Supplementary Materials for

A glycine-specific N-end rule pathway mediates the quality control of protein *N*-myristoylation

Richard T. Timms, Zhiqian Zhang, David Y. Rhee, J. Wade Harper, Itay Koren, Stephen J. Elledge

correspondence to: ikoren@bwh.harvard.edu, selledge@genetics.med.harvard.edu

This PDF file includes:

Materials and Methods
Figs. S1 to S16
Captions for databases S1 to S7

Other Supplementary Materials for this manuscript includes the following:

Databases S1 to S7 as zipped archives

Materials and Methods

Cell Culture

HEK-293T (ATCC® CRL-3216™) cells were grown in Dulbecco's Modified Eagle's Medium (DMEM) (Life Technologies) supplemented with 10% fetal bovine serum (HyClone) and penicillin/streptomycin (Thermo Fisher Scientific).

Transfection and lentivirus production

Lentivirus was generated through the transfection of HEK-293T cells using PolyJet In Vitro DNA Transfection Reagent (SignaGen Laboratories). Cells seeded at approximately 80% confluency were transfected as recommended by the manufacturer with the lentiviral transfer vector plus four plasmids encoding Gag-Pol, Rev, Tat and VSV-G. The media was changed 24 h post-transfection and lentiviral supernatants collected a further 24 h later. Cell debris was removed by centrifugation (800 x *g*, 5 min) and virus was stored in single-use aliquots at -80°C. Transduction of target cells was achieved by adding the virus in the presence of 8 μg/ml hexadimethrine bromide (Polybrene, Sigma-Aldrich).

Inhibitors

The proteasome inhibitor Bortezomib was obtained from APExBio and the pan-CRL inhibitor MLN4924 was obtained from Active Biochem; both were used at a final concentration of 1 μM. Cyclohexmide was obtained from Merck. The NMT1/2 inhibitor IMP-1088 was purchased from Cayman Chemical and was used at a final concentration of 1 μM for 24 h.

Antibodies

Primary antibodies used in this study were: rabbit anti-UBR1 (Bethyl, A302-988A), rabbit anti-UBR2 (Bethyl, A305-416A), rabbit anti-UBR4 (Bethyl, A302-278A), rabbit anti-GFP (Abcam, ab290), rabbit anti-FOXJ3 (Bethyl, A303-107A), rabbit anti-ALKBH1 (Abcam, ab195376), rabbit anti-CHMP3 (Bethyl, A305-397A), mouse anti-vinculin (Sigma, V9131), rabbit anti-Fyn (Cell Signaling, 4023T), rabbit anti-LAMTOR (Cell Signaling, 8975T), rabbit anti-Yes (Cell Signaling, 3201S), rabbit anti-Lyn (Bethyl, A302-683A-T), rabbit anti-NDUFAF4 (ABclonal, A14345), rabbit anti-Src (Cell Signaling, 2123T) and rabbit anti-GAPDH (Cell Signaling, 5174). The HA and FLAG epitope tags were detected using rat anti-HA peroxidase (Sigma-Aldrich, 12013819001) and rabbit anti-FLAG peroxidase (Cell Signaling, #2044S). HRP-conjugated goat anti-rabbit IgG secondary antibody was obtained from Jackson ImmunoResearch (#111-035-003).

Plasmids

Lentiviral vectors encoding dominant-negative Cullin constructs were a generous gift from W. Harper. For exogenous expression of CRL2 substrate adaptors, the pHRSIN-P_{SFFV}-GFP-WPRE-P_{PGK}-Hygro vector was used (a gift from P. Lehner), with the constructs cloned in in place of GFP using the Gibson assembly method (NEBuilder HiFi Cloning Kit). Plasmids encoding ZYG11A and ZYG11B were obtained from Addgene (plasmids #110550 and #110551, deposited by E. Kipreos (*47*)), while an entry

vector encoding ZER1 was obtained from the Ultimate ORF Clone collection (Thermo Fisher Scientific). A plasmid encoding TEV protease was also obtained from Addgene (plasmid #64276, deposited by X. Shu (*28*)).

For individual CRISPR/Cas9-mediated gene disruption experiments, the lentiCRISPR v2 vector was used (Addgene #52961, deposited by Feng Zhang). Oligonucleotides encoding the top and bottom strands of the sgRNAs were synthesized (IDT), annealed and phosphorylated (T4 PNK; NEB) and cloned into the lentiCRISPR v2 vector as described (*48*). Nucleotide sequences of the sgRNAs used were:

sg-AAVS1: GGGGCCACTAGGGACAGGAT
sg1-UBR1: GTGAGAGGATGGAAATCAGCG
sg2-UBR1: GATTCTAACTTGTGGACCGAA
sg1-UBR2: GAGGAGGAGAGAAGATGGCGT
sg2-UBR2: GTACCCAAAATCTACTGCAG
sg1-UBR4: GCCTCTCGAAGATGAACACCG
sg2-UBR4: GCTGACCCCTGGACAGACAG
sg-ATE1: GTATCAGGATCTCATAGACCG
sg1-ZYG11B: GCGCTCGTAAGGATCCTCGA
sg2-ZYG11B: GAAGCTCGAAGGCCAGAAAGC
sg1-ZER1: GTATGAGGAGGAGAACCCAGG
sg2-ZER1: GCCGCAGCAGGGACTCCACA
sg1-NMT1: GCAGGGTGTAGGGCTCCTGG
sg2-NMT1: GAAGCTCTACCGACTGCCAG
sg1-NMT2: GATAGACGGGGACAATGAGG
sg2-NMT2: GGACACGTGCGGGATAGACG

Flow cytometry

Analysis of HEK-293T cells by flow cytometry was performed on a BD LSRII instrument (Becton Dickinson) and the resulting data was analyzed using FlowJo. Cell sorting was performed on a MoFlo Astrios (Beckman Coulter).

Generation of Ub-GPS libraries

Human protein coding sequences were downloaded from the Gencode database (release 27). The first 72 nucleotides of entries that (1) started with a methionine residue, (2) had a transcript_support_level equal to 1 or 2 and (3) were common to both the Ensembl and Havana databases were included in the oligonucleotide design. After removal of identical 72-mer sequences, the final N-terminome library consisted of a total of 24,638 sequences. The oligonucleotide pool was synthesized by Agilent Technologies and amplified by PCR (Q5 Hot Start High-Fidelity DNA Polymerase, NEB). The PCR product was then cloned into the Ub-GPS vector between a unique SalI site engineered into the 3' end of the ubiquitin gene and a unique NdeI site at the 5'end of GFP using the Gibson assembly method (NEBuilder HiFi Cloning Kit), such that the resulting vector encoded the peptides immediately downstream of ubiquitin, followed by a short linker (ATSALGT) and GFP (commencing SKGEEL-). At least 100-fold representation of the library was maintained at each step.

Ub-GPS libraries for saturation mutagenesis were generated in an identical manner. For each peptide selected for analysis, each amino acid encoded at position 2 through to position 10 was mutated to all other possible 19 amino acids. For each peptide 9 reference sequences were also synthesized, in which the same wild-type amino acid sequence was encoded by different nucleotide sequences.

Two databases were used to collate cleavage products for the caspase cleavage site Ub-GPS library: Degrabase (*25*) and PROSPER (*26*). All annotated cleavage sites in human proteins occurring after aspartic acid in each databases were included for oligonucleotide design, which, after removal of duplicates, resulted in a total of 2,234 sequences. Amino acid sequence were converted into nucleotide sequences using the following codons:
A: GCC, C: TGC, D: GAC, E: GAG, F: TTC, G: GGC, H: CAC, I: ATC, K: AAG, L: CTG, M: ATG, N: AAC, P: CCC, Q: CAG, R: AGA, S: TCC, T: ACC, V: GTG, W: TGG and Y: TAC. The oligonucleotide pool was synthesized by Agilent Technologies and amplified and cloned into the Ub-GPS vector as above.

Ub-GPS screens

GPS plasmid libraries were packaged into lentiviral particles which were used to transduce HEK-293T cells at a multiplicity of infection of ~0.2 (achieving approximately 20% $DsRed^+$ cells) and at sufficient scale to achieve ~500-fold coverage of the library (a total of ~12 million transduced cells in the case of the human N-terminome library). Puromycin (1.5 μg/ml) was added two days post-transduction to eliminate untransduced cells. Surviving cells were pooled, expanded and then partitioned by FACS into six bins 7 days post-transduction based on the GFP/DsRed ratio. Genomic DNA was extracted from each of the pools separately (Gentra Puregene Cell Kit, Qiagen) and the fusion peptides amplified by PCR (Q5 Hot Start Polymerase, NEB) using a forward primer annealing to the end of the ubiquitin gene and a reverse primer annealing to the front of GFP; sufficient reactions were performed to amplify a total mass of DNA equivalent to the mass of genomic DNA from cells representing 500-fold coverage of the library. All PCR products were pooled, and one-tenth of the mix was purified using a spin column (Qiagen PCR purification kit). Finally, 200 ng of the purified PCR product was used as the template for a second PCR reaction using primers to add the Illumina P5 sequence and a 7 bp 'stagger' region to the 5' end, and Illumina indexes and P7 sequence at the 3' end. Samples to be multiplexed were then pooled, purified on an agarose gel (QIAEXII Gel Extraction Kit, Qiagen) and sequenced on an Illumina NextSeq instrument.

CRISPR screens

A custom sgRNA library was designed targeting 43 E2 enzymes, 11 core CRL components and 109 CRL2/5 adaptors at a depth of 6 sgRNAs per gene. The sgRNA sequences together with flanking BbsI restriction enzymes recognition sites were synthesized by Twist Bioscience. The oligonucleotide pool was amplified by PCR (Q5 Hot Start Polymerase, NEB) and the product purified (Qiagen PCR purification kit) and digested with BbsI (NEB). The digested product was concentrated by ethanol precipitation and then visualized on a 10% TBE PAGE gel (Thermo Fisher Scientific) stained with SYBR Gold (Thermo Fisher Scientific). DNA was isolated from the 28 bp

band using the 'crush-and-soak' method, concentrated by ethanol precipitation, and then cloned into lentiCRISPR v2 (Addgene #52961) digested with BsmBI (NEB).

The sgRNA library DNA was packaged into lentiviral particles. HEK-293T cells stably expressing unstable peptide-GFP fusion proteins were transduced at a multiplicity of infection of ~0.3 at sufficient scale to maintain at least 1000-fold representation of the library. Untransduced cells were eliminated through puromycin selection commencing two days post-transduction. The top ~5% of the surviving cells based on the GFP/DsRed ratio were isolated by FACS, which was performed 7 days post-transduction. For each screen genomic DNA was extracted from both the sorted cells and the unselected library as a reference. The sgRNAs in both pools were amplified by PCR and sequenced on an Illumina NextSeq instrument.

Immunoprecipitation and immunoblotting

HEK-293T cells stably expressing epitope-tagged CRL2 substrate adaptors and peptide-GFP fusions were grown in 10 cm plates. Following treatment with Bortezomib (1 μM, 5 h), cells were lysed in ice-cold lysis buffer (50 mM Tris, 100 mM NaCl, 0.5% NP-40, pH 7.5 supplemented with EDTA-free protease inhibitor tablet and Phos-Stop phosphatase inhibitor tablet (Roche)) for 30 min on ice. Nuclei were pelleted by centrifugation (14,000 x *g*, 10 min, 4°C). Beads coated with anti-HA (Pierce anti-HA magnetic beads, Thermo Fisher Scientific) or anti-FLAG (anti-FLAG M2 magnetic beads, Sigma-Aldrich) antibodies were added to the supernatants and incubated with rotation overnight at 4°C. The beads were then washed three times with lysis buffer before bound proteins were eluted upon incubation with SDS-PAGE sample buffer (95°C, 10 min). Proteins were subsequently resolved by SDS-PAGE (NuPAGE Bis-Tris gels, Thermo Fisher Scientific) and transferred to a nitrocellulose membrane (Trans-Blot Turbo System, Bio-Rad) which was then blocked in 10% nonfat dry milk in PBS + 0.1% Tween-20 (PBS-T). The membrane was incubated with primary antibody overnight at 4°C, and then, following three washes with PBS-T, HRP-conjugated secondary antibody was added for 1 h at room temperature. Following a further three washes in PBS-T, reactive bands were visualized using Western Lightning Plus ECL (Perkin Elmer) and HyBlot CL film (Denville Scientific).

