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

Richard T. Timms[1,2], Zhiqian Zhang[1,2], David Y. Rhee[3], J. Wade Harper[3], Itay Koren[1,2]*†, Stephen J. Elledge[1,2]*

[1]Division of Genetics, Department of Medicine, Howard Hughes Medical Institute, Brigham and Women's Hospital, Boston, MA 02115, USA
[2]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA
[3]Department of Cell Biology, Harvard Medical School, Boston, MA 02115, USA
†Present address: The Mina and Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat-Gan, 5290002, Israel

*Correspondence: itay.koren@biu.ac.il, selledge@genetics.med.harvard.edu

Abstract:

The N-terminal residue influences protein stability through N-degron pathways. Here, through stability profiling of the human N-terminome, we uncover multiple additional features of N-degron pathways. In addition to uncovering extended specificities of UBR E3 ligases, we characterized two related Cullin-RING E3 ligase complexes, $Cul2^{ZYG11B}$ and $Cul2^{ZER1}$, that act redundantly to target N-terminal glycine. N-terminal glycine degrons are depleted at native N-termini but strongly enriched at caspase cleavage sites, suggesting roles for the substrate adaptors ZYG11B and ZER1 in protein degradation during apoptosis. Furthermore, ZYG11B and ZER1 participate in the quality control of *N*-myristoylated proteins, wherein N-terminal glycine degrons are conditionally exposed following a failure of *N*-myristoylation. Thus, an additional N-degron pathway specific for glycine regulates the stability of metazoan proteomes.

One Sentence Summary:

Stability profiling of the human N-terminome uncovers an N-degron pathway through which Cullin-RING E3 ligases target N-terminal glycine.

Main text:

The ubiquitin-proteasome system (UPS) is the major route through which eukaryotic cells achieve selective protein degradation (*1*). The specificity of this system is provided by E3 ubiquitin ligases, of which more than 600 are encoded in the human genome. E3 ligases recognize specific sequence elements, known as degrons, that are present in substrate proteins (*2*). However, whilst a detailed knowledge of the specificity of E3 ligases for degrons will be essential for achieving a systems-level understanding of the UPS, our current knowledge of degron motifs remains remarkably sparse (*3*).

The first degrons to be discovered were located at the N-terminus of proteins (*4*). N-terminal degrons are targeted by N-degron pathways (formerly known as N-end rule pathways (*5*)), of which there are two main branches: the Arg/N-degron pathway, through which UBR-family E3 ligases target N-termini typically generated through endoproteolytic cleavage (*6, 7*), and the Ac/N-degron pathway, through which proteins bearing acetylated N-termini are targeted for degradation by the E3 ligase MARCH6 (also known as TEB4) (*8*, *9*). In addition, a Pro/N-degron pathway was recently described, through which proteins harboring an N-terminal proline residue are degraded by the GID E3 ligase complex (*10*) **(fig. S1)**. Theoretically these pathways have the capacity to target the majority of cellular proteins, but the extent to which they impact protein stability in a physiological context remains unclear. For example, loss of N-terminal acetyltransferase (NAT) enzymes has minimal effects on protein stability in yeast (*11*), which is inconsistent with a widespread role for the Ac/N-degron pathway.

Previously we modified the Global Protein Stability (GPS) system (*12*) to develop a high-throughput method to characterize degron motifs in human proteins (*13*). This approach is based on a lentiviral expression vector encoding two fluorescent proteins: DsRed, which

serves as an internal reference, and GFP fused to a short peptide of interest, which is translated from an internal ribosome entry site (IRES). Because both DsRed and the GFP-peptide fusion protein are expressed from the same transcript, the GFP/DsRed ratio can be used to quantify the effect of the peptide sequence on the stability of GFP (*13*). Here, we exploit the ubiquitin-fusion technique (*4*) to adapt this "GPS-peptidome" approach to search for N-terminal degron motifs. We thereby directly examined the contribution of N-terminal sequences to protein stability in human cells.

Stability profiling of the human N-terminome using GPS-peptidome technology

We synthesized an oligonucleotide library encoding the first 24 amino acids of the primary isoform(s) of all human proteins, both with and without an initiator methionine (~50,000 sequences). These were cloned into the "Ub-GPS" expression vector between the ubiquitin gene and GFP **(Fig. 1A)**. Upon expression of the constructs in HEK-293T cells, proteolytic cleavage of the ubiquitin moiety by endogenous deubiquitinating enzymes led to the exposure of the peptides at the N-terminus of GFP **(Fig. 1A)**. Fluorescence-activated cell sorting (FACS) was used to partition the population into six bins of equal size based on the stability of the peptide-GFP fusion. The stability of each fusion was then quantified by Illumina sequencing, with each peptide assigned a protein stability index (PSI) score ranging between 1 (maximally unstable) and 6 (maximally stable) based on the proportion of sequencing reads in each bin **(data S1)**.

We began by assessing the effect of the initiator methionine on protein stability. Overall, peptide-GFP fusions lacking an initiator methionine were much less stable than their counterparts with an initiator methionine **(Fig. 1B)**. However, this effect was only observed for certain N-terminal residues **(Fig. 1C)**. Reporters commencing with amino acids bearing

small side chains (C, V, G, P, T, A and S) were generally relatively stable, and exhibited little or no difference in overall stability whether or not they were preceded by an upstream methionine residue. This is consistent with efficient cleavage of the initiator methionine by methionine aminopeptidases when the following amino acid has a sufficiently small radius of gyration (*14*). In contrast, peptide-GFP fusions commencing with all other residues (except methionine itself) were generally stable only when preceded by an upstream methionine residue, and were greatly destabilized in the absence of an initiator methionine **(Fig. 1C-F)**.

Overall, these data provide strong support for a central role of the Arg/N-degron pathway in protein quality control. Whereas proteins bearing native N-termini (methionine itself, or C/V/G/P/T/A/S, from which methionine is normally removed (*14*)) are broadly stable, proteins bearing aberrant N-termini (R/K/H/W/Y/F/L/I/D/E/N/Q, without a preceding methionine) are all highly unstable. The latter residues correspond perfectly to the primary type I (R/K/H), primary type II (W/Y/F/L/I), secondary (D/E) and tertiary (N/Q) N-terminal degrons of the Arg/N-degron pathway **(fig. S1A)**. Crucially, however, when these residues were preceded by methionine – as they would be in the context of normal protein synthesis – broad stabilization was observed **(Fig. 1F)**.

Computational identification of destabilizing N-terminal motifs

Subsequently we focused on understanding the factors that determined the stability of peptide-GFP fusions synthesized with an initiator methionine. Stability scores for these fusions were distributed bimodally, with approximately one-third of the library exhibiting significant instability **(Fig. 1B, blue histogram)**. One key factor that strongly influenced stability was amino acid composition **(Fig. 2, A and B)**. For example, aspartic acid and glutamic acid were depleted from unstable peptides and enriched among the stable peptides,

while hydrophobic residues such as tryptophan, phenylalanine and leucine showed the opposite pattern. This effect is not specific to the N-terminus, however, as we find that similar rules govern the stability of reporter constructs in which peptides are fused at the C-terminus of GFP (*13*).

Most amino acids exerted a similar effect on stability regardless of their position across the 24-mer peptide, but we noticed that certain residues exerted differing effects specifically when encoded at the second position **(Fig. 2, A and B)**. We therefore performed a computational analysis to identify motifs that might promote instability specifically when located at or near the N-terminus of the peptide. For all possible combinations of di-peptide motifs, we compared the mean stability of all peptide-GFP fusions harboring the motif within the first seven N-terminal amino acids versus those harboring the motif at an internal position in the 24-mer peptide **(Fig. 2C and data S2)**. Over 80% of the top 100 candidate destabilizing N-terminal motifs could be grouped into four categories based solely on the identity of the second residue: lysine was present downstream of the initiator methionine in 26 motifs, arginine in 24 motifs, glycine in 22 motifs and cysteine in 9 motifs **(Fig. 2D)**. Reporters encoding these residues at the second position were significantly less stable than reporters containing these residues at any internal position **(Fig. 2E)**, and, globally, peptide-GFP fusions commencing MC-, MR-, MG- and MK- exhibited the lowest mean stability **(Fig. 2F)**. Thus, considering initiator methionine removal, this analysis identified N-terminal glycine and cysteine in addition to MR- and MK- as candidate destabilizing N-terminal motifs.

Exploring the substrate repertoire of UBR family E3 ligases

Next we sought to identify the cellular machinery targeting each class of putative N-terminal degron. We began by investigating a role for UBR family E3 ligases. UBR1, UBR2 and UBR4 have been shown functionally to participate in the recognition of N-degrons (*15*), and so, through sequential rounds of CRISPR/Cas9-mediated gene disruption, we attempted to create a single cell clone lacking all three of these UBR proteins. Despite screening ~40 clones, we were unable to identify a clone in which simultaneous ablation of UBR1, UBR2 and UBR4 proteins was observed by immunoblot, suggesting that such a triple mutant cell may not be viable. However, we were able to generate clones expressing substantially reduced levels of two or more of the proteins **(Fig. 3A)**. Ub-GPS reporters in which the initiator methionine of GFP was replaced with either arginine (R), lysine (K) or tyrosine (Y) were strongly destabilized in wild-type cells, but this effect was abrogated in UBR KO clone #1 and clone #3 **(fig. S2A)** and completely abolished in clone #2 **(Fig. 3B)**.

