R., L.C.M., M.A.K., D.F.C., S.O., Y. Fang. and H.K. performed experiments. M.M.J., R.C.P, A.S.B., T.J.P, A.C.G., F.S. and A.M. provided supervision. P.B., C.K.C., F.S. and A.M. wrote the original draft.

Competing interests

FS is a founder and consultant for Repare Therapeutics and Induxion Therapeutics. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-52009-x.

Correspondence and requests for materials should be addressed to Frank Sicheri or Alberto Martin.

Peer review information *Nature Communications* thanks Weiwei Peng, Bernardo Reina-San-Martin and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.