Mass spectrometry

UBR KO clone #2 cells stably expressing peptide-GFP fusions growing in 15 cm plates were lysed as described above, and immunoprecipitation performed in a similar way using GFP-Trap_MA magnetic agarose beads (Chromotek). Elution of the peptide-GFP fusion proteins was achieved by treatment with 2 M glycine for 1 min, followed by neutralization with 1 M Tris base, pH 10.4. Eluted proteins were reduced using DTT (Thermo Fisher) and alkylated with iodoacetamide (Sigma). Following TCA precipitation (Sigma), proteins were digested with Glu-C (Thermo Fisher) then cleaned up on C-18 stage tips (3M).

Mass spectrometry data were collected using a Q Exactive mass spectrometer (Thermo Fisher) coupled with a Famos Autosampler (LC Packings) and an Accela600 liquid chromatography pump (Thermo Fisher). Peptides were separated on a 100 μm inner diameter microcapillary column packed with ~25 cm of Accucore C18 resin (2.6 μm, 150 Å, Thermo Fisher). Peptides were separated using a 120 gradient of 5 to 25%

acetonitrile in 0.125% formic acid at a flow rate of ~300 nl/min. The scan sequence began with an Orbitrap MS1 spectrum with resolution 70,000, scan range 300−1500 Th, automatic gain control (AGC) target 1×10^5, maximum injection time 250 ms, and centroided data type. The top twenty precursors were selected for MS2 analysis which consisted of HCD (high-energy collision dissociation) with the following parameters: resolution 17,500, AGC 1×10^5, maximum injection time 100 ms, isolation window 1.6 Th, normalized collision energy (NCE) 27, and centroid spectrum data type. Unassigned charge states were excluded from MS^2 analysis, but singly charged species were included. Dynamic exclusion was set to automatic. Mass spectra were processed using a Sequest-based in-house software pipeline.

Bioinformatics

N-terminome Ub-GPS screen. Raw Illumina reads derived from each GPS bin were first trimmed of constant sequences derived from the Ub-GPS vector backbone using Cutadapt (*49*). Resulting 72 nt reads were mapped to the reference input library using Bowtie 2 (*50*), and count tables were generated from reads that aligned perfectly to the reference sequence. Following correction for sequencing depth, the protein stability index (PSI) metric was calculated for each peptide-GFP fusion. The PSI score is given by the sum of multiplying the proportion of reads in each bin by the bin number (1-6 in this case), thus yielding a stability score between 1 (maximally unstable) and 6 (maximally unstable):

PSI = $\sum_{i=1}^{6} Ri * i$

(where i represents the number of the bin and R_i represents the proportion of Illumina reads present for a peptide in that given bin i). Read counts and associated stability score for each peptide-GFP fusion are detailed in **data S1**.

Prediction of destabilizing N-terminal motifs. The stability data derived from the Ub-GPS N-terminome screen was used to identify potential destabilizing N-terminal degron motifs **(Fig. 2, C and D)**. Varying exactly two residues at a time between position 2 and position 7, for all possible combinations of di-peptide motifs (allowing gaps) the mean PSI of all peptides containing that motif at the front of the peptide (that is, immediately following the initiator methionine) was compared to the mean PSI of all peptides containing the same motif at any internal location within the 24-mer peptide. The full data for all possible N-terminal motifs is tabulated in **data S2**.

N-terminome Ub-GPS screen in different genetic backgrounds. The Ub-GPS N-terminome screens with the UBR mutant clones and the ZYG11B and ZER1 mutant cells were performed in a similar manner as above, except that only the half of the N-terminome library comprising peptides bearing an initiator methionine was used. Read counts for all peptide-GFP fusions are detailed in **data S3A**. Subsequently, comparisons between wild-type cells and combined ZYG11B/ZER1 mutant cells **(data S3B)** and between wild-type, control (AAVS1) mutant and three NMT1/2 mutant clones **(data S3C)** were performed in a similar way. In each case, a ΔPSI score was generated for each peptide-GFP fusion reflecting the difference in raw PSI scores between the wild-type or control mutant cells and the experimental mutant cells. For the plots shown in Fig. 3 and fig. S3, peptide-GFP fusions were defined as UBR substrates if they were stabilized >0.8

PSI units **(Fig. 3E)** or >0.6 PSI units **(fig. S3A-D)** in the UBR KO clone compared to control cells, but also not stabilized >0.3 PSI units in either ZYG11B or ZER1 mutant cells. The logoplots shown in **fig. S3E-H** were generated with iceLogo (*51*) and rendered using Seq2Logo (*52*): the 'experimental set' comprised all peptides starting with the indicated motif that were identified as a UBR substrate in any of the three UBR KO clones, the 'reference set' comprised all peptides in the N-terminome library starting with that same motif, and the 'percentage difference' scoring system was used. Residues significantly enriched at $P < 0.05$ are displayed. For the heatmap shown in **Fig. 7D**, peptide-GFP fusions destabilized >0.5 PSI units in all three NMT1/2 mutant clones were included for analysis.

Saturation mutagenesis Ub-GPS screens. The heatmaps displayed in **Fig. 3G-I, fig. S4A-C, Fig. 4B and fig. S5** illustrate the difference between the PSI for each individual mutant peptide and the median PSI of all the unmutated peptides; the darker the red color, the greater the stabilizing effect of the mutation. For the heatmaps shown in **Fig. 5A-C and fig. S13** the color scales indicate the raw PSI stability measurement, which lies between 1 (maximally unstable; dark blue) and 6 (maximally stable; dark red); the exception is the comparison between ZYG11B mutant and ZER1 mutant cells (right columns) where a ΔPSI score reflecting the difference between raw PSI score in ZYG11B mutant cells and ZER1 mutant cells for each peptide is depicted. The full data for all mutant peptides in all genetic backgrounds is detailed in **data S4A-C**.

N-terminome Ub-GPS screen with MLN4924. The Ub-GPS N-terminome screen with MLN4924 described in **fig. S6** was analyzed in the same way as above, ultimately yielding a ΔPSI score reflecting the difference in raw PSI scores between the untreated and MLN4924-treated conditions **(data S5)**. For the heatmap shown in **fig. S6E**, the intensity of the colors represent the depletion (blue) or enrichment (red) of each amino acid comparing the pool of CRL substrates to all peptides detected in the N-terminome library.

CRISPR screens. Constant regions derived from the backbone of the lentiCRISPR v2 expression vector were removed from Illumina reads using Cutadapt, and count tables were generated from the remaining variable portion of the sgRNA sequences using Bowtie 2. The Model-based Analysis of Genome-wide CRISPR/Cas9 Knockout (MAGeCK) algorithm (*46*) was used to rank the performance of individual genes targeted by multiple sgRNAs enriched in the selected cells versus the unsorted populations. The full MAGeCK output for each screen is detailed in **data S6**. For the scatterplots shown in **Fig. 4E** and **fig. S8B**, the MAGeCK score plotted on the y-axis is calculated as the negative log10 of the 'pos|score' value generated by MAGeCK.

Proteome composition analysis. Canonical protein sequences were downloaded from the Swiss-Prot database. For each position between position 2 and position 10 at the N-terminus of the proteins the total abundance of each amino acid was quantified, expressed as a proportion of the total number of protein sequences, and then normalized to the mean proportion observed across the 9 N-terminal residues between position 2 and position 10. For the analysis of N-terminal glycine degrons shown in **Fig. 5, F and G**, we

further categorized glycine residues as either 'favored' for CRL2-mediated degradation through ZYG11B and ZER1 if they were followed by F, G, H, L, M or Y, or 'disfavored' if followed by D, E, I, N, P, R, S or T. The mean abundance of $G_{favored}$ and $G_{disfavored}$ across the 8 N-terminal residues between position 3 and position 10 was then compared to their abundance at position 2. For the analysis of human caspase cleavage sites presented in **Fig. 6A**, all unique cleavage sites occurring downstream of aspartic acid (D) annotated in Degrabase 1.0 (*25*) were analyzed using iceLogo. In **Fig. 6B**, all unique cleavage sites occurring downstream of aspartic acid (D) and upstream of glycine (G) were analyzed; the frequency of $G_{favored}$ and $G_{disfavored}$ at the N-terminus of these caspase cleavage products was compared to the frequency of $G_{favored}$ and $G_{disfavored}$ at all glycine residues in the human proteome.

Caspase cleavage site Ub-GPS screen. The Ub-GPS screen with peptides derived from human caspase cleavage sites was analyzed in the same way as above, generating a ΔPSI score for each peptide reflecting the difference in raw PSI scores between either control (AAVS1) mutant cells or combined ZYG11B/ZER1 double mutant cells versus wild-type cells **(data S7)**. For the heatmap shown in **Fig. 6D**, peptide-GFP fusions stabilized >0.5 PSI units in both ZYG11B/ZER1 double mutant cell lines but <0.25 PSI units in control knockout cells were included; the intensity of the colors represent the depletion (blue) or enrichment (red) of each amino acid comparing this pool of ZYG11B/ZER1 substrates to all peptides detected in the caspase cleavage site library.

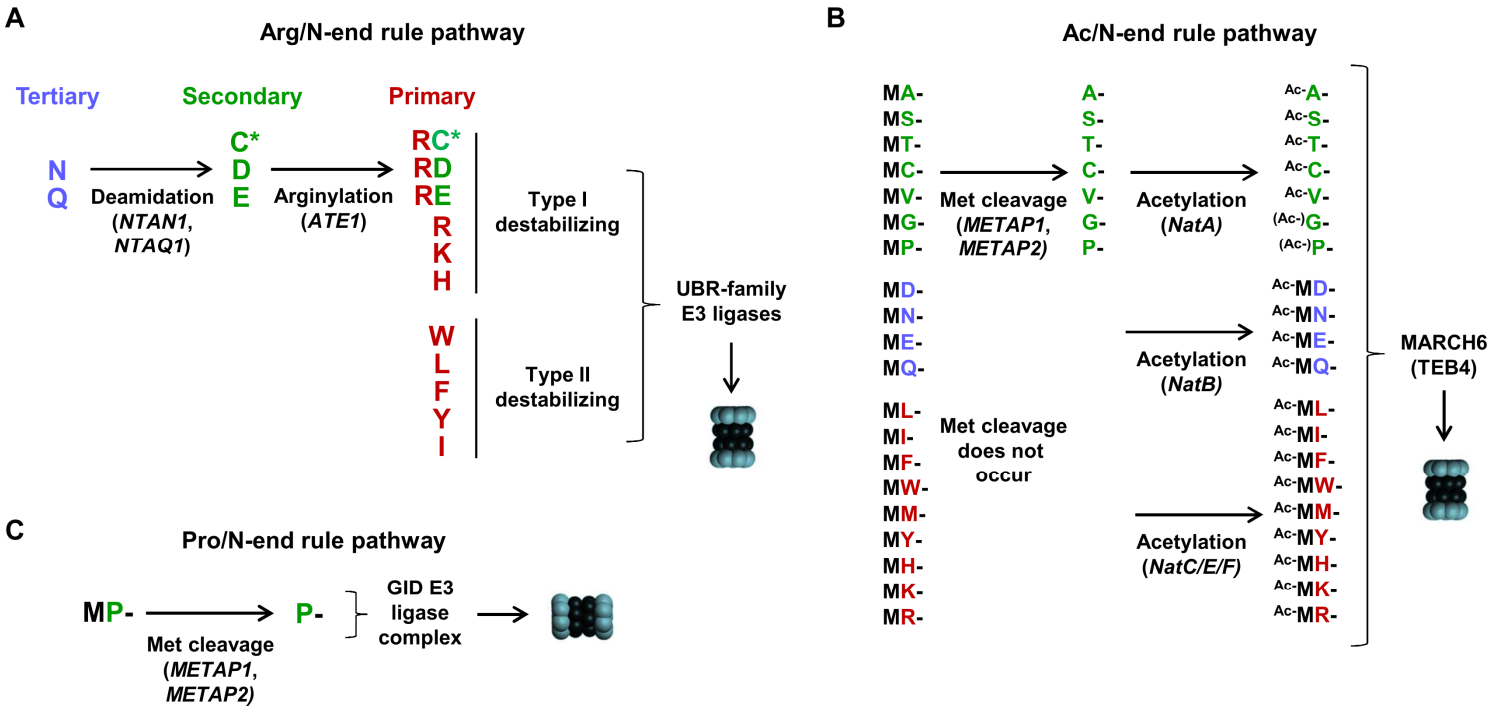

Fig. S1. Summary of known N-end rule pathways.

(A) The Arg/N-end rule. Proteins harboring primary type I (basic) or type II (bulky hydrophobic) residues at their N-terminus are targeted for proteasomal degradation by UBR family proteins. Secondary (D and E) destabilizing residues at the N-terminus can also become Arg/N-end rule substrates through N-terminal arginylation, as can tertiary destabilizing residues (N and Q) at the N-terminus following prior deamidation (*6*). Furthermore, oxidation of N-terminal cysteine by nitric oxide, indicated by an asterisk, can render it a substrate for arginylation and hence UBR-mediated degradation (*17*).