We created a panel of Ub-GPS constructs in which either 23-mer peptides **(Fig. 2C)** or 3-mer peptides **(fig. S2B)** harboring example degron motifs downstream of an initiator methionine were fused to the N-terminus of GFP. In both cases, loss of UBR proteins resulted in the stabilization of reporters bearing three of the classes of degrons motifs: MK-, MR- and N-terminal cysteine. However, loss of UBR proteins had little or no effect on the stability of the GFP-fusion proteins bearing N-terminal glycine, suggesting a role for additional E3 ligase(s) in the recognition of this particular N-terminal degron.

It was not surprising that UBR E3 ligases targeted N-terminal cysteine, given that nitric oxide-mediated oxidation and subsequent arginylation of N-terminal cysteine renders it a substrate for the Arg/N-degron pathway (*16*). That said, ATE1 disruption only led to modest stabilization of two peptide-GFP substrates exposing N-terminal cysteine **(fig. S2, C and D)**, suggesting that additional routes to UBR-mediated degradation must also exist. UBR-mediated degradation of proteins commencing MK- and MR- was unexpected,

however, suggesting that, in addition to targeting truncated proteins bearing abnormal N-termini, UBR ligases might also target certain intact proteins bearing their initiator methionine. To confirm that the initiator methionine of these substrates was indeed intact, and thus rule out the possibility that methionine removal was instead exposing canonical Arg/N-degrons, we used mass spectrometry to examine the N-terminus of two example peptide-GFP UBR substrates expressed in UBR KO clone #2 **(fig. S3A)**. In both cases we were readily able to detect the intact N-terminal peptide with the initiator methionine present, while we could not detect any peptides corresponding to a putative processed form without an initiator methionine **(fig. S3B)**.

To further examine this property of UBR proteins, we directly compared the stability of the entire Ub-GPS N-terminome library in wild-type cells versus UBR KO clones #1, #2 and #3 **(Fig. 3D and data S3A)**. Loss of UBR proteins had little effect on the overall stability of reporters synthesized with an N-terminal methionine: only 570 peptide-GFP fusion proteins (<3% of the N-terminome library) exhibited substantial stabilization (>0.8 PSI units) in any of the UBR mutant clones compared to control cells **(Fig. 3E)**. Sequence analysis of the UBR substrates revealed a clear preference for particular N-terminal degron motifs **(Fig. 3F and fig. S4A-H)**. Consistent with our previous data **(Fig. 2C and fig. S2B)**, peptides commencing MC-, MK- and MR- were all enriched. Peptides commencing ML- and MI- were also overrepresented, and for three example peptides in each case we validated that they were indeed stabilized in UBR KO clone #2 **(fig. S4I)**. In *S. cerevisiae* Ubr1 has been shown to target proteins commencing MΦ- (where Φ is a bulky hydrophobic residue, W/F/Y/L/I) for degradation (*17*); however, unlike peptides starting ML- and MI-, we did not observe enrichment for peptides starting MF- or MY- among the UBR substrates, and only weak enrichment for peptides starting MW-.

Finally, we noted that only a small proportion of all peptides in the library commencing MK-, MR-, ML- or MI- were UBR substrates, suggesting that additional residues were essential for degron recognition. Indeed, analysis of the composition of all the UBR substrates identified in each category highlighted preferred residues enriched at downstream positions **(fig. S4E-H)**. Furthermore, for some example peptides starting MK-, MR- and MC- we defined the N-terminal UBR degron in detail by performing saturation mutagenesis experiments. We created a Ub-GPS library in which each of the residues from position 2 to position 10 of the 24-mer peptide were mutated to all other possible amino acids, and measured the stability of the resulting peptide-GFP fusions by FACS and Illumina sequencing **(data S4A)**. These experiments confirmed the critical importance of the lysine, arginine or cysteine residue encoded at the second position, but also demonstrated that certain mutations at the third or fourth position could prevent degron recognition **(Fig. 3G-I and fig. S5)**. These data also confirmed the requirement for these degron motifs to be positioned at the extreme N-terminus, as addition of just a single upstream amino acid (that is, immediately after the initiator methionine) resulted in stabilization of the peptide-GFP fusions **(Fig. 3G-I and fig. S5, column labeled 'add')**.

N-terminal glycine can act as a potent degron

We next focused on the one class of N-terminal degron motif that was not a substrate for UBR-mediated degradation: N-terminal glycine. To validate that N-terminal glycine did indeed constitute a degron motif, we performed a series of mutagenesis experiments on a panel of unstable Ub-GPS reporters in which 24-mer peptides commencing MG- were fused to the N-terminus of GFP **(Fig. 4A and fig. S6A)**. In each case the glycine residue was indeed critical for instability, as a single substitution converting the glycine residue to serine

(G2S) was sufficient to inhibit degradation **(Fig. 4A and fig. S6A, left)**. Moreover, the position of the glycine residue at the extreme N-terminus was also critical, as addition of a single serine residue upstream of the glycine (add S) stabilized the peptide-GFP fusions to a similar extent **(Fig. 4A and fig. S6A, center)**. Finally, consistent with the notion that the initiator methionine is constitutively cleaved when followed by a small residue such as glycine, deletion of the initiator methionine (ΔMet) had no stabilizing effect on any of the peptide-GFP fusions **(Fig. 4A and fig. S6A, right)**.

For some example peptides we defined the N-terminal glycine degron in detail by performing saturation mutagenesis experiments **(data S4A)**. This confirmed the absolute requirement for the exposure of glycine at the extreme N-terminus, as addition of any single amino acid upstream of the glycine resulted in stabilization of the peptide-GFP fusion **(Fig. 4B and fig. S6B-G, column labeled 'add')**. The size of the degron motif appeared to be relatively small, but some substitutions at the residues immediately downstream of the exposed glycine did exert a stabilizing effect **(Fig. 4B and fig. S6B-G)**.

$Cul2^{ZYG11B}$ and $Cul2^{ZER1}$ target N-terminal glycine

We began the search for the E3 ligase(s) responsible for targeting N-terminal glycine by using the small molecule MLN4924. MLN4924 acts as a broad inhibitor of Cullin-RING ligases (CRLs) by blocking Cullin neddylation (*18*), thus allowing us to narrow the search to either CRL or non-CRL ligase families. All our example Ub-GPS constructs bearing N-terminal glycine were stabilized upon treatment with MLN4924, implicating CRLs in the recognition of N-terminal glycine **(Fig. 4C and fig. S7A)**.

Next we sought to identify the specific CRL adaptor(s) responsible for recognition of the N-terminal glycine degron. Employing dominant-negative constructs to inhibit each of the major Cullins, we determined that either Cul2 or Cul5 was responsible for the degradation of example Ub-GPS reporters harboring N-terminal glycine degrons **(Fig. 4D and fig. S7B)**. Using these reporter substrates, we performed a series of CRISPR/Cas9-mediated genetic screens using a library of single guide RNAs (sgRNAs) targeting known CRL2/5 substrate adaptor proteins **(fig. S8A)**. Together these screens identified ZYG11B as the CRL2 substrate adaptor responsible for recognition of the N-terminal glycine degron motif **(Fig. 4E, fig. S8B and data S5)**. Intriguingly, ZER1, which is closely related to ZYG11B (29% amino acid identity) **(fig. S8C)**, was enriched at or approaching the level of statistical significance in several screens, suggesting that these two related adaptors may collaborate in the degradation of proteins exposing N-terminal glycine **(Fig. 4F)**. The third member of the ZYG11 family, ZYG11A, did not score in any of the screens, consistent with RNA-seq data (*19*) suggesting that it is rarely expressed across human tissues **(fig. S8, C and D)**.

To examine the possibility of cooperation between ZYG11B and ZER1, we performed individual CRISPR/Cas9-mediated gene disruption experiments, ablating the function of ZYG11B or ZER1 either alone or in combination. Loss of ZYG11B alone did indeed stabilize all of the peptide-GFP fusion proteins **(Fig. 4G and fig. S9A)**, but, whilst complete stabilization was observed for two of the reporters **(fig. S9A)**, only partial stabilization was observed for the others. In contrast, loss of ZER1 alone had little stabilizing effect on any of the reporters; however, simultaneous disruption of both ZER1 and ZYG11B resulted in complete stabilization **(Fig. 4G and fig. S9A)**. Furthermore, ZYG11B and ZER1 both associated with putative substrates bearing N-terminal glycine degrons **(fig. S9B)**, and exogenous expression of either ZYG11B or ZER1 alone in ZYG11B/ZER1 double mutant cells fully restored the degradation of a peptide-GFP fusion whose stabilization required

ablation of both endogenous ZYG11B and ZER1 **(Fig. 4H)**. Finally, we validated that $Cul2^{ZYG11B}$ and $Cul2^{ZER1}$ were able to mediate the degradation of full-length proteins bearing exposed glycine residues at their N-termini **(Fig. 4, I and J and fig. S9C-E)**.