(B) The Ac/N-end rule. Acetylation of N-terminal residues can create degrons recognized by the MARCH6 E3 ubiquitin ligase (*44*). Cleavage of the initiator methionine upstream of amino acids with sufficiently small side chains results (colored green) in the exposure of the second residue, which can be a substrate for acetylation on the α-amino group mediated by NatA. The uncleaved initiator methionine upstream of the other amino acids (colored blue and red) can also serve as a substrates for other Nat enzymes (*45*). It is estimated that up to 80% of all human proteins are acetylated to some extent, with the degree of acetylation varying depending on the sequence context; N-terminal alanine and serine are thought to be almost completely acetylated, while N-terminal glycine and proline are rarely acetylated (*45*).

(C) The Pro/N-end rule. Proline exposed at the N-terminus of proteins following methionine cleavage can act as degron recognized by GID4, the substrate recognition component of the multi-subunit GID E3 ligase complex (*9*).

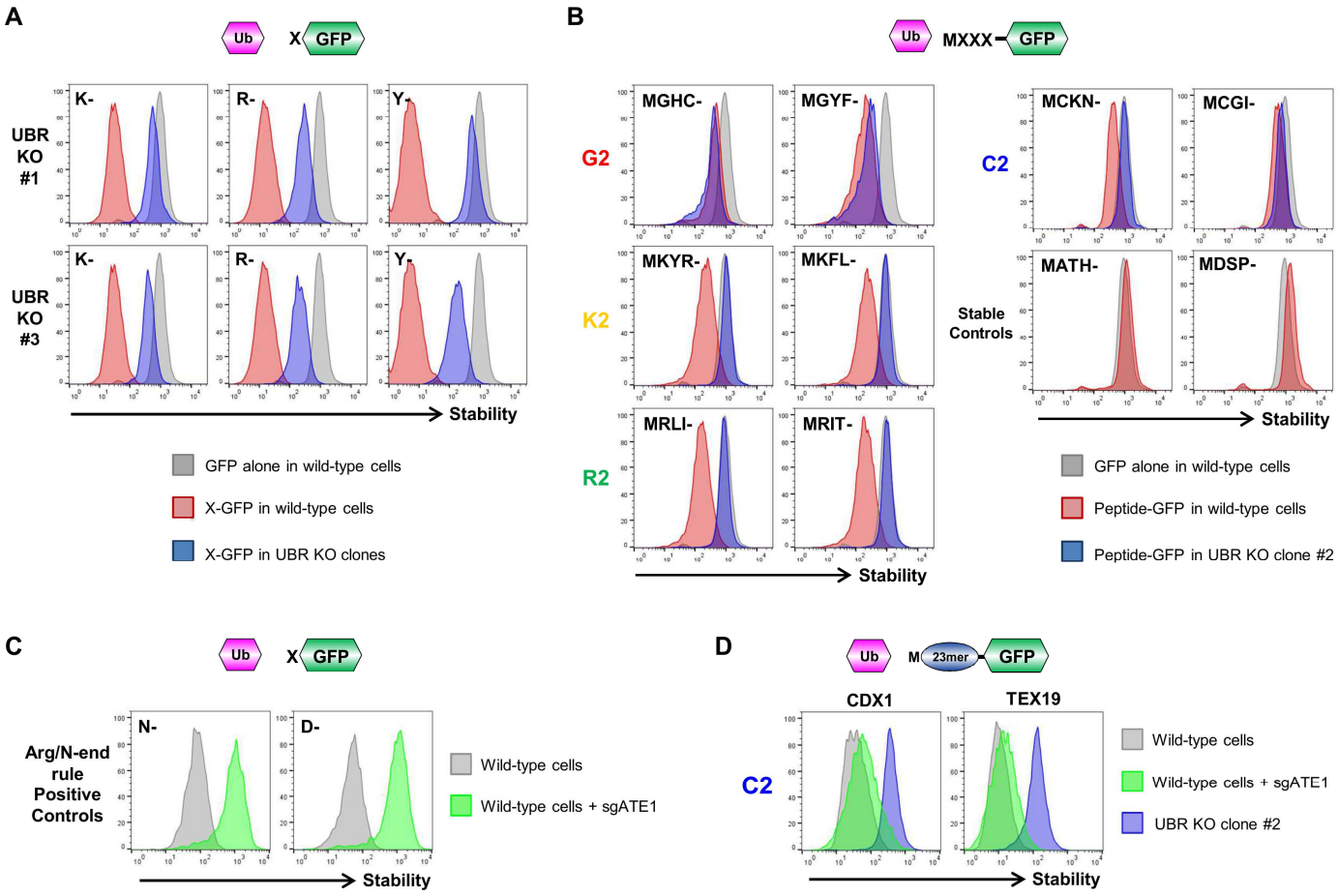

Fig. S2. Assessing the role of UBR family proteins in the degradation of proteins bearing N-terminal degron motifs.

(A) Functional validation of UBR KO clone #1 and clone #3. Optimal Arg/N-end rule substrates expressed in the Ub-GPS vector were highly unstable in wild-type HEK-293T cells, but were stabilized in the UBR KO clones, as measured by flow cytometry.

(B) Assessing the role of UBR proteins in the degradation of proteins bearing candidate N-terminal degron motifs. Three amino acids 'triplets' containing putative N-terminal degrons were fused to the N-terminus of GFP and expressed in the context of the Ub-GPS system in wild-type cells or UBR KO clone #2. In wild-type cells, all of the substrates (red histograms) exhibited some degree of instability compared to GFP alone (gray histograms). For three of the motifs – MR-, MK- and N-terminal cysteine – stabilization to the level of GFP alone was observed in UBR KO clone #2, whereas little or no stabilization was observed for the fusions harboring N-terminal glycine.

(C and D) Assessing the role of N-terminal arginylation in the degradation of substrates bearing N-terminal cysteine. **(C)** Functional validation of efficient ATE1 disruption. Ub-GPS substrates bearing tertiary (N) or secondary (D) Arg/N-end rule degrons at their N-terminus were unstable in wild-type HEK-293T cells, but exhibited marked stabilization upon CRISPR/Cas9-mediated ablation of the arginyltransferase ATE1. **(D)** Loss of ATE1 does not fully stabilize unstable substrates bearing N-terminal cysteine. CRISPR-mediated disruption of ATE1 resulted in only partial stabilization of two example Ub-GPS substrates exposing N-terminal cysteine, which did not reach the level of stabilization achieved in UBR KO clone #2.

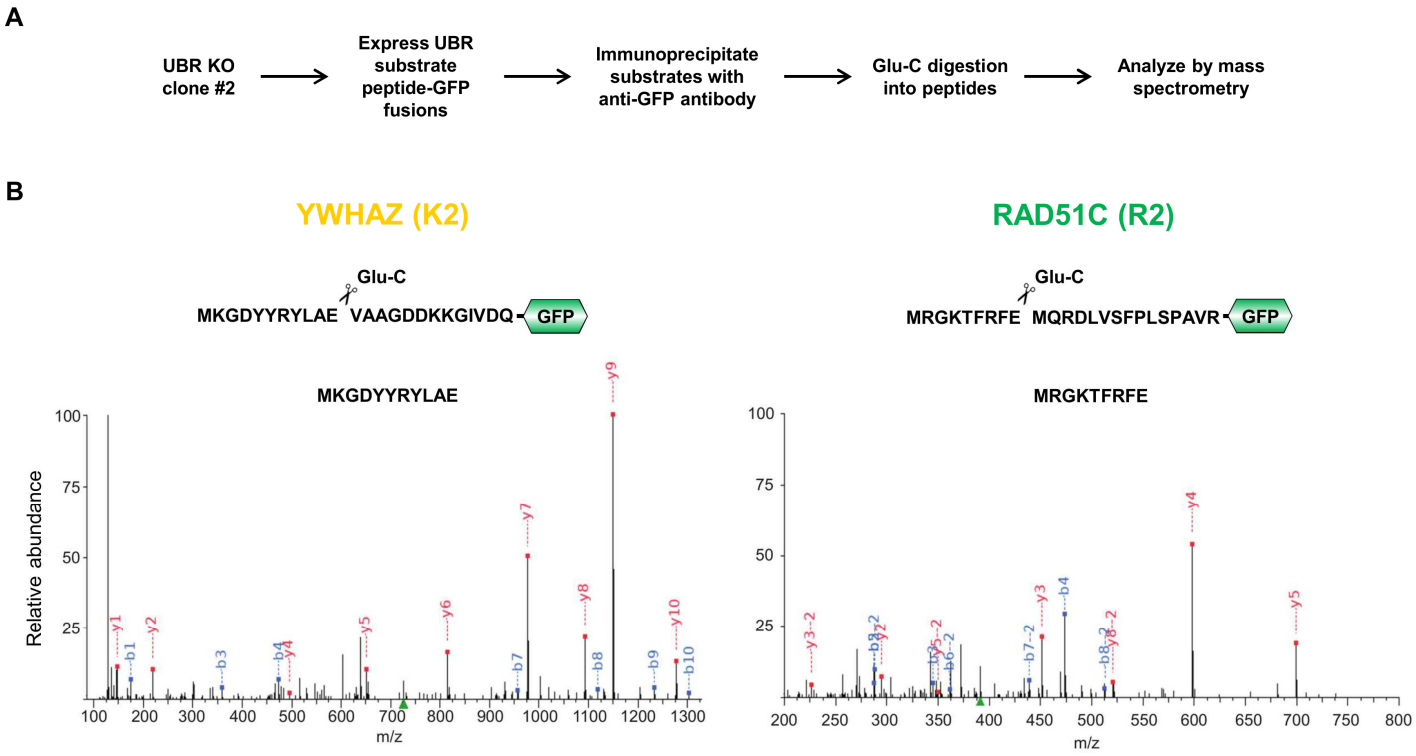

Fig. S3. UBR substrates commencing MK- and MR- retain their initiator methionine.

(A) Overview of the experiment. Peptide-GFP UBR substrates were expressed in UBR KO clone #2, immunoprecipitated using an anti-GFP antibody and then analyzed by mass spectrometry.

(B) Mass spectra showing the N-terminal peptides identified from YWHAZ (commencing MK-) and RAD51C (commencing MR-). In each case, only species with the initiator methionine intact were detected.

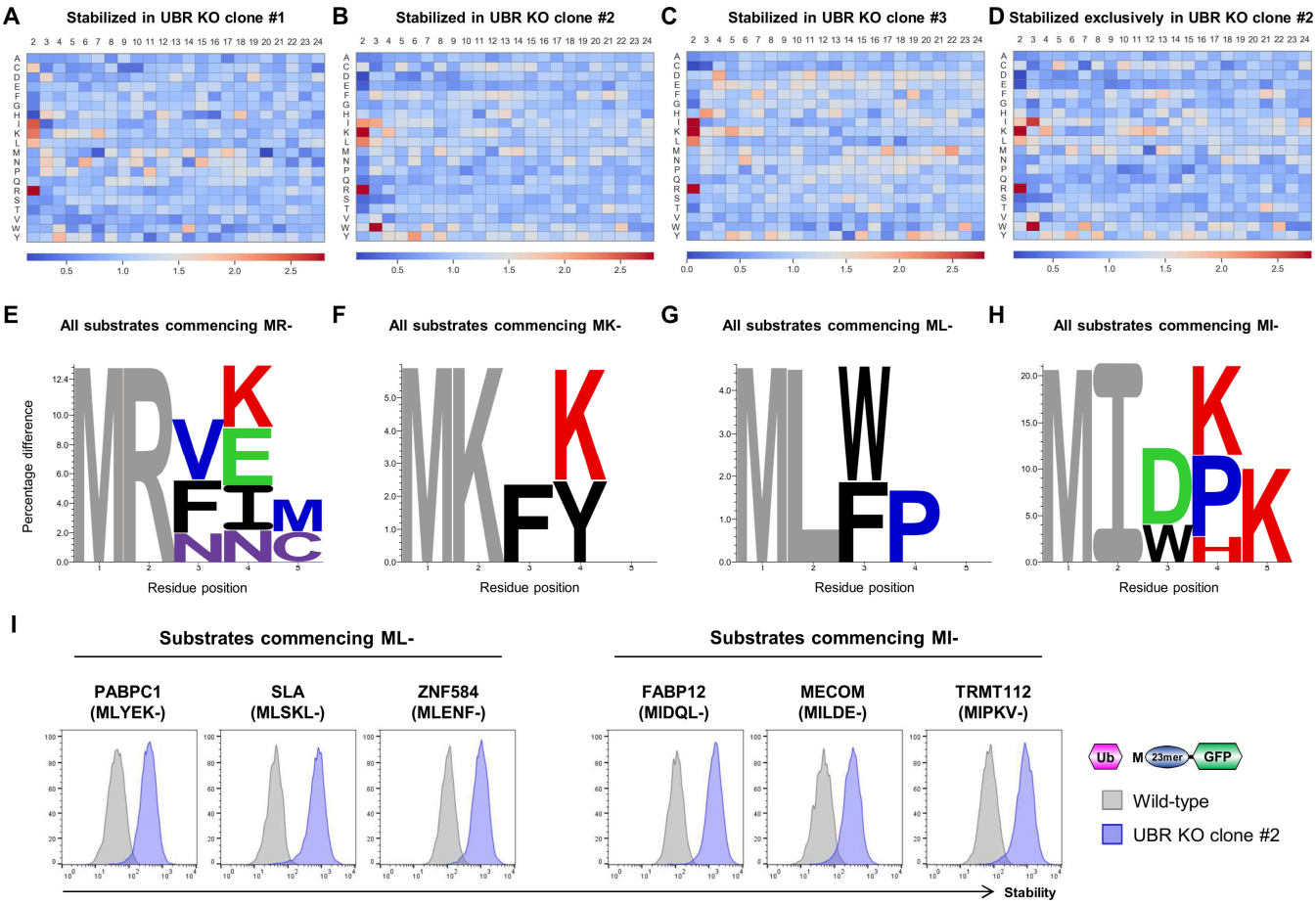

Fig. S4. Global identification of UBR substrates across the human N-terminome.