To obtain a global view of the substrates targeted by these Cul2 complexes, we compared the stability of the Ub-GPS N-terminome library in wild-type cells versus cells lacking either ZYG11B, ZER1 or both ZYG11B and ZER1 **(fig. S10A and data S3B)**. First, this revealed that ZYG11B and ZER1 share the majority of their substrates: there were 115 fusions stabilized in ZYG11B mutant cells and 36 stabilized in ZER1 mutant cells, while 488 were stabilized in the double mutant cells. Sequence analysis of these shared substrates confirmed that N-terminal glycine was the most enriched feature, whilst also highlighting preferred (F, G, H, K and Y) and disfavored (D, E, I, P, S and T) residues at the following position **(fig. S10B)**. Of the substrates that were targeted solely by ZYG11B, over 90% encoded a glycine residue at the second position **(fig. S10C)**. Intriguingly, there was no enrichment of N-terminal glycine among the substrates exclusively targeted by ZER1 **(fig. S10D)**. This finding suggested that (1) any ZER1 substrates bearing an N-terminal glycine were also substrates for ZYG11B, and hence were still targeted for degradation in ZER1 mutant cells, and (2) whilst N-terminal glycine was indispensable for recognition by ZYG11B, in some contexts ZER1 might recognize substrates commencing with residues other than glycine. We characterized one such substrate – the N-terminal peptide derived from KCNT2 (commencing MPYL) - in detail **(fig. S11)**. In particular, saturation mutagenesis revealed that the hydrophobic residues encoded at the third and fourth position formed a critical part of the ZER1 degron, while some more flexibility was tolerated at the second position **(fig. S11G)**. However, the location of these residues relative to the front of the peptide remained critical, because the addition of a single amino acid upstream of the proline residue prevented degradation **(fig. S11G)**.

Defining the N-terminal glycine degrons recognized by ZYG11B and ZER1

To gain further insight into the specific degron motifs recognized by ZYG11B and ZER1, we examined a larger number of potential peptide-GFP substrates commencing with glycine **(fig. S12)**. These could be divided into three categories: peptides containing degrons fully stabilized upon ZYG11B mutant alone **(fig. S12A)**, peptides containing degrons stabilized partially upon ZYG11B mutant alone, but which required combined ZYG11B and ZER1 mutant for complete stabilization **(fig. S12B)**, and peptides containing degrons for which full redundancy was observed between ZYG11B and ZER1 **(fig. S12C)**. For the vast majority of the peptides in the latter two categories an aromatic residue (H, F or Y) was located downstream of the terminal glycine, supporting the idea that ZER1 might preferentially recognize bulky residues located further along the peptide chain **(fig. S12D)**.

We tested this hypothesis more rigorously by repeating the saturation mutagenesis experiments in the genetic background of either ZYG11B ablation or ZER1 ablation **(data S4, B and C)**. The results for some representative peptides are shown in **Fig. 5A-D and fig. S13**. Mutations conferring instability in wild-type cells were identical to those conferring instability in ZER1 mutant cells **(Fig. 5, A and B)**. Therefore, these residues comprise the minimal N-terminal glycine degron, which is recognized by ZYG11B. Conversely, the ZER1 degron (as revealed in ZYG11B mutant cells) is more extensive, because mutations two or more residues downstream of the terminal glycine interfered with degradation **(Fig. 5, C and D)**. Overall, these data support a model whereby both ZYG11B and ZER1 target substrates with exposed glycine residues at their N-termini; however, the recognition motif for ZYG11B is relatively small, comprising just the terminal glycine and the following residue, whereas the recognition motif for ZER1 may extend three or more residues along the polypeptide chain and preferentially comprises amino acids with bulky aromatic side chains.

N-terminal glycine degrons are depleted from metazoan proteomes

GPS-peptidome technology has already identified a suite of degron motifs lying at the C-terminus of human proteins (*13*). All of these degron motifs are depleted from the human proteome (*13*), suggesting evolutionary pressure to avoid degradation by E3 ligases that target terminal degrons. We thus examined the abundance of N-terminal glycine degrons in eukaryotic proteomes. As is the case for the residue at the extreme C-terminus of eukaryotic proteins (*13*), the identity of the residue following the initiator methionine at the N-terminus was far more variable than at all neighboring positions, suggesting that its properties are particularly important **(Fig. 5E)**. Nonetheless, glycine was encoded at almost exactly the expected frequency at the second position across a range of metazoan model organisms **(Fig. 5F, blue dots)**. However, classifying glycine residues as those favored (G followed by F, G, H, L, M or Y) or disfavored (G followed by D, E, I, N, P, R, S or T) for CRL2-mediated degradation revealed that, compared to sequences located internally, N-terminal glycine degron motifs are depleted from animal proteomes **(Fig. 5F, orange dots)**, while N-terminal glycine motifs that are not efficiently recognized by ZYG11B and ZER1 are correspondingly enriched **(Fig. 5F, green dots)**. As a control we performed a similar analysis on a panel of reference fungal proteomes, which possess Cul2 but no ZYG11B-family ortholog (*20*). Consistent with the idea that there should be no selective pressure to avoid N-terminal glycine degrons in the absence of $Cul2^{ZYG11B}$ and $Cul2^{ZER1}$, no such relationship was observed as in animal proteomes **(Fig. 5G)**. Thus, the avoidance of N-terminal glycine motifs appears to have shaped the composition of metazoan proteomes.

ZYG11B and ZER1 target protein fragments bearing N-terminal glycine following proteolytic cleavage

Endoproteolysis generates an additional source of terminal degrons (*21–23*). Caspase cleavage preferentially occurs immediately upstream of glycine residues **(Fig. 6A)**. Indeed, of the ~1800 known human caspase cleavage sites, approximately one-third result in the exposure of glycine at the N-terminus of the downstream fragment (*24*), suggesting a potential role for ZYG11B and ZER1 in the degradation of proteins cleaved during apoptosis. Moreover, in contrast to the situation at the native N-termini of human proteins **(Fig. 5F)**, we found that N-terminal glycine degrons favoring CRL2-mediated degradation were enriched at caspase cleavage sites **(Fig. 6B)**.

We used GPS to assess a potential role for ZYG11B and ZER1 in the removal of proteolytic fragments. We generated a Ub-GPS peptide library in which the 24 residues downstream of all caspase cleavage events annotated in Degrabase (*24*) and PROSPER (*25*) were fused to the N-terminus of GFP, and profiled the stability of these peptide-GFP fusions in wild-type cells versus combined ZYG11B/ZER1 mutant cells **(Fig. 6C and data S6)**. The results confirmed that $Cul2^{ZYG11B}$ and $Cul2^{ZER1}$ could target many caspase cleavage products: 225 substrates were stabilized >0.5 PSI units in both ZYG11B/ZER1 double mutant lines, of which 219 (97%) harbored an N-terminal glycine residue **(Fig. 6D;** the GPS profiles of some example substrates are shown in **Fig. 6E and fig. S14A)**.

We validated these findings in two ways. First, for a panel of example cleavage products exposing N-terminal glycine degrons, we verified that the full-length protein fragments downstream of the cleavage site were stabilized in ZYG11B/ZER1 double mutant cells **(Fig. 6F)**. Second, we demonstrated that these fragments would also be substrates for ZYG11B and ZER1 following endoproteolytic cleavage. Our initial attempts to perform these

experiments by inducing the dimerization of caspase 9 (*26*) resulted in rapid cell death. Therefore, in order to decouple proteolytic cleavage from cell death, we engineered mutant versions of four example substrates in which the caspase cleavage site was replaced with the Tobacco Etch Virus (TEV) protease cleavage site **(Fig. 6G)**. TEV protease recognizes the amino acid sequence ENLYFQ/G (where / represents the cleavage position), thus exposing an N-terminal glycine on the downstream fragment, and is active when expressed in mammalian cells (*27*, *28*). Upon expression of TEV protease, we observed destabilization of the downstream cleavage products bearing N-terminal glycine degrons in wild-type cells, but this effect was abrogated in ZYG11B/ZER1 double mutant cells **(Fig. 6G and fig. S14B)**. Thus, ZYG11B and ZER1 are likely to be involved in the clearance of proteolytic fragments following caspase cleavage during apoptosis.

ZYG11B and ZER1 function in the quality control of *N*-myristoylated proteins

Finally we considered whether the recognition of N-terminal glycine degrons might be conditionally regulated through post-translational modifications. Intriguingly, *N*-myristoylation, the process through which the 14-carbon fatty acid myristate is attached to the N-terminus of a subset of eukaryotic proteins (*29*), occurs exclusively on N-terminal glycine **(Fig. 7A)**. Given that our mutagenesis experiments showed that addition of just a single amino acid to the N-terminus prevented ZYG11B- and ZER1-mediated recognition, we reasoned that *N*-myristoylation would prevent CRL2-mediated degradation via N-terminal glycine. Thus we hypothesized that ZYG11B and ZER1 might play an important role in "myristoylation quality control", degrading proteins bearing N-terminal glycine degrons conditionally exposed following a failure of *N*-myristoylation.

Given that the *N*-myristoyltransferase enzymes (NMT1 and NMT2 in human cells) require less than the first 20 residues for substrate recognition (*29*), we reasoned that the peptide-GFP fusion proteins expressed from our N-terminome Ub-GPS library should undergo native *N*-myristoylation. In order to examine the effect of *N*-myristoylation on protein stability, we profiled the N-terminome Ub-GPS library in the presence or absence of NMT1/2 **(Fig. 7B and data S3C)**. Although we were not able to generate clones in which both NMT1 and NMT2 were completely ablated following CRISPR/Cas9-mediated gene disruption – a finding consistent with the notion that *N*-myristoylation is an essential process (*30*) – we did isolate three clones that retained only residual levels of one NMT enzyme as assessed by immunoblot **(Fig. 7C)**. When we analyzed the composition of all the peptide-GFP fusion proteins whose stability was significantly reduced in all three NMT1/2 mutant clones, N-terminal glycine was the most enriched feature **(Fig. 7D)**. Thus, a failure to undergo *N*-myristoylation can lead to instability of the unmodified protein.