(A-C) Sequence composition of UBR substrates. The heatmaps show the relative enrichment (red) or depletion (blue) of each amino acid across all positions of the 23-mer peptide comparing peptides stabilized >0.6 PSI units in UBR KO clone #1 **(A)**, clone #2 **(B)** or clone #3 **(C)** relative to the whole N-terminome library.

(D) Assessing potential UBR4 substrates among the human N-terminome. As we observed a large number of reporters that were exclusively stabilized in UBR clone #2, which was the only clone that lacked UBR4 expression, we also analyzed the composition of these peptides separately: the heatmap shows the relative enrichment (red) or depletion (blue) of each amino acid across all positions of the 23-mer peptide comparing peptides stabilized >0.6 PSI units in UBR KO clone #2 but not in either clone #1 or clone #3. The similar preferences observed suggested that UBR4 might collaborate with UBR1 and UBR2 to enhance the degradation of their substrates, thus resulting in a greater number of peptide-GFP fusions showing significant stabilization.

(E-H) Assessing the sequence context of UBR substrates. Logoplots highlighting residues significantly enriched at positions 3, 4 or 5, comparing all peptides identified as a UBR substrate in any of the UBR KO clones commencing MR- **(E)**, MK- **(F)**, ML- **(G)** or MI- **(H)** relative to the whole N-terminome library.

(I) UBR proteins target substrates commencing ML- and MI-. The first 24 residues of the indicated proteins were fused to the N-terminus of GFP and expressed in wild-type cells and UBR KO clone #2, and the stability of the reporter constructs were assessed by flow cytometry.

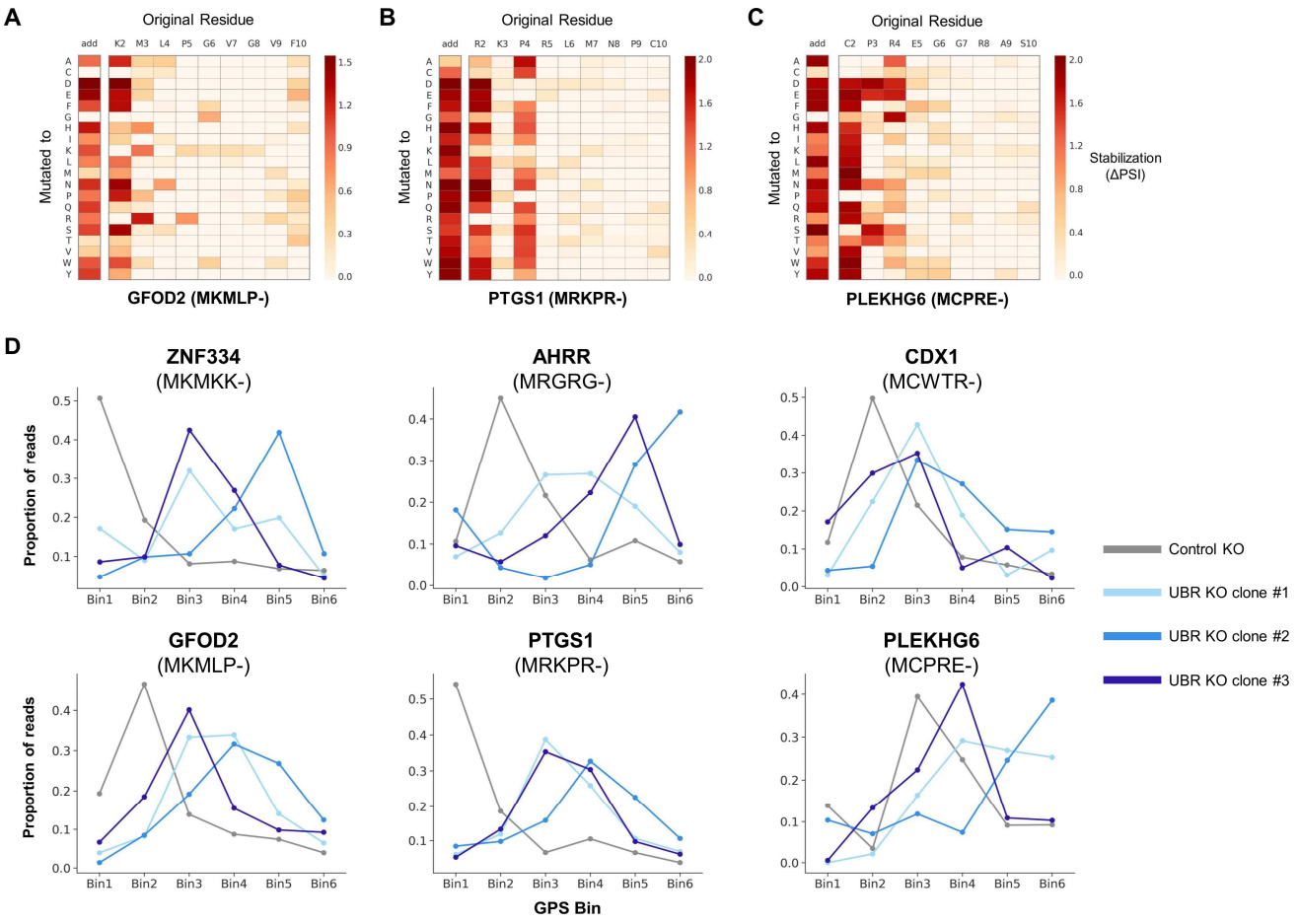

Fig. S5. Characterizing N-terminal UBR degrons through synthetic mutagenesis.

(A-C) Saturation mutagenesis of N-terminal peptides derived from GFOD2 (commencing MK-) **(A)**, PTGS1 (commencing MR-) **(B)** and PLEKHG6 (commencing MC-) **(C)**. A Ub-peptide-GPS library was generated in which each residue from the second position to the tenth position of the peptide was mutated to all other possible amino acids; the library was then expressed in HEK-293T cells and the stability of each mutant was measured by FACS followed by Illumina sequencing. The color scale reflects the degree of stabilization (measured in PSI units) for each mutant peptide-GFP fusion compared to the median stability of all the wild-type peptides. In addition, in order to assess the requirement for the degron to be positioned at the extreme N-terminus, mutants were also created in which each amino acid was added to the front of the peptide (that is, immediately after the initiator methionine).

(D) GPS screen profiles for each of the six UBR peptide-GFP substrates selected for saturation mutagenesis, showing the distribution of Illumina sequencing reads across the six bins in control cells (gray line) versus the three UBR mutant clone (blue lines).

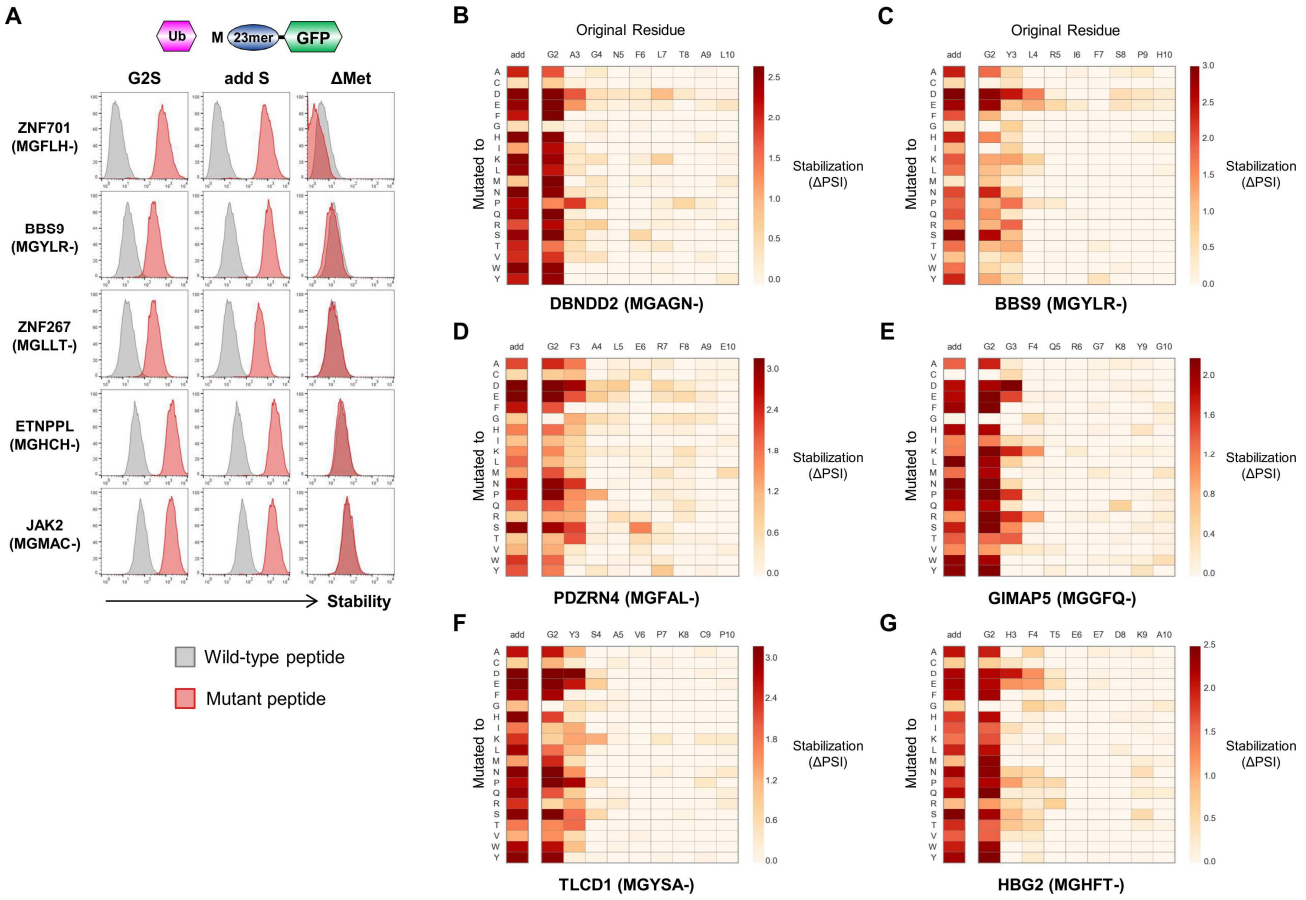

Fig. S6. Synthetic mutagenesis defines the composition of N-terminal glycine degrons.

(A) N-terminal glycine can serve as a degron. N-terminal peptides derived from the indicated genes were expressed in the context of the Ub-GPS system in HEK-293T cells. In each case, the wild-type peptide and a mutant version lacking the initiator methionine (ΔMet) were highly unstable, whereas mutant versions in which the terminal glycine was mutated to serine (G2S) or in which a serine residue was added between the initiator methionine and the glycine residue (add S) were not.

(B-G) Saturation mutagenesis of N-terminal peptides derived from the **(B)** *DBNDD2*, **(C)** *BBS9*, **(D)** *PDZRN4*, **(E)** *GIMAP5*, **(F)** *TLCD1*, and **(G)** *HBG2* genes. A Ub-peptide-GPS library was generated in which each residue from the second position to the tenth position of the peptide was mutated to all other possible amino acids; the library was then expressed in HEK-293T cells and the stability of each mutant was measured by FACS followed by Illumina sequencing. The color scale reflects the degree of stabilization (measured in PSI units) for each mutant peptide-GFP fusion compared to the median stability of all the wild-type peptides. In addition, in order to assess the requirement for the degron to be positioned at the extreme N-terminus, mutants were also created in which each amino acid was added at the front of the peptide, between the initiator methionine and the following glycine residue.