To investigate a possible role for ZYG11B and ZER1 in this process, we examined the stability of a panel of example substrates **(fig. S15A)** wherein N-terminal peptides derived from proteins known to undergo *N*-myristoylation (*31*) were expressed in the presence and absence of both NMT1/2 and ZYG11B/ZER1. These peptide-GFP fusion proteins were indeed efficiently myristoylated, as evidenced by membrane localization in wild-type cells but not in NMT1/2 mutant cells **(fig. S15B)**. Validating the screen results, in each case we observed destabilization of the peptide-GFP fusion protein upon loss of NMT1/2 **(Fig. 7E, gold histograms)**; moreover, ZYG11B and ZER1 were primarily responsible for this instability, because complete or near-complete re-stabilization was observed upon ablation of both NMT1/2 and ZYG11B/ZER1 **(Fig. 7E, purple histograms)**. The true magnitude of this effect is likely to be even greater, because addition of the small molecule NMT1/2 inhibitor IMP-1088 (*32*) to the NMT1/2 mutant clones, thereby inhibiting the residual *N*-

myristoyltransferase activity remaining in the cell, further enhanced the destabilization of the peptide-GFP substrates **(fig. S15C)**. Moreover, the small degree of stabilization observed with some of the fusion proteins upon ablation ZYG11B and ZER1 in wild-type (that is, NMT1/2-sufficient) cells **(Fig. 7E, top row)** suggested that some fraction of protein molecules do normally escape *N*-myristoylation, emphasizing the necessity for a degradative mechanism to remove these aberrant species.

Lastly, we wanted to validate that endogenous *N*-myristoylated proteins behaved in a similar manner. Indeed, we observed a significant reduction in the steady-state levels of a panel of example substrates in NMT1/2 mutant cells, which was abrogated upon concurrent ablation of ZYG11B and ZER1 **(Fig. 7F)**. However, unlike the complete or near-complete stabilization that we observed using the peptide-GFP fusion constructs **(Fig. 7E)**, here dual ZYG11B/ZER1 mutant only resulted in partial re-stabilization. Thus, in the context of full-length proteins, multiple degrons in addition to N-terminal glycine may be exposed following a failure of *N*-myristoylation, rendering them substrates for additional E3 ligases. Altogether, these data demonstrate a physiological role for ZYG11B and ZER1 in the surveillance of myristoylated proteins: successful *N*-myristoylation shields proteins from degradation, but a failure to undergo *N*-myristoylation results in the exposure of N-terminal glycine degrons and CRL2-mediated degradation **(Fig. 7G)**.

Discussion

Here we exploited GPS technology to examine directly the contribution of N-terminal sequences to protein stability across the human proteome. Unexpectedly, in addition to targeting abnormal proteins lacking an initiator methionine, we discovered that UBR-family E3 ligases also targeted proteins with a native N-terminus in which an arginine or lysine

residue follows an intact initiator methionine. We also found that cysteine exposed at the N-terminus of GFP conferred instability in a UBR-dependent manner. Nitric oxide-mediated oxidation of N-terminal cysteine renders it a substrate for arginylation by ATE1 and hence UBR-mediated degradation (*16*). However, here substrates bearing N-terminal cysteine were not stabilized to the same extent in ATE1 mutant cells as in cells lacking UBR proteins; thus, if UBR proteins do not directly bind N-terminal cysteine, an ATE1-independent pathway must exist that permits this class of degrons to be recognized by UBR E3 ligases.

Most significantly we uncovered an additional N-degron pathway centered on N-terminal glycine. There are intriguing mechanistic similarities between the ZYG11B- and ZER1-mediated recognition of N-terminal glycine degrons and the KLHDC2-, KLHDC3 and KLHDC10-mediated recognition of C-terminal glycine degrons (*13*), with both processes involving multiple related members of CRL2 substrate adaptor families. Like the Kelch repeats found in the KLHDC family proteins, the leucine-rich repeats and the armadillo-like repeats present in the ZYG11 family adaptors also have the propensity to form solenoid structures (*33*), raising the possibility of a common structural mode through which terminal glycine residues are engaged (*34*). Furthermore, like their C-terminal counterparts, the ZYG11 family of substrate adaptors have also shaped the proteome, with N-terminal glycine degrons being broadly avoided across metazoa.

Our data suggests two contexts in which the targeting of N-terminal glycine degrons may play an important physiological role. *N*-myristoylation is a post-translational modification regulating the membrane localization and other properties of several hundred human proteins (*29*), a group which comprises notable members including Arf family GTPases, G protein alpha subunits and Src family tyrosine kinases (*35*). We propose a model whereby a failure of *N*-myristoylation conditionally exposes N-terminal glycine degrons to ZYG11B and ZER1, which are normally occluded upon successful modification. Further

work will be required to ascertain whether other classes of terminal degrons function in analogous quality control pathways to ensure the efficient deposition of post-translational modifications.

Furthermore, the strong enrichment for favored ZYG11B/ZER1 glycine degrons at the N-termini of known caspase cleavage products suggested a potential role for these CRL2 complexes during apoptosis, and we confirmed experimentally that many caspase cleavage events would generate substrates efficiently degraded by $Cul2^{ZYG11B}$ and $Cul2^{ZER1}$. After glycine, the next most commonly generated N-terminal residue following caspase cleavage is serine, accounting for ~28% of annotated caspase sites **(Fig. 6A)**. Intriguingly, in complete contrast to glycine, serine is the most stabilizing residue when exposed at the N-terminus **(Fig. 2A)**. Indeed, our caspase cleavage product GPS screen showed that fragments bearing N-terminal serine were generally extremely stable **(fig. S14C)**. This may be useful where caspases need to activate a target, such as in the case of ATM, whose C-terminal cleavage product acts in a dominant-negative manner to prevent DNA repair during apoptosis (*36*), or RAD21, whose C-terminal cleavage product acts as a pro-apoptotic factor (*37*).

The significance of this glycine-specific N-degron pathway to human biology is underscored by the fact that the frequency of heterozygous loss-of-function mutations in humans for both ZYG11B and ZER1 is far lower than would be predicted. ZYG11B and ZER1 both have a pLi value of 1 in the ExAC database (*38*), indicating that loss-of-function variants are strongly selected against in the heterozygous state thereby demonstrating potent haploinsufficiency and counter selection in humans. Misregulation of Src-family tyrosine kinases could be deleterious to development. In *C. elegans*, the ZYG11 ortholog is required for the metaphase to anaphase transition and M phase exit at meiosis II (*20*, *39*, *40*). In humans ZYG11B and ZER1 are both expressed in the testes and ovaries, and hence a similar role in the regulation of meiosis could also explain the strong selection against loss-of-

function mutations. Altogether, the comprehensive analysis of N-terminal degrons presented here has illuminated multiple new aspects of N-degron proteolytic pathways and revealed that a family of E3 ligases specific for N-terminal glycine has shaped the human proteome.

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. 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 (*41*)), 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 (*27*)).

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 (*42*). 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 (*24*) and PROSPER (*25*). 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 (*43*). Resulting 72 nt reads were mapped to the reference input library using Bowtie 2 (*44*), 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):

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

(where $\boldsymbol{i}$ 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 (*45*) and rendered using Seq2Logo (*46*): 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. S12** 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**.

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 (*47*) 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 S5**. 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 (*24*) 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 S6)**. 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.

References and Notes:

1. G. Kleiger, T. Mayor, Perilous journey: a tour of the ubiquitin-proteasome system. *Trends Cell Biol.* **24**, 352–9 (2014).
2. T. Ravid, M. Hochstrasser, Diversity of degradation signals in the ubiquitin–proteasome system. *Nat. Rev. Mol. Cell Biol.* **9**, 679–689 (2008).
3. B. Mészáros, M. Kumar, T. J. Gibson, B. Uyar, Z. Dosztányi, Degrons in cancer. *Sci. Signal.* **10**, eaak9982 (2017).
4. A. Bachmair, D. Finley, A. Varshavsky, In vivo half-life of a protein is a function of its amino-terminal residue. *Science*. **234**, 179–86 (1986).
5. A. Varshavsky, N-degron and C-degron pathways of protein degradation. *Proc. Natl. Acad. Sci.* **116**, 358–366 (2019).
6. B. Bartel, I. Wünning, A. Varshavsky, The recognition component of the N-end rule pathway. *EMBO J.* **9**, 3179–89 (1990).
7. A. Varshavsky, The N-end rule pathway and regulation by proteolysis. *Protein Sci.* **20**, 1298–1345 (2011).
8. C.-S. Hwang, A. Shemorry, A. Varshavsky, N-Terminal Acetylation of Cellular Proteins Creates Specific Degradation Signals. *Science (80-.).* **327**, 973–977 (2010).
9. A. Shemorry, C.-S. Hwang, A. Varshavsky, Control of Protein Quality and Stoichiometries by N-Terminal Acetylation and the N-End Rule Pathway. *Mol. Cell*. **50**, 540–551 (2013).
10. S.-J. Chen, X. Wu, B. Wadas, J.-H. Oh, A. Varshavsky, An N-end rule pathway that recognizes proline and destroys gluconeogenic enzymes. *Science (80-.).* **355**, eaal3655 (2017).
11. I. Kats *et al.*, Mapping Degradation Signals and Pathways in a Eukaryotic N-terminome. *Mol. Cell*. **70**, 488–501.e5 (2018).
12. H.-C. S. Yen, Q. Xu, D. M. Chou, Z. Zhao, S. J. Elledge, Global Protein Stability Profiling in Mammalian Cells. *Science (80-.).* **322**, 918–923 (2008).
13. I. Koren *et al.*, The Eukaryotic Proteome Is Shaped by E3 Ubiquitin Ligases Targeting C-Terminal Degrons. *Cell*. **173**, 1622–1635 (2018).
14. F. Sherman, J. W. Stewart, S. Tsunasawa, Methionine or not methionine at the beginning of a protein. *BioEssays*. **3**, 27–31 (1985).
15. T. Tasaki *et al.*, A Family of Mammalian E3 Ubiquitin Ligases That Contain the UBR Box Motif and Recognize N-Degrons. *Mol. Cell. Biol.* **25**, 7120–7136 (2005).
16. R.-G. Hu *et al.*, The N-end rule pathway as a nitric oxide sensor controlling the levels of multiple regulators. *Nature*. **437**, 981–6 (2005).
17. H.-K. Kim *et al.*, The N-Terminal Methionine of Cellular Proteins as a Degradation Signal. *Cell*. **156**, 158–169 (2014).
18. T. A. Soucy *et al.*, An inhibitor of NEDD8-activating enzyme as a new approach to treat cancer. *Nature*. **458**, 732–736 (2009).