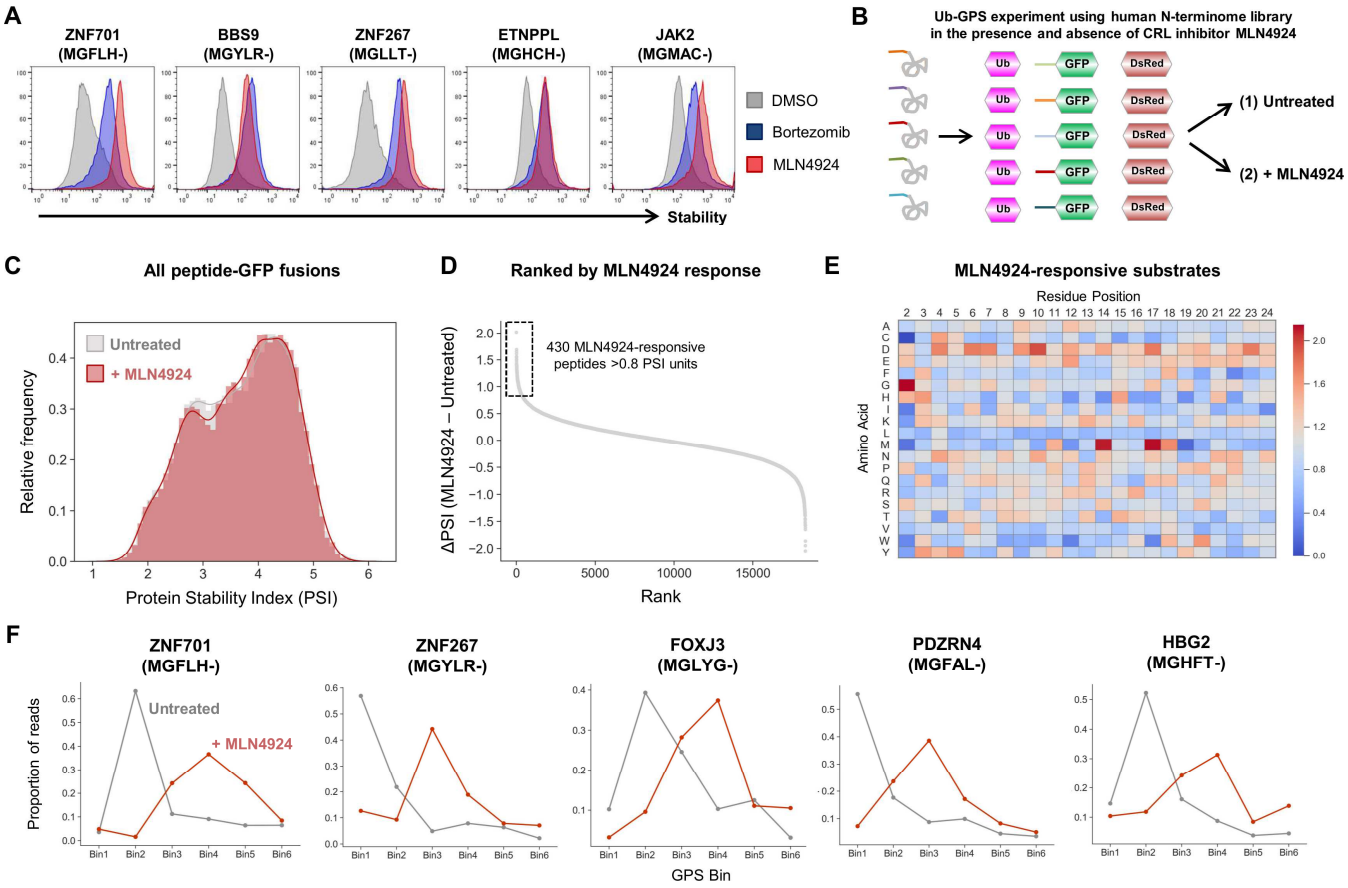

Fig. S7. Global identification of Cullin-RING E3 ligase substrates across the human N-terminome.

(A) Cullin-RING ligases target N-terminal glycine. N-terminal 24-mer peptides from the indicated genes were expressed in HEK-293T cells in the context of the Ub-GPS system. The stability of the peptide-GFP fusion proteins was measured by flow cytometry following treatment with the proteasome inhibitor Bortezomib or the pan-CRL inhibitor MLN4924.

(B) Schematic representation of a GPS screen to identify CRL substrates across the human N-terminome. The Ub-GPS N-terminome library was expressed in HEK-293T cells, which were then partitioned into six bins by FACS based on the GFP/DsRed ratio either with or without prior treatment with MLN4924.

(C and D) Effect of MLN4924 on the stability of the Ub-GPS N-terminome library. Distribution of protein stability scores for all peptide-GFP fusions with an initiator methionine in the presence (red) or absence (gray) of MLN4924 **(C)**; a total of 430 peptide-GFP fusions were stabilized > 0.8 PSI units upon drug treatment **(D)**.

(E) Sequence composition of N-terminal CRL substrates. Heatmap shows the relative enrichment (red) or depletion (blue) of each amino acid across all positions of the 23-mer peptide, comparing the 430 CRL substrates from (D) relative to the whole N-terminome library; glycine at the N-terminus is the most enriched feature. Occasionally relatively rare amino acids such as methionine appear to give a strong enrichment signal in certain positions, but these are likely insignificant statistical fluctuations.

(F) GPS profiles of example peptide-GFP CRL substrates encoding glycine at the second position, both in the presence (red) and absence (gray) of MLN4924.

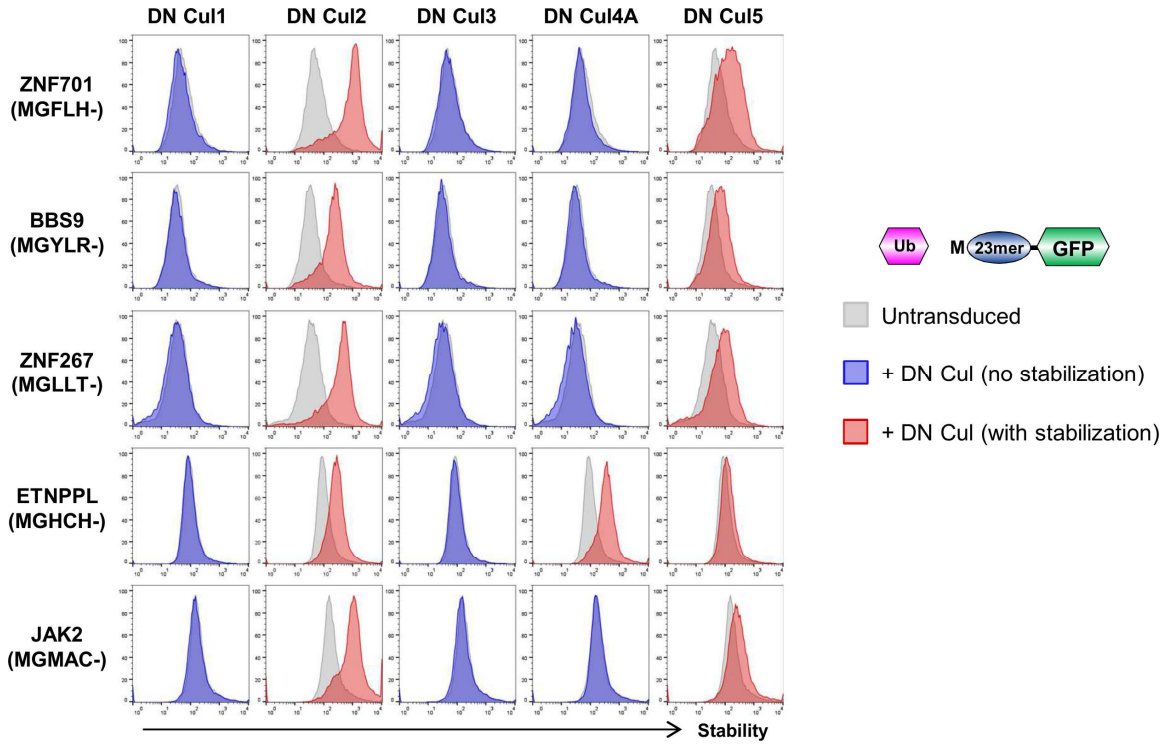

Fig. S8. CRL2 complexes target N-terminal glycine.

N-terminal 24-mer peptides derived from the indicated genes were expressed in HEK-293T cells in the context of the Ub-GPS system. The cells were then transduced with lentiviral vectors expressing dominant-negative (DN) version of the indicated Cullins and stabilization of the peptide-GFP fusions was assessed by flow cytometry. Except for the peptide derived from the ETNPPL gene, which was also stabilized upon expression of dominant-negative Cul4A, expression of dominant-negative Cul2 resulted in the greatest degree of stabilization in each case.

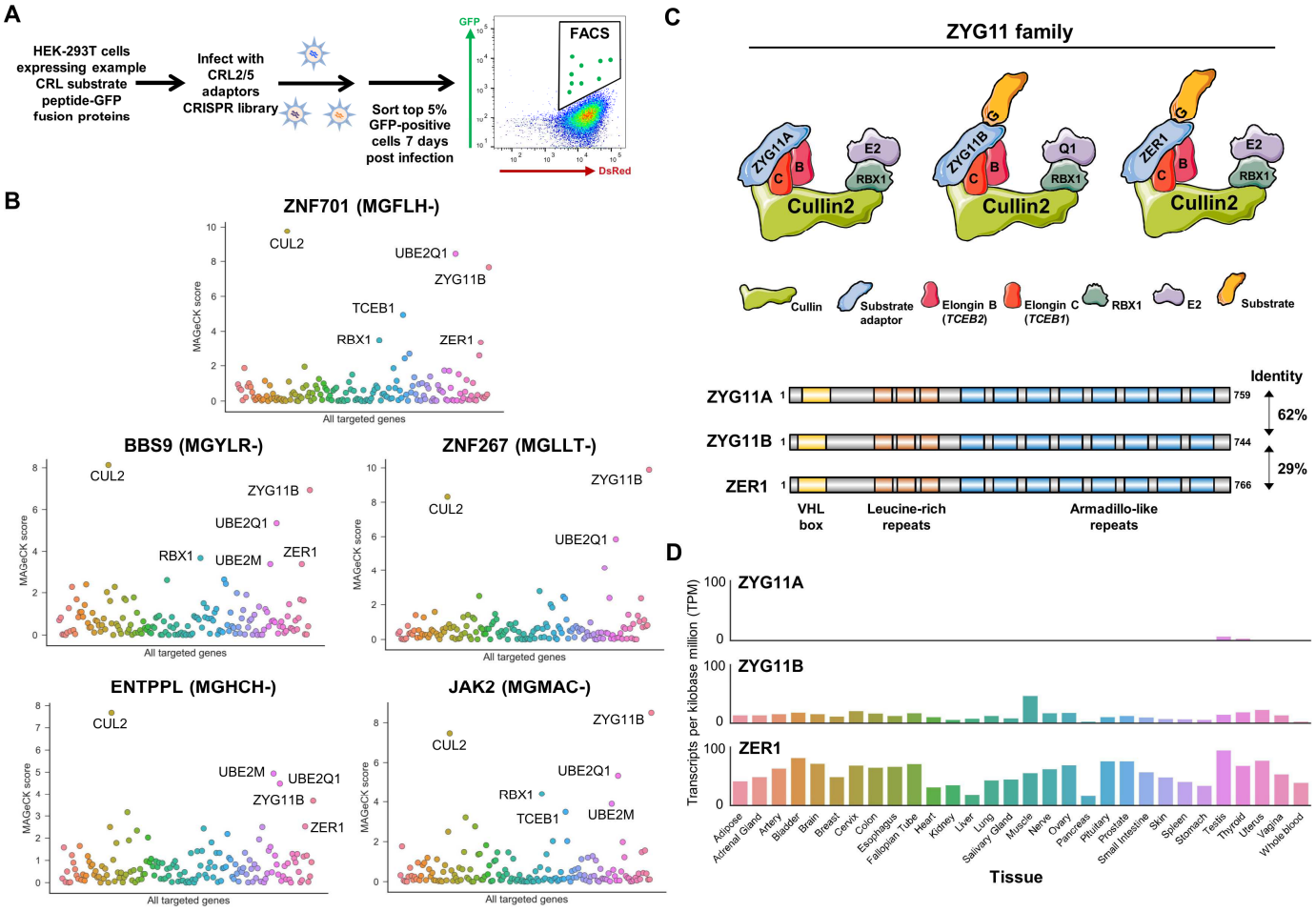

Fig. S9. The ZYG11 family of Cul2 adaptors target N-terminal glycine.