19. K. G. Ardlie *et al.*, The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science (80-.).* **348**, 648–660 (2015).

20. S. Vasudevan, N. G. Starostina, E. T. Kipreos, The Caenorhabditis elegans cell-cycle regulator ZYG-11 defines a conserved family of CUL-2 complex components. *EMBO Rep.* **8**, 279–286 (2007).

21. K. I. Piatkov, C. S. Brower, A. Varshavsky, The N-end rule pathway counteracts cell death by destroying proapoptotic protein fragments. *Proc. Natl. Acad. Sci.* **109**, E1839–E1847 (2012).

22. K. I. Piatkov, L. Colnaghi, M. Békés, A. Varshavsky, T. T. Huang, The Auto-Generated Fragment of the Usp1 Deubiquitylase Is a Physiological Substrate of the N-End Rule Pathway. *Mol. Cell.* **48**, 926–933 (2012).

23. K. I. Piatkov, J.-H. Oh, Y. Liu, A. Varshavsky, Calpain-generated natural protein fragments as short-lived substrates of the N-end rule pathway. *Proc. Natl. Acad. Sci.* **111**, E817–E826 (2014).

24. E. D. Crawford *et al.*, The DegraBase: a database of proteolysis in healthy and apoptotic human cells. *Mol. Cell. Proteomics.* **12**, 813–24 (2013).

25. J. Song *et al.*, PROSPER: an integrated feature-based tool for predicting protease substrate cleavage sites. *PLoS One.* **7**, e50300 (2012).

26. K. C. Straathof *et al.*, An inducible caspase 9 safety switch for T-cell therapy. *Blood.* **105**, 4247–4254 (2005).

27. T.-L. To *et al.*, Rationally designed fluorogenic protease reporter visualizes spatiotemporal dynamics of apoptosis in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 3338–43 (2015).

28. T.-L. To *et al.*, Rational Design of a GFP-Based Fluorogenic Caspase Reporter for Imaging Apoptosis In Vivo. *Cell Chem. Biol.* **23**, 875–882 (2016).

29. M. H. Wright, W. P. Heal, D. J. Mann, E. W. Tate, Protein myristoylation in health and disease. *J. Chem. Biol.* **3**, 19–35 (2010).

30. S. H. Yang *et al.*, *N* -Myristoyltransferase 1 Is Essential in Early Mouse Development. *J. Biol. Chem.* **280**, 18990–18995 (2005).

31. E. Thinon *et al.*, Global profiling of co- and post-translationally N-myristoylated proteomes in human cells. *Nat. Commun.* **5**, 4919 (2014).

32. A. Mousnier *et al.*, Fragment-derived inhibitors of human N-myristoyltransferase block capsid assembly and replication of the common cold virus. *Nat. Chem.* **10**, 599–606 (2018).

33. I. S. Gul, P. Hulpiau, Y. Saeys, F. van Roy, Metazoan evolution of the armadillo repeat superfamily. *Cell. Mol. Life Sci.* **74**, 525–541 (2017).

34. D.-V. Rusnac *et al.*, Recognition of the Diglycine C-End Degron by CRL2KLHDC2 Ubiquitin Ligase. *Mol. Cell.* **72**, 813–822.e4 (2018).

35. E. Thinon *et al.*, Global profiling of co- and post-translationally N-myristoylated proteomes in human cells. *Nat. Commun.* **5**, 4919 (2014).

36. G. C. Smith, F. d'Adda di Fagagna, N. D. Lakin, S. P. Jackson, Cleavage and inactivation of ATM during apoptosis. *Mol. Cell. Biol.* **19**, 6076–84 (1999).

37. F. Chen *et al.*, Caspase proteolysis of the cohesin component RAD21 promotes apoptosis. *J. Biol. Chem.* **277**, 16775–81 (2002).

38. M. Lek *et al.*, Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. **536**, 285–291 (2016).

39. R. Sonneville, P. Gönczy, zyg-11 and cul-2 regulate progression through meiosis II and polarity establishment in C. elegans. *Development*. **131**, 3527–3543 (2004).

40. J. Liu, S. Vasudevan, E. T. Kipreos, CUL-2 and ZYG-11 promote meiotic anaphase II and the proper placement of the anterior-posterior axis in C. elegans. *Development*. **131**, 3513–3525 (2004).

41. R. S. Balachandran *et al.*, The ubiquitin ligase CRL2 ^{ZYG11} targets cyclin B1 for degradation in a conserved pathway that facilitates mitotic slippage. *J. Cell Biol.* **215**, 151–166 (2016).

42. N. E. Sanjana, O. Shalem, F. Zhang, Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods*. **11**, 783–784 (2014).

43. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*. **17**, 10 (2011).

44. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods*. **9**, 357–9 (2012).

45. N. Colaert, K. Helsens, L. Martens, J. Vandekerckhove, K. Gevaert, Improved visualization of protein consensus sequences by iceLogo. *Nat. Methods*. **6**, 786–787 (2009).

46. M. C. F. Thomsen, M. Nielsen, Seq2Logo: a method for construction and visualization of amino acid binding motifs and sequence profiles including sequence weighting, pseudo counts and two-sided representation of amino acid enrichment and depletion. *Nucleic Acids Res.* **40**, W281-7 (2012).

47. W. Li *et al.*, MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol.* **15**, 554 (2014).

Acknowledgements:

We are grateful to C. Araneo and his team for FACS and J. Paulo for mass spectrometry. We thank A. Varshavsky, J. Wells and E. Tate for advice. **Funding:** R.T.T. is a Sir Henry Wellcome Postdoctoral Fellow (201387/Z/16/Z). Z.Z. is a Croucher Foundation Honorary PhD Scholar. This work was supported by an NIH grant (AG11085) to S.J.E. and J.W.H. S.J.E. is an Investigator with the Howard Hughes Medical Institute. **Author Contributions:** Conceptualization, R.T.T., I.K. and S.J.E.; Investigation, R.T.T., I.K., D.Y.R., Z.Z.; Writing, R.T.T., I.K. and S.J.E; Supervision, J.W.H. and S.J.E. **Competing interests:** the authors declare no competing interests. **Data and materials availability:** all data is available in the main text or the supplementary materials.

Figure Legends:

Figure 1. GPS profiling of the human N-terminome.

(A) Schematic representation of the N-terminome GPS screen, in which the first 24 residues of all human proteins were expressed in the Ub-GPS vector as N-terminal fusions to GFP.

(B) Distribution of protein stability scores observed from the screen depicted in (A).

(C) Boxplots showing the distribution of stability scores for all peptides commencing with the indicated amino acid, when encoded either with (blue boxes) or without (orange) an upstream methionine residue.

(D-F) Heatmaps depicting the mean stability score for all peptides commencing with the indicated two amino acids, when encoded either with **(E)** or without **(D)** an upstream methionine residue; **(F)** illustrates the difference between the two.

Figure 2. Identification of degron motifs located at protein N-termini.

(A and B) The effect of peptide composition on protein stability. Heatmaps showing the relative depletion (blue) or enrichment (red) of each amino acid across all positions of the 23-mer peptide amongst unstable peptides **(A)** versus stable peptides **(B)**.

(C-F) Computational prediction of N-terminal degrons. (**C**) For all possible combinations of di-peptide motifs, the mean difference in stability between peptides containing the motif at the extreme N-terminus (that is, immediately following the initiator methionine) was compared to all peptides containing the motif at any other internal position in the peptide. (**D**) Classes of N-terminal degrons. The majority of the top 100 predicted destabilizing N-terminal motifs encoded either glycine (G2), lysine (K2), arginine (R2) or cysteine (C2) at the second position; some example motifs are annotated. (**E**) Boxplots showing the distribution of stability scores for all peptides in which the indicated residues were encoded at the second position (colored boxes) versus any other internal position within the peptide (gray boxes). (**F**) Boxplots showing the distribution of stability scores for all peptides with the indicated residues encoded at the second position.

Figure 3. Assessing the repertoire of UBR substrates among the human N-terminome.

(A-C) Assessing UBR-mediated degradation via N-terminal degron motifs. **(A)** CRISPR/Cas9-mediated generation of clones expressing reduced levels of UBR1, UBR2 and UBR4. **(B)** Functional validation of UBR KO clones. Optimal Arg/N-end rule substrates were highly unstable in wild-type cells, but not in UBR KO clone #2 as measured by flow cytometry. See also fig. S2A. **(C)** UBR proteins target example peptide-GFP reporters in which lysine, arginine or cysteine, but not glycine, are encoded at the second position. N-terminal peptides derived from the indicated genes were expressed in wild-type or UBR KO clone #2 using the Ub-GPS system and their stability assessed by flow cytometry.