(A) Schematic representation of the CRISPR screens. HEK-293T cells expressing unstable peptide-GFP fusions were transduced with Cas9 and a library of sgRNAs targeting known Cul2 and Cul5 adaptors plus sgRNAs targeting core CRL components and E2 enzymes. One week later the top 5% of cells based on GFP expression were isolated by FACS, and the sgRNAs enriched in this population were identified by Illumina sequencing.

(B) CRISPR screen results, as determined using the MAGeCK algorithm to identify genes targeted by multiple sgRNAs enriched in the selected cells versus the unselected library (*46*).

(C) Schematic representation of the Cul2 complexes formed using ZYG11 family substrate adaptors.

(D) Expression of ZYG11 family substrate adaptors across human tissues. RNA-seq data from the GTEx project (*20*) shows that while ZYG11B and ZER1 are broadly expressed across all human tissues, ZYG11A shows only low expression in the testes and thymus.

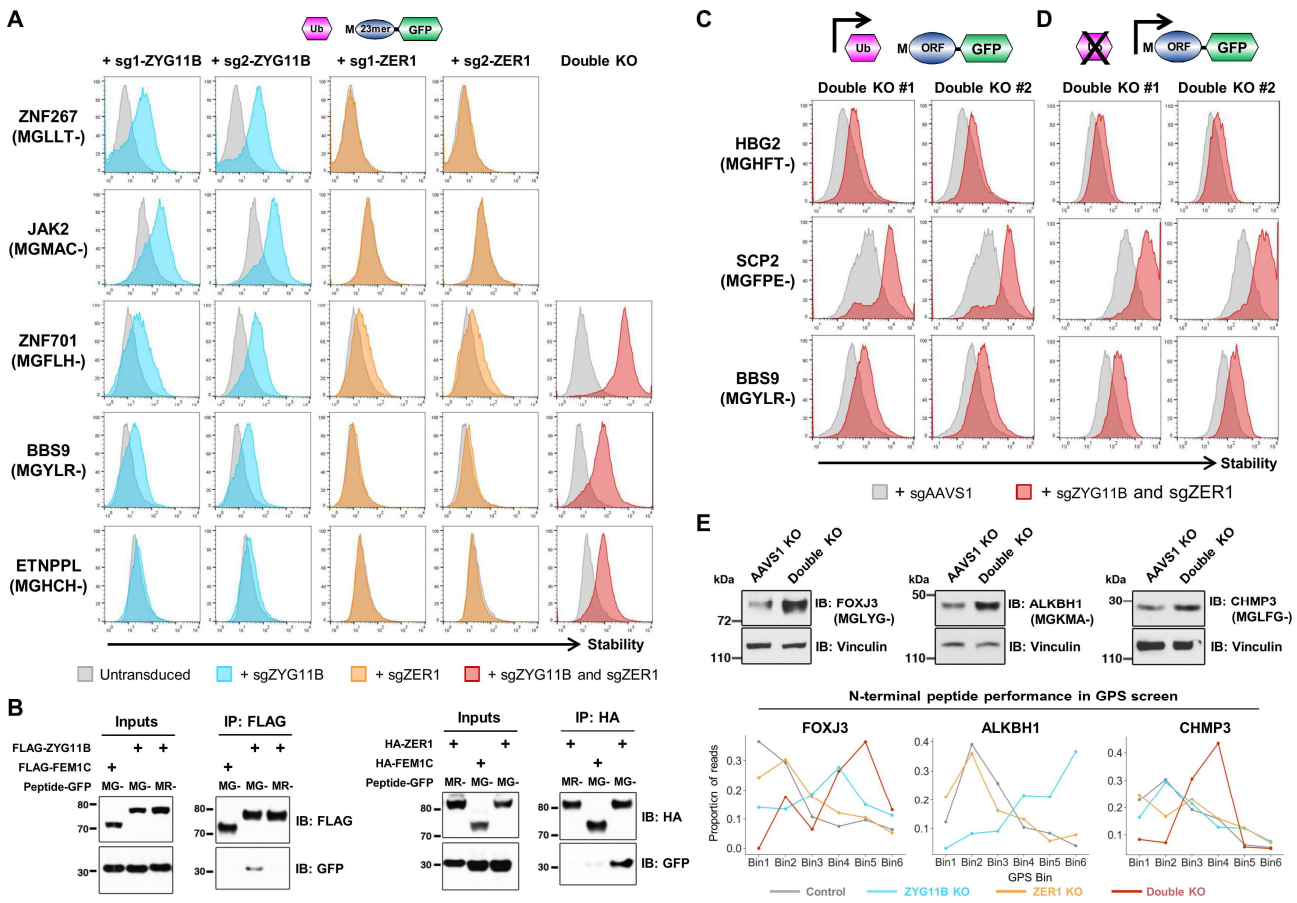

Fig. S10. Cul2^{ZYG11B} and Cul2^{ZER1} redundantly target N-terminal glycine degrons.

(A) Ablation of ZYG11B and ZER1 stabilizes peptide-GFP fusion proteins bearing N-terminal glycine degrons. N-terminal 24-mer peptides derived from the indicated genes were stably expressed in the context of the Ub-GPS system in HEK-2923T cells. For the peptides derived from ZNF267 and JAK2, disruption of ZYG11B alone resulted in full stabilization (blue histograms), whereas for the other peptide-GFP fusions combined mutation of both substrate adaptors was required for full stabilization (red histograms).

(B) ZYG11B and ZER1 associate with reporter substrates bearing N-terminal glycine degrons. HEK-293T cells were generated stably expressing epitope-tagged ZYG11B or ZER1 and peptide-GFP reporters (commencing MG-) derived from the N-terminus of SNX11 (left) or ZNF267 (right). Following immunoprecipitation of the substrate adaptor, co-immunoprecipitation of the reporter substrate was assessed by immunoblot for GFP. An irrelevant Cul2 adaptor (FEM1C) and an irrelevant peptide-GFP reporter derived from the N-terminus of RAD51C (commencing MR-) served as negative controls.

(C and D) Cul2^{ZYG11B} and Cul2^{ZER1} target full-length proteins for degradation. Full-length open reading frames (ORFs) corresponding to the indicated genes were expressed as N-terminal fusions to GFP either in the context of the ubiquitin fusion system (C) or without upstream ubiquitin (D) and expressed in wild-type (gray histograms) or double mutant cells lacking ZYG11B and ZER1 expression (red histograms).

(E) Endogenous substrates of Cul2^{ZYG11B} and Cul2^{ZER1}. The abundance of the indicated proteins in AAVS1 knockout (control) or combined ZYG11B/ZER1 double mutant cells was assessed by immunoblot (upper panels). The GPS screen performance of the corresponding N-terminal peptides (see fig. S10) is shown in the bottom panels; insufficient sequencing reads prevented quantification of the stability of the N-terminal peptide from ALKBH1 in the double mutant cells.

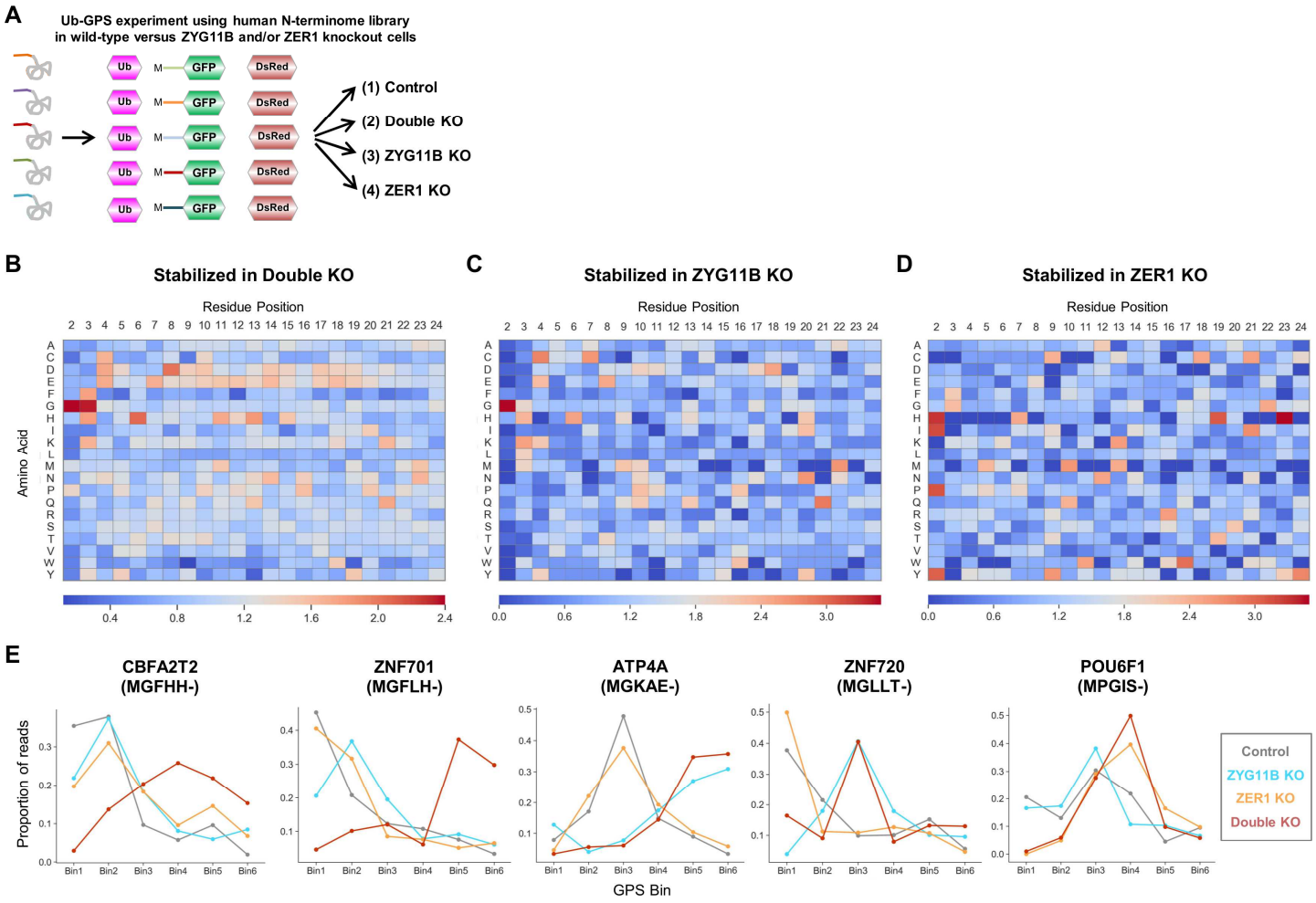

Fig. S11. Global identification of Cul2^{ZYG11B} and Cul2^{ZER1} substrates across the human N-terminome.

(A) Schematic representation of the GPS experiment, in which the stability of the Ub-GPS N-terminome library with initiator methionine was compared in wild-type HEK-293T cells versus ZYG11B mutant, ZER1 mutant and combined ZYG11B and ZER1 mutant cells.

(B-D) Sequence composition of ZYG11B and ZER1 substrates. Heatmaps show the relative enrichment (red) or depletion (blue) of each amino acid across all positions of the 23-mer peptide, comparing the composition of substrates stabilized in double ZYG11B and ZER1 mutant cells **(B)**, ZYG11B mutant cells **(C)**, or ZER1 mutant cells **(D)** relative to the whole N-terminome library.

(E) GPS profiles of example substrates, showing the read distribution across the bins in control cells (gray) versus ZYG11B mutant cells (blue), ZER1 mutant cells (orange) or dual ZYG11B/ZER1 double mutant cells (red). These examples illustrate substrates redundantly targeted by both ZYG11B and ZER1 (N-terminal peptides derived from CBFA2T2 and ZNF701), substrates targeted solely by ZYG11B (N-terminal peptides derived from ATP4A and ZNF720), and substrates targeted solely by ZER1 (N-terminal peptide derived from POU6F1).

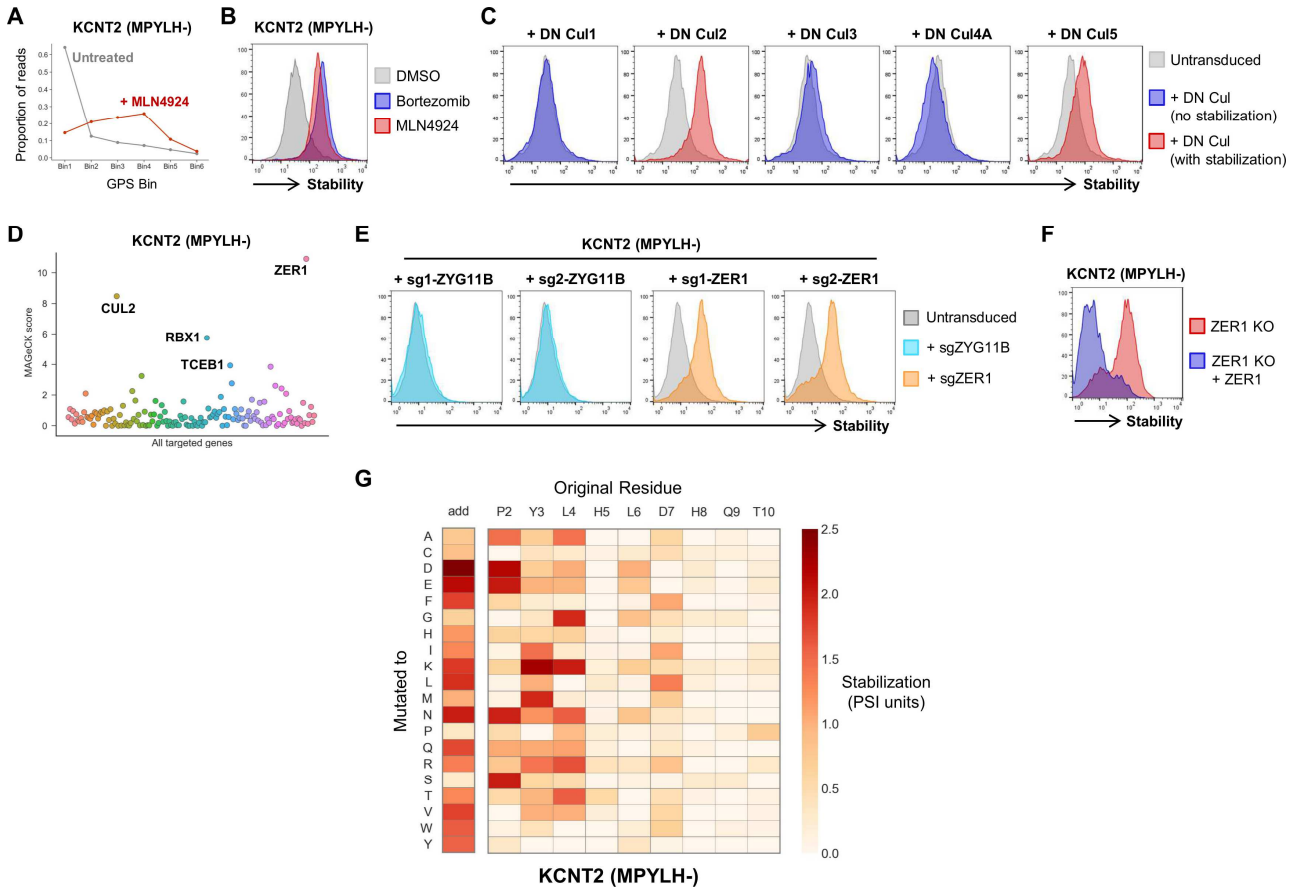

Fig. S12. An N-terminal glycine residue is not absolutely critical for substrate recognition by Cul2^{ZER1}.

(A-C) Characterization of an N-terminal degron targeted by Cullin-RING E3 ligases commencing with proline. **(A)** The N-terminal peptide derived from *KCNT2*, commencing MPYLH-, was identified as a CRL substrate in the GPS screen with MLN4924 (fig. S6). We validated that this peptide fused to GFP was indeed stabilized by MLN4924 **(B)**, and, employing dominant-negative Cullin constructs, found that it was a substrate for CRL2 **(C)**.

(D) CRISPR screen identifies ZER1 as the CRL2 adaptor responsible for the degradation of the KCNT2 peptide-GFP fusion. HEK-293T cells stably expressing the KCNT2 peptide-GFP fusion in the context of the Ub-GPS system were transduced with Cas9 and a CRISPR sgRNA library targeting Cul2/5 substrate adaptors, and cells in which the GFP fusion were stabilized were isolated by FACS. Guides significantly enriched in the sorted cells compared to the unselected library were identified using the MAGeCK algorithm.

(E) Individual CRISPR-mediated gene disruption experiments validating ZER1-mediated degradation of the KCNT2 peptide-GFP fusion.

(F) Rescue of reporter degradation in ZER mutant cells upon exogenous expression of ZER1.

(G) Saturation mutagenesis of the *KCNT2* peptide identifies the critical residues for ZER1-mediated recognition. This revealed that the hydrophobic residues encoded at the third and fourth position were particularly important for ZER1 binding, with some more flexibility allowed at the second position. However, the positioning of the degron motif relative to the N-terminus of the protein remaining critical, as the addition of almost any single amino acid downstream of the initiator methionine resulted in stabilization.

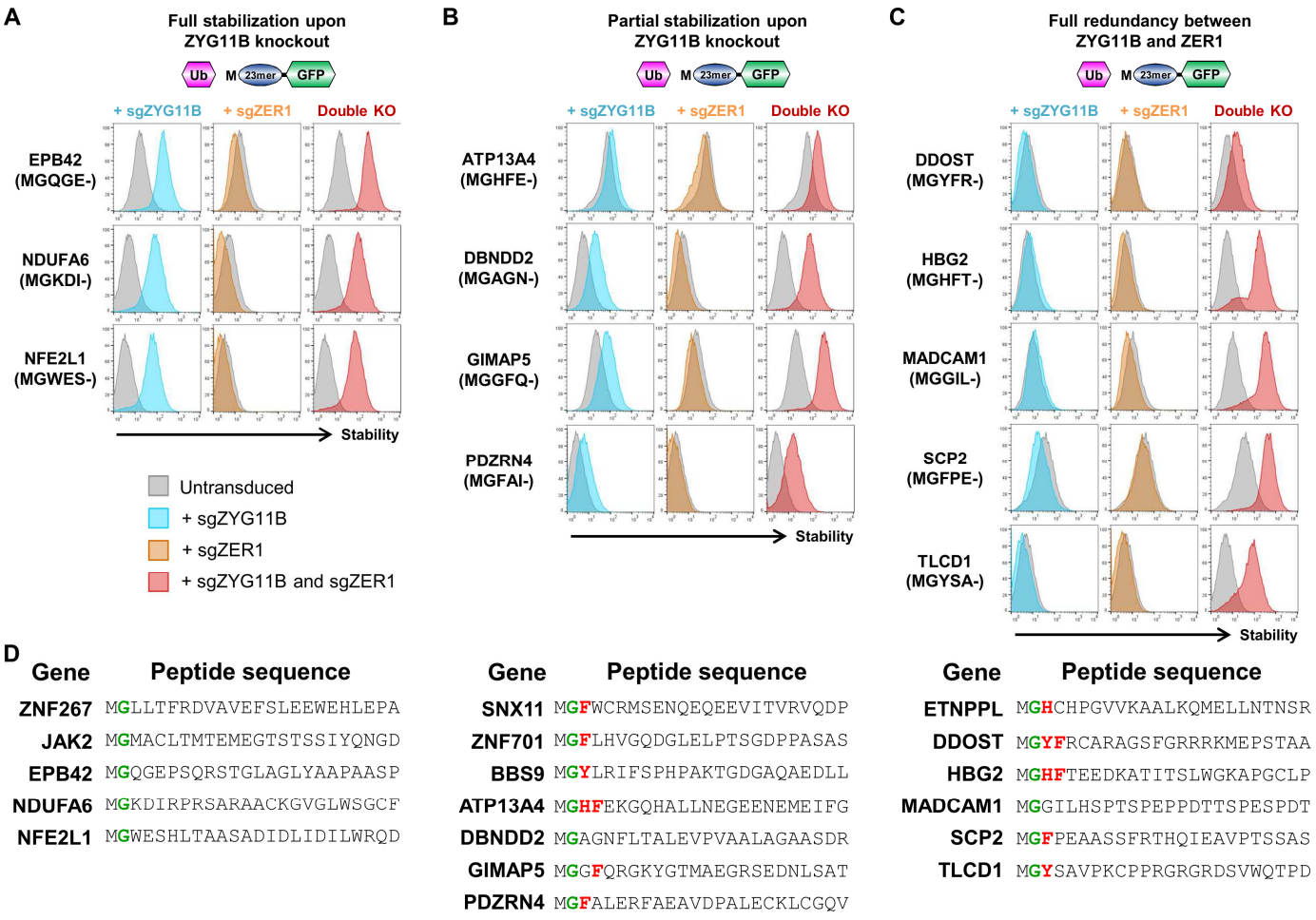

Fig. S13. Exploring the degron requirements for Cul2^{ZYG11B} versus Cul2^{ZER1}.

(A-D) Individual CRISPR/Cas9-mediated gene disruption experiments to identify additional substrates of Cul2^{ZYG11B} and Cul2^{ZER1} bearing N-terminal glycine. In each case, N-terminal peptides derived from the indicated genes were expressed in the context of the Ub-GPS system in either wild-type, ZYG11B mutant, ZER1 mutant or dual ZYG11B/ZER1 mutant HEK-293T cells and their stability assessed by flow cytometry. Substrates could be grouped into three categories: **(A)** peptide-GFP fusions fully stabilized upon ablation of ZYG11B, for which no additional stabilization was observed in the double mutant cells, **(B)** peptide-GFP fusions which exhibited some degree of stabilization in ZYG11B single mutant cells, but for which full stabilization was only observed in the double mutant cells, and **(C)** peptide-GFP fusions which displayed no stabilization at all in either ZYG11B or ZER1 mutant cells but were stabilized in the double mutant cells. For the majority of the substrates in the latter two categories, an aromatic residue (histidine, phenylalanine or tyrosine) was encoded at the third and/or fourth position in the peptide **(D)**, suggesting that this feature may be an important part of the degron recognized specifically by ZER1.

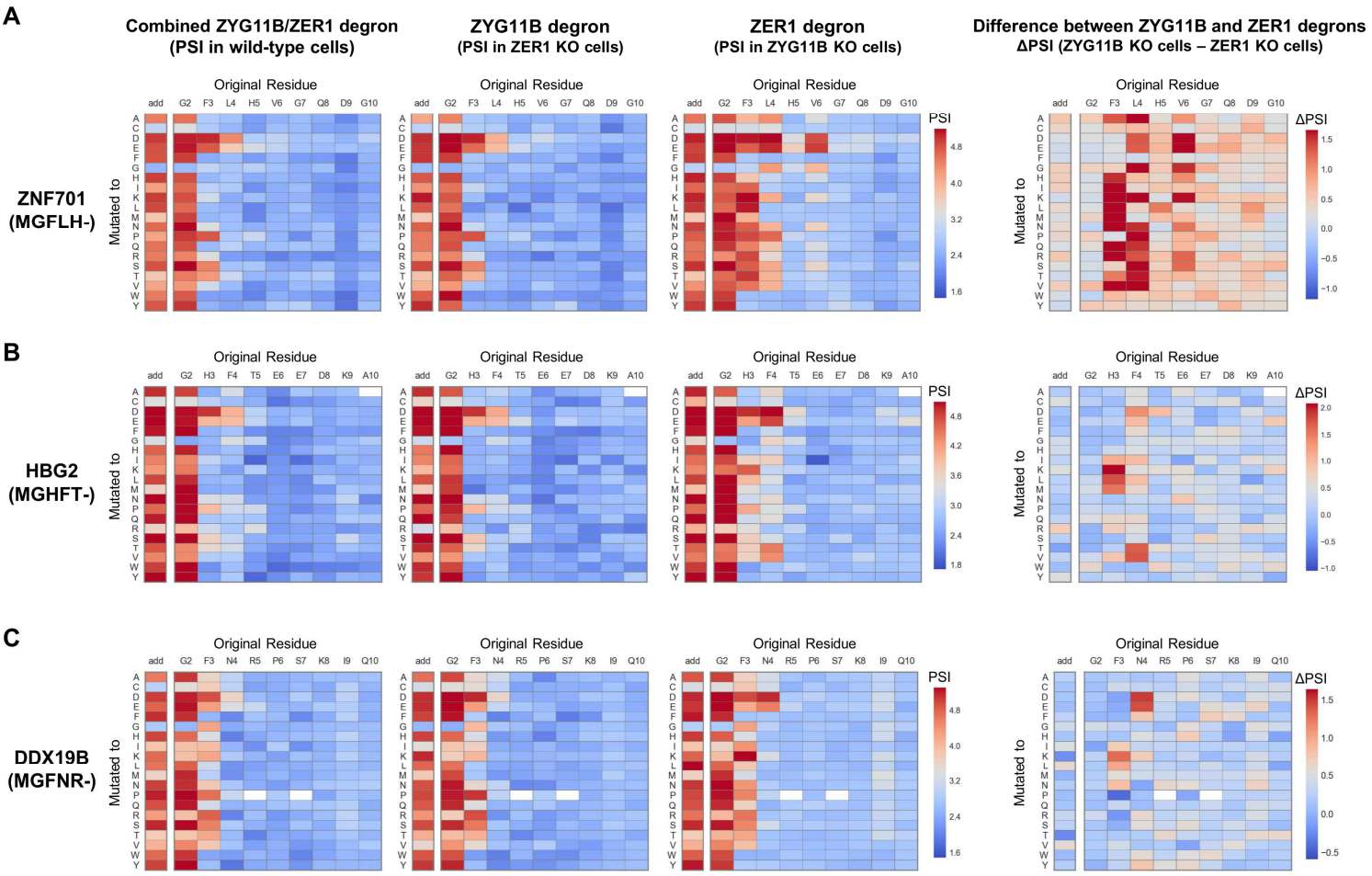

Fig. S14. Saturation mutagenesis defines the degrons recognized by Cul2^{ZYG11B} and Cul2^{ZER1}.