(D-F) Global identification of N-terminal UBR substrates. **(D)** Schematic representation of the GPS screen. **(E)** Venn diagram summarizing the substrates stabilized >0.8 PSI units across the three UBR KO clones. **(F)** Heatmap showing the relative enrichment (red) or depletion (blue) of each amino acid across all positions of the 23-mer peptide comparing peptides stabilized in two or more of the UBR KO clones relative to the whole N-terminome library. See also fig. S4.

(G-I) Characterization of N-terminal UBR degrons through saturation mutagenesis. Each of the first ten residues of the N-terminal peptides derived from ZNF334 (commencing MK-) **(G)**, AHRR (commencing MR-) **(H)** and CDX1 (commencing (MC-) **(I)** were mutated to all other possible residues and their stabilities measured by FACS and Illumina sequencing. The darker the color, the greater the degree of stabilization compared to the wild-type sequence. See also fig. S5.

Figure 4. Cul2^{ZYG11B} and Cul2^{ZER1} target N-terminal glycine.

(A) Glycine at the N-terminus can act as a potent degron. The N-terminal peptide derived from *SNX11* and a mutant version lacking the initiator methionine (ΔMet) were highly unstable, whilst 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. See also fig. S6A.

(B) Defining N-terminal glycine degrons through saturation mutagenesis. Each of the first ten residues of the *SNX11* peptide were mutated to all other possible residues and their stabilities measured by FACS and Illumina sequencing. The darker the color, the greater the degree of stabilization compared to the wild-type sequence. See also fig. S6B.

(C and D) Cul2 complexes target N-terminal glycine. Stabilization of the SNX11-GFP reporter upon treatment with the CRL inhibitor MLN4924 **(C)**, and following expression of dominant-negative versions of Cullins **(D)**. See also fig. S7.

(E and F) CRISPR screens identify the Cul2 substrate adaptors responsible for the recognition of N-terminal glycine. **(E)** Results of the SNX11-GFP reporter screen, which highlighted two CRL2 complexes **(F)**. See also fig. S8.

(G-K) Cul2^{ZYG11B} and Cul2^{ZER1} cooperate to target N-terminal glycine. **(G)** CRISPR-mediated ablation of both ZYG11B and ZER1 was required for full stabilization of the SNX11-GFP reporter. **(H)** Exogenous expression of either ZYG11B or ZER1 rescued degradation of the SNX11-GFP reporter in cells lacking endogenous ZYG11B and ZER1. **(I and J)** Knockout of ZYG11B and ZER1 stabilized full-length SNX11 fused to the N-terminus of GFP, both when expressed in the context of the Ub-GPS system **(I)** or without upstream ubiquitin fusion **(J)**. See also fig. S9.

Figure 5. N-terminal glycine degrons are depleted from metazoan proteomes.

(A-D) Defining the degrons recognized by ZYG11B and ZER1 through saturation mutagenesis. Each of the first ten residues of the SNX11 N-terminal peptide were mutated to all possible amino acids; the stability of each mutant in the resulting Ub-GPS library was measured in wild-type **(A)**, ZER1 mutant **(B)** or ZYG11B mutant **(C)** cells. The color scale reflects the raw PSI measurement for each peptide-GFP fusion, such that the greater the intensity of the red color, the greater the stabilizing effect of the mutation. The heatmap in **(D)** illustrates the difference between the PSI in ZYG11B mutant cells versus ZER1 mutant cells; 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. See also fig. S13.

(E) Normalized amino acid frequencies across the first ten residues (following the initiator methionine) of human proteins.

(F and G) Depletion of N-terminal glycine degrons in metazoan proteomes. The normalized amino acid frequency of glycine encoded at the second position in the indicated proteomes is shown by the blue dots, and is further categorized depending on whether the glycine residue is followed by a residue favoring (orange dots) or disfavoring (green dots) CRL2-mediated degradation. The relationship observed across animal proteomes **(F)** is not apparent across fungal proteomes **(G)**, which do not possess a ZYG11 ortholog. (***$P < 0.001$, Fisher's exact test)

Figure 6. ZYG11B and ZER1 target N-terminal glycine degrons generated through endoproteolytic cleavage.

(A and B) Caspase cleavage preferentially generates fragments bearing N-terminal glycine. **(A)** Logoplot depicting the consensus sequence surrounding all caspase cleavage sites annotated in Degrabase (*24*). **(B)** Compared to their frequency across the human proteome, preferred glycine degrons (orange bar) are enriched at known caspase cleavage sites, while disfavored glycine degrons (green bar) are depleted.

(C-E) Caspase cleavage events generate N-terminal glycine degrons targeted by ZYG11B and ZER1. **(C)** Schematic representation of the caspase cleavage product Ub-GPS screen. **(D)** Heatmap showing the relative enrichment (red) or depletion (blue) of each amino acid across all positions of the 23-mer peptide comparing peptides stabilized in both ZYG11B/ZER1 double mutant cells compared to the whole caspase cleavage site library. **(E)** Profiles of example substrates. Residues flanking the caspase cleavage site (indicated by arrows) are shown. See also fig. S14A.

(F and G) CRL2-mediated degradation of proteolytic cleavage products bearing N-terminal glycine degrons. **(F)** The full-length downstream caspase cleavage products of the indicated proteins were expressed using the Ub-GPS system, and their stability assessed in wild-type (gray) and dual ZYG11B/ZER1 mutant cells (red) by flow cytometry. **(G)** The caspase site in the indicated full-length ORFs was replaced with a TEV protease cleavage site. Upon TEV expression (blue histograms), destabilization of the downstream cleavage products bearing N-terminal glycine degrons was observed in wild-type cells (top row), but not in dual ZYG11B/ZER1 mutant cells (bottom). See also fig. S14B.

Figure 7. Cul2^{ZYG11B} and Cul2^{ZER1} target proteins that fail to undergo *N*-myristoylation for proteasomal degradation.

(A) *N*-myristoylation occurs on N-terminal glycine.

(B) Schematic representation of the GPS screen designed to assess the effect of loss of *N*-myristoylation on protein stability.

(C) Immunoblot validation of NMT1/2 knockout clones. Arrowheads indicate bands of the expected molecular weight.

(D) Loss of *N*-myristoylation destabilizes peptide-GFP fusions commencing with glycine. The heatmap shows the relative enrichment (red) or depletion (blue) of each amino acid across all positions of the 23-mer peptide, comparing the 91 peptides exhibiting significant destabilization in all three NMT1/2 mutant clones relative to the whole N-terminome library.

(E and F) ZYG11B and ZER1 target N-terminal glycine degrons exposed following a failure of *N*-myristoylation. **(E)** The first 24 amino acids from the indicated proteins were expressed as N-terminal fusions to GFP, and their stability in the indicated genetic backgrounds was measured by flow cytometry. The destabilization observed upon loss of NMT1/2 (gold histograms) is rescued in the absence of ZYG11B and ZER1 (purple histogram). **(F)** The abundance of the indicated myristoylated proteins was assessed by immunoblot in either control (AAVS1) or ZYG11B/ZER1 double mutant cells (DKO), either with or without simultaneous ablation of NMT1/2. Src, which did not exhibit significant destabilization in NMT1/2 mutant cells in the GPS screen, is shown as a negative control.

(G) Model depicting the role of Cul2^{ZYG11B} and Cul2^{ZER} in the quality control of *N*-myristoylated proteins.

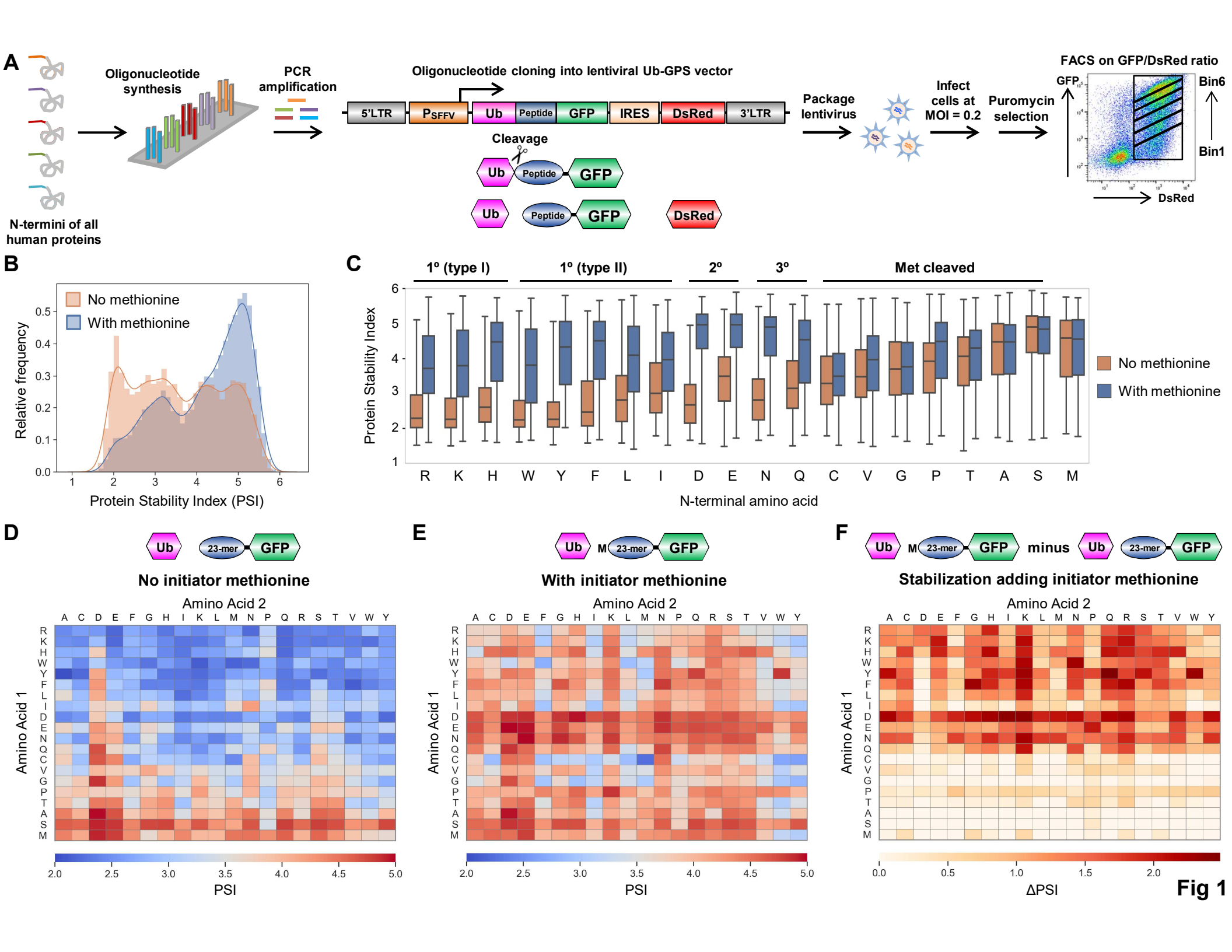
A
N-termini of all human proteins
Oligonucleotide synthesis
PCR amplification
Oligonucleotide cloning into lentiviral Ub-GPS vector
5'LTR
PSFFV
Ub
Peptide
GFP
IRES
DsRed
3'LTR
Cleavage
Package lentivirus
Infect cells at MOI = 0.2
Puromycin selection
FACS on GFP/DsRed ratio
Bin6
Bin1
B
No methionine
With methionine
Relative frequency
Protein Stability Index (PSI)
C
1º (type I)
1º (type II)
2º
3º
Met cleaved
Protein Stability Index
N-terminal amino acid
D
No initiator methionine
E
With initiator methionine
F
minus
Stabilization adding initiator methionine
Amino Acid 1
Amino Acid 2
PSI
ΔPSI