(A-C) Saturation mutagenesis of the N-terminal peptides derived from the **(A)** *ZNF701*, **(B)** *HBG2* and **(C)** *DDX19B* genes in different genetic backgrounds. A Ub-peptide-GPS library was generated in which each of the first ten residues of the peptides were mutated to all possible amino acids; the library was then expressed in either wild-type, ZYG11B mutant or ZER1 mutant HEK-293T cells and the stability of each mutant was measured by FACS followed by Illumina sequencing. For the left three columns, the color scale reflects the raw PSI measurement for each peptide-GFP fusion and thus the greater the intensity of the red color, the greater the stabilizing effect of the mutation. For the rightmost column, the color scale reflects the difference between the PSI in ZYG11B mutant cells versus ZER1 mutant cells and thus a dark red color indicates mutations which prevent recognition by ZER1 but not by ZYG11B, while a dark blue color indicates mutations which permit recognition by ZER1 but not by ZYG11B. In all plots, the column labeled 'add' represents the addition of an amino acid between the initiator methionine and the glycine residue encoded at the second position.

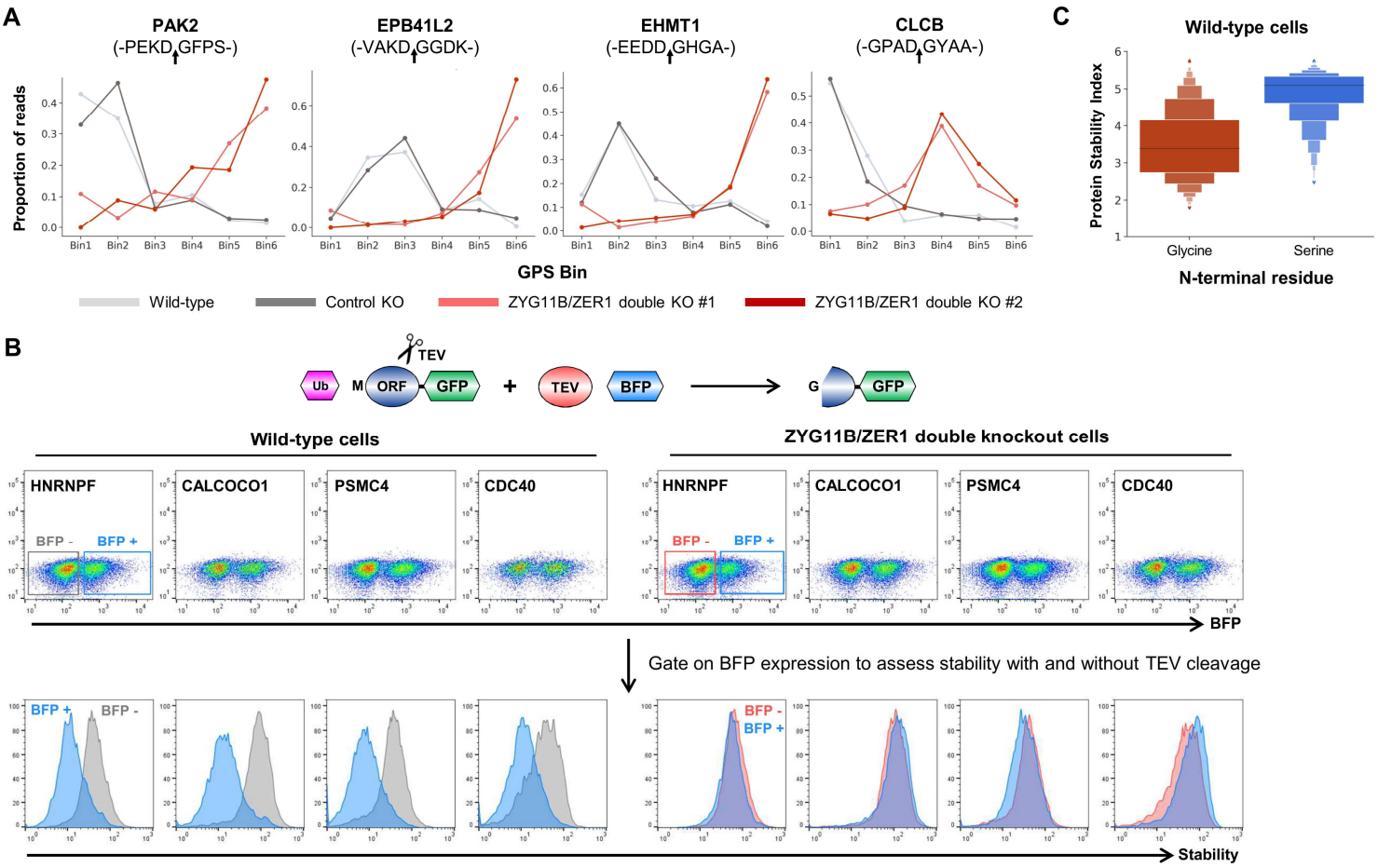

Fig. S15. Caspase cleavage generates N-terminal glycine degrons targeted by ZYG11B and ZER1.

(A) Caspase cleavage sites represent a source of N-terminal glycine degrons. Profiles for four example substrates from the Ub-GPS screen with the caspase cleavage product library are shown, illustrating the distribution of Illumina sequencing reads across the six bins. The amino acids around the caspase site are indicated, with the position of cleavage indicated by an arrow.

(B) Destabilization via N-terminal glycine degrons following proteolytic cleavage. The full-length ORFs encoding the indicated proteins were modified such that their annotated caspase cleavage site was replace with a TEV protease cleavage site. Upon TEV expression, as monitored by the co-expression of BFP (top row), the stability of the downstream cleavage product decreased in wild-type cells (bottom row, left), but not in combined ZYG11B/ZER1 mutant cells (bottom row, right).

(C) N-terminal glycine and serine have contrasting effects on protein stability. The letter-value boxplots depict the distribution of stability scores for all peptide-GFP fusions in the caspase cleavage site library commencing with either an N-terminal glycine (red) or serine (blue) residue.

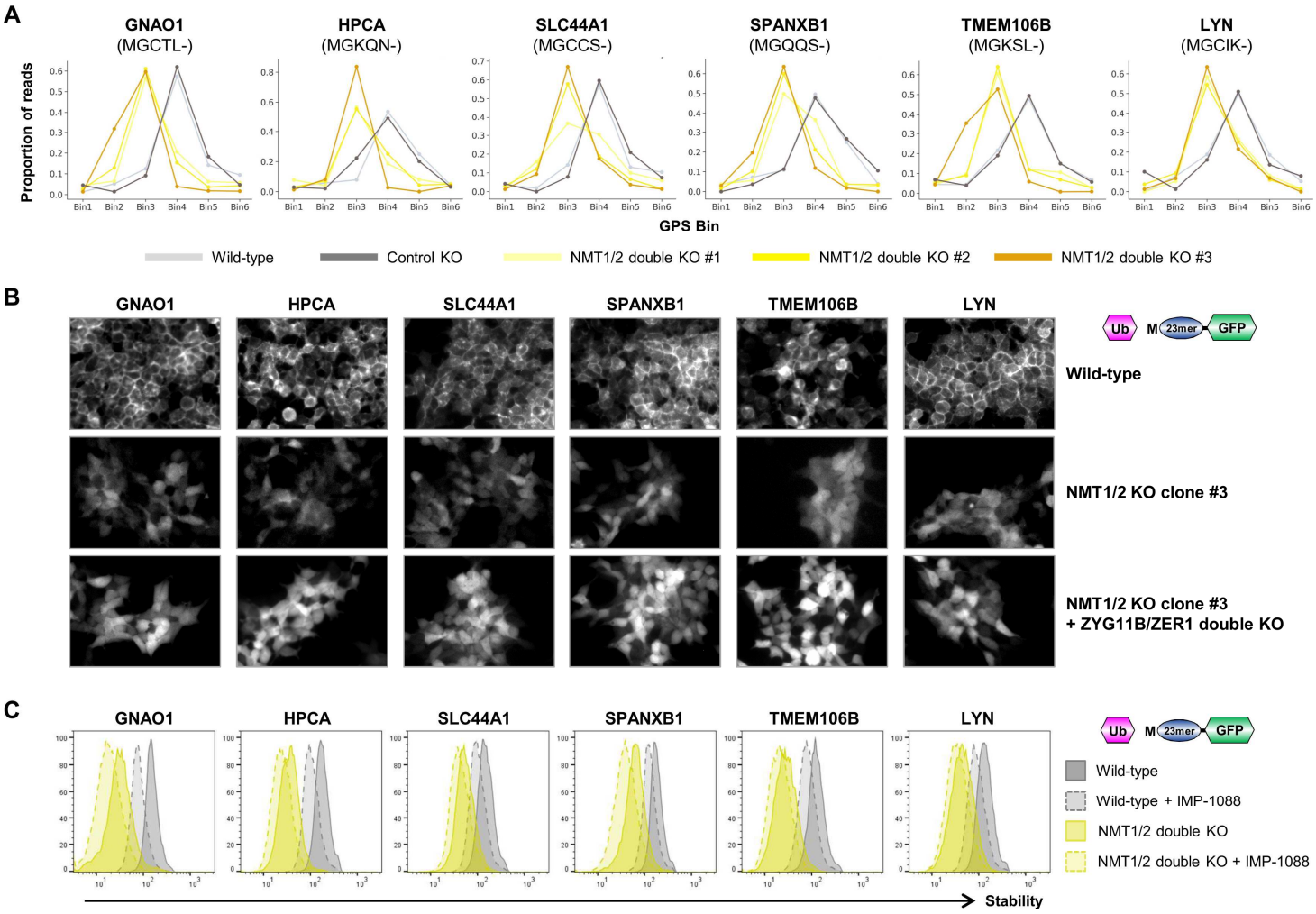

Fig. S16. A failure of N-myristoylation exposes N-terminal glycine degrons targeted by ZYG11B and ZER1.

(A) GPS screen profiles for the six peptides selected for individual validation; concordant destabilization was observed in all three NMT1/2 mutant clones.

(B) Ablation of NMT1/2 impairs the membrane localization of example peptide-GFP substrates. The first 24 amino acids from the indicated proteins were expressed as N-terminal fusions to GFP in HEK-293T cells and their localization assessed by fluorescence microscopy. The membrane localization apparent in wild-type cells (top row) is abolished in NMT1/2 mutant clone #3 (middle row), and the abundance of the peptide-GFP fusions in the NMT1/2 mutant clones is increased upon mutation of ZYG11B and ZER1.

(C) Small molecule NMT inhibition further destabilizes example peptide-GFP substrates. Treatment with the NMT1/2 inhibitor IMP-1088 (dotted lines) decreases the stability of the peptide-GFP substrates, both in wild-type cells (gray histograms) and NMT1/2 mutant clone #3 (gold histograms).

Additional Data table S1 (separate file)
N-terminome GPS screen data in wild-type cells.

Additional Data table S2 (separate file)
Computational prediction of N-terminal degrons.

Additional Data table S3 (separate file)
N-terminome GPS screen data in different genetic backgrounds.

Additional Data table S4 (separate file)
Saturation mutagenesis data.

Additional Data table S5 (separate file)
Identifying N-terminal peptides targeted by Cullin-RING E3 ligases.

Additional Data table S6 (separate file)
CRISPR screen data.

Additional Data table S7 (separate file)
Caspase cleavage product GPS screen data.