Fig 1

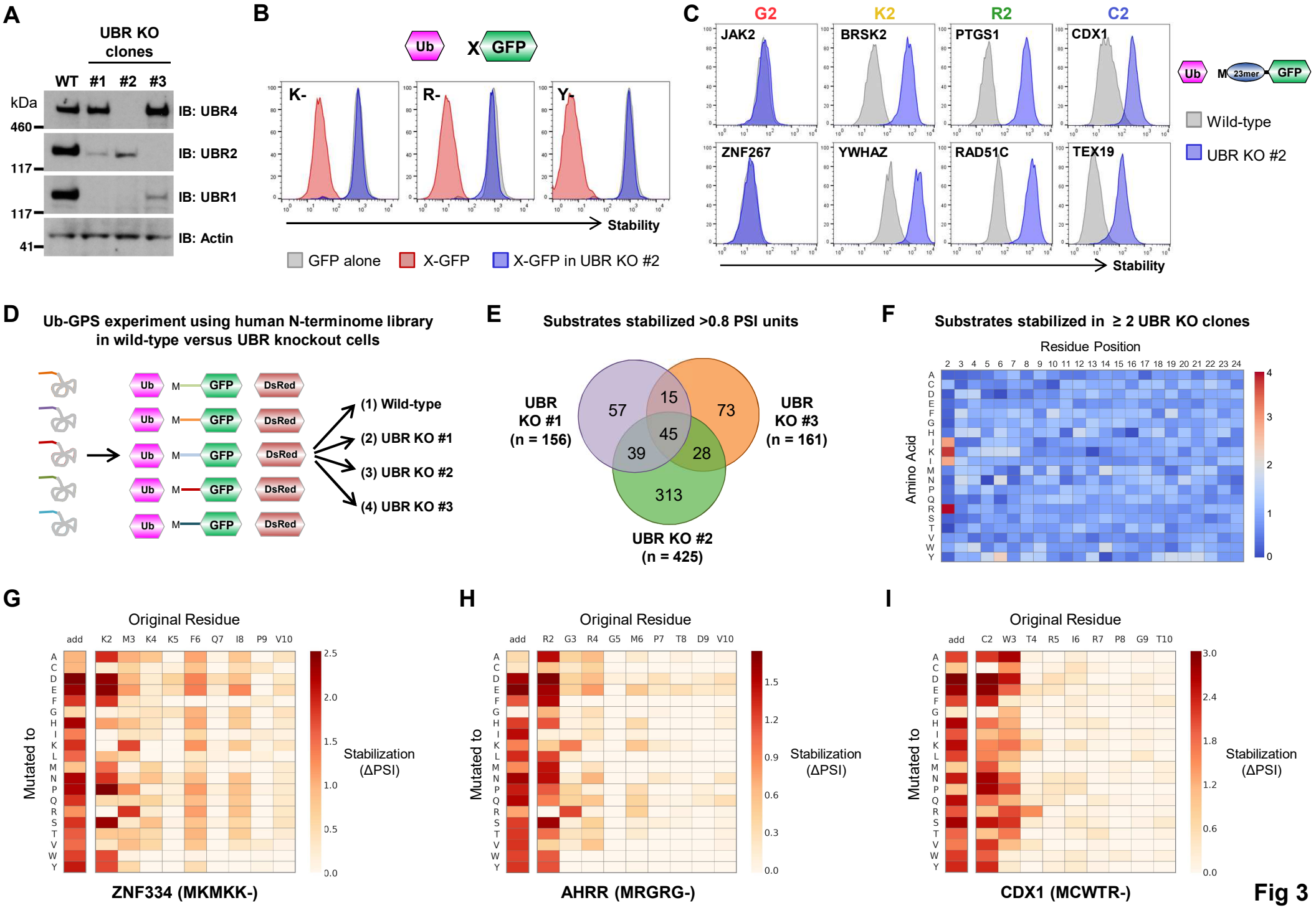
A
UBR KO clones
WT #1 #2 #3
kDa
460
117
117
41
IB: UBR4
IB: UBR2
IB: UBR1
IB: Actin
B
Ub
X GFP
K-
R-
Y-
Stability
GFP alone
X-GFP
X-GFP in UBR KO #2
C
G2
K2
R2
C2
JAK2
BRSK2
PTGS1
CDX1
ZNF267
YWHAZ
RAD51C
TEX19
Stability
Ub
M
23mer
GFP
Wild-type
UBR KO #2
D
Ub-GPS experiment using human N-terminome library in wild-type versus UBR knockout cells
Ub
M
GFP
DsRed
(1) Wild-type
(2) UBR KO #1
(3) UBR KO #2
(4) UBR KO #3
E
Substrates stabilized >0.8 PSI units
UBR KO #1 (n = 156)
UBR KO #3 (n = 161)
UBR KO #2 (n = 425)
57
15
73
45
39
28
313
F
Substrates stabilized in ≥ 2 UBR KO clones
Residue Position
Amino Acid
G
Original Residue
add K2 M3 K4 K5 F6 Q7 I8 P9 V10
Mutated to
Stabilization (ΔPSI)
ZNF334 (MKMKK-)
H
Original Residue
add R2 G3 R4 G5 M6 P7 T8 D9 V10
Mutated to
Stabilization (ΔPSI)
AHRR (MRGRG-)
I
Original Residue
add C2 W3 T4 R5 I6 R7 P8 G9 T10
Mutated to
Stabilization (ΔPSI)
CDX1 (MCWTR-)

Fig 3

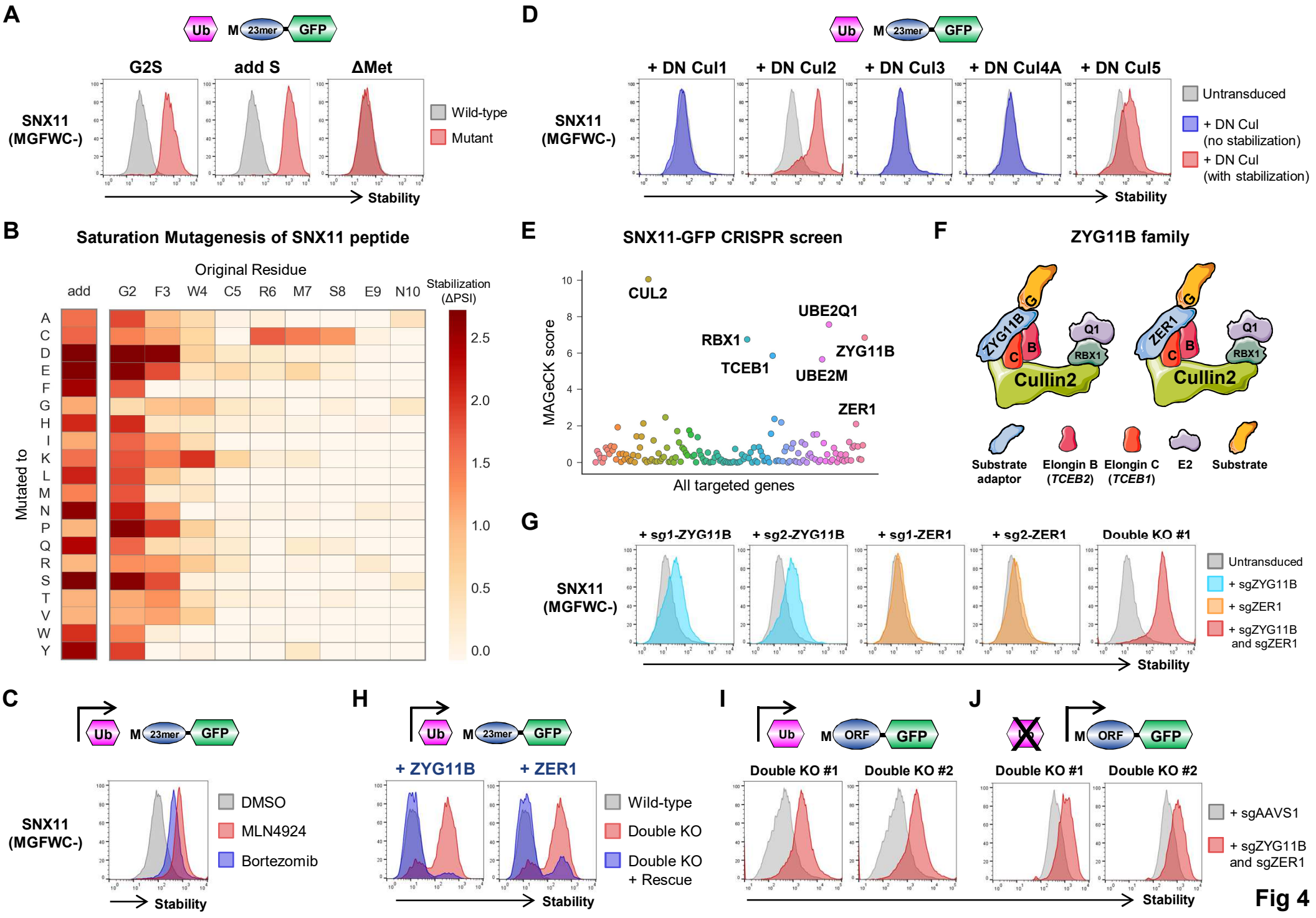
A
Ub
M
23mer
GFP
G2S
add S
ΔMet
SNX11
(MGFWC-)
Wild-type
Mutant
Stability
B
Saturation Mutagenesis of SNX11 peptide
Original Residue
add
G2
F3
W4
C5
R6
M7
S8
E9
N10
Stabilization
(ΔPSI)
Mutated to
A
C
D
E
F
G
H
I
K
L
M
N
P
Q
R
S
T
V
W
Y
2.5
2.0
1.5
1.0
0.5
0.0
C
Ub
M
23mer
GFP
SNX11
(MGFWC-)
DMSO
MLN4924
Bortezomib
Stability
D
Ub
M
23mer
GFP
+ DN Cul1
+ DN Cul2
+ DN Cul3
+ DN Cul4A
+ DN Cul5
SNX11
(MGFWC-)
Untransduced
+ DN Cul
(no stabilization)
+ DN Cul
(with stabilization)
Stability
E
SNX11-GFP CRISPR screen
MAGeCK score
All targeted genes
CUL2
UBE2Q1
RBX1
ZYG11B
TCEB1
UBE2M
ZER1
F
ZYG11B family
ZYG11B
ZER1
C
B
Q1
RBX1
Cullin2
Substrate
adaptor
Elongin B
(TCEB2)
Elongin C
(TCEB1)
E2
Substrate
G
+ sg1-ZYG11B
+ sg2-ZYG11B
+ sg1-ZER1
+ sg2-ZER1
Double KO #1
SNX11
(MGFWC-)
Untransduced
+ sgZYG11B
+ sgZER1
+ sgZYG11B
and sgZER1
Stability
H
Ub
M
23mer
GFP
+ ZYG11B
+ ZER1
SNX11
(MGFWC-)
Wild-type
Double KO
Double KO
+ Rescue
Stability
I
Ub
M
ORF
GFP
Double KO #1
Double KO #2
Stability
J
Ub
M
ORF
GFP
Double KO #1
Double KO #2
+ sgAAVS1
+ sgZYG11B
and sgZER1
Stability

Fig 4

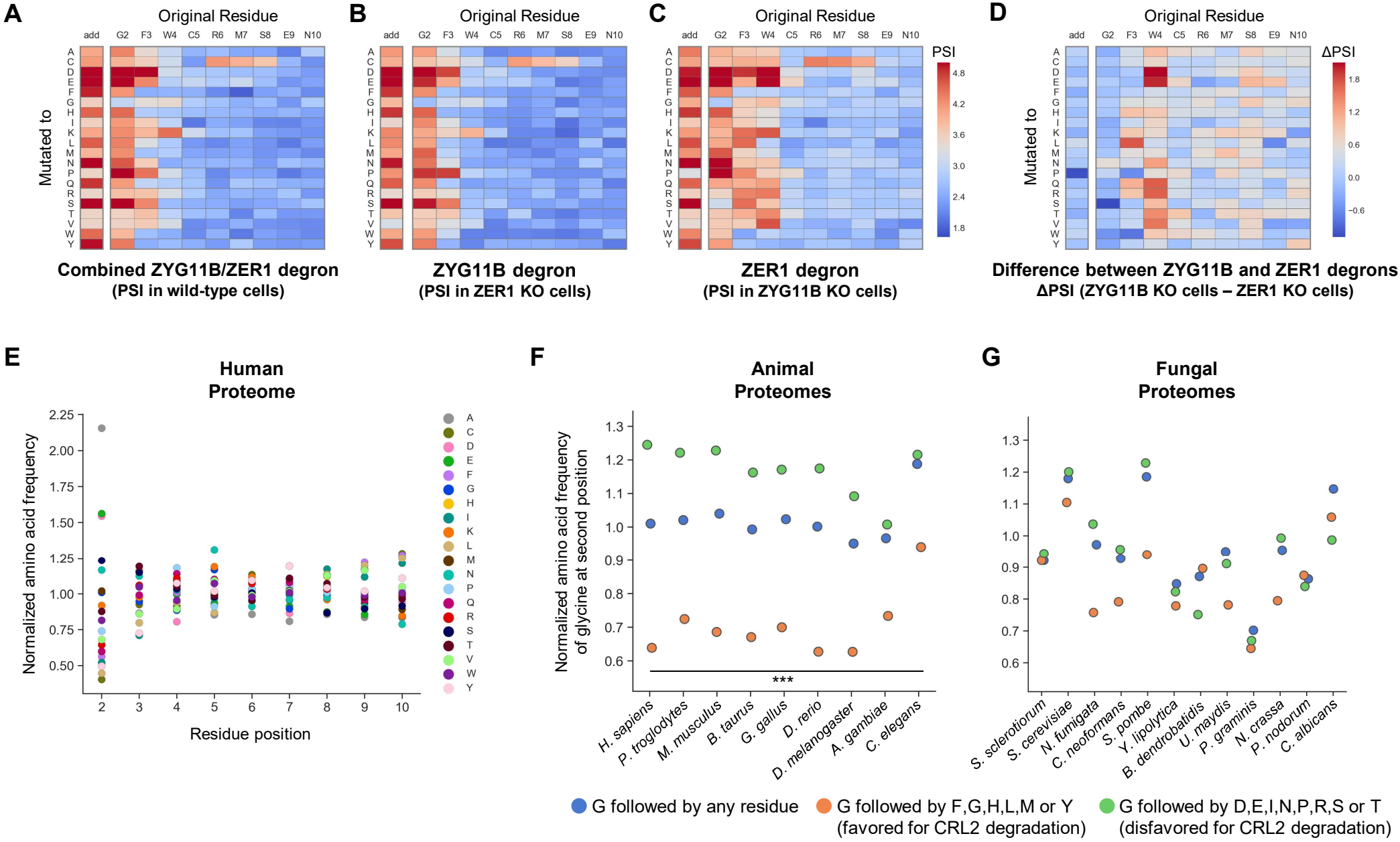
A
Original Residue
add G2 F3 W4 C5 R6 M7 S8 E9 N10
Mutated to
A C D E F G H I K L M N P Q R S T V W Y
Combined ZYG11B/ZER1 degron
(PSI in wild-type cells)
B
Original Residue
ZYG11B degron
(PSI in ZER1 KO cells)
C
Original Residue
PSI
4.8 4.2 3.6 3.0 2.4 1.8
ZER1 degron
(PSI in ZYG11B KO cells)
D
Original Residue
ΔPSI
1.8 1.2 0.6 0.0 −0.6
Difference between ZYG11B and ZER1 degrons
ΔPSI (ZYG11B KO cells – ZER1 KO cells)
E
Human Proteome
Normalized amino acid frequency
2.25 2.00 1.75 1.50 1.25 1.00 0.75 0.50
2 3 4 5 6 7 8 9 10
Residue position
A C D E F G H I K L M N P Q R S T V W Y
F
Animal Proteomes
Normalized amino acid frequency of glycine at second position
1.3 1.2 1.1 1.0 0.9 0.8 0.7 0.6
***
H. sapiens
P. troglodytes
M. musculus
B. taurus
G. gallus
D. rerio
D. melanogaster
A. gambiae
C. elegans
G
Fungal Proteomes
1.3 1.2 1.1 1.0 0.9 0.8 0.7 0.6
S. sclerotiorum
S. cerevisiae
N. fumigata
C. neoformans
S. pombe
Y. lipolytica
B. dendrobatidis
U. maydis
P. graminis
N. crassa
P. nodorum
C. albicans
G followed by any residue
G followed by F,G,H,L,M or Y
(favored for CRL2 degradation)
G followed by D,E,I,N,P,R,S or T
(disfavored for CRL2 degradation)

Fig 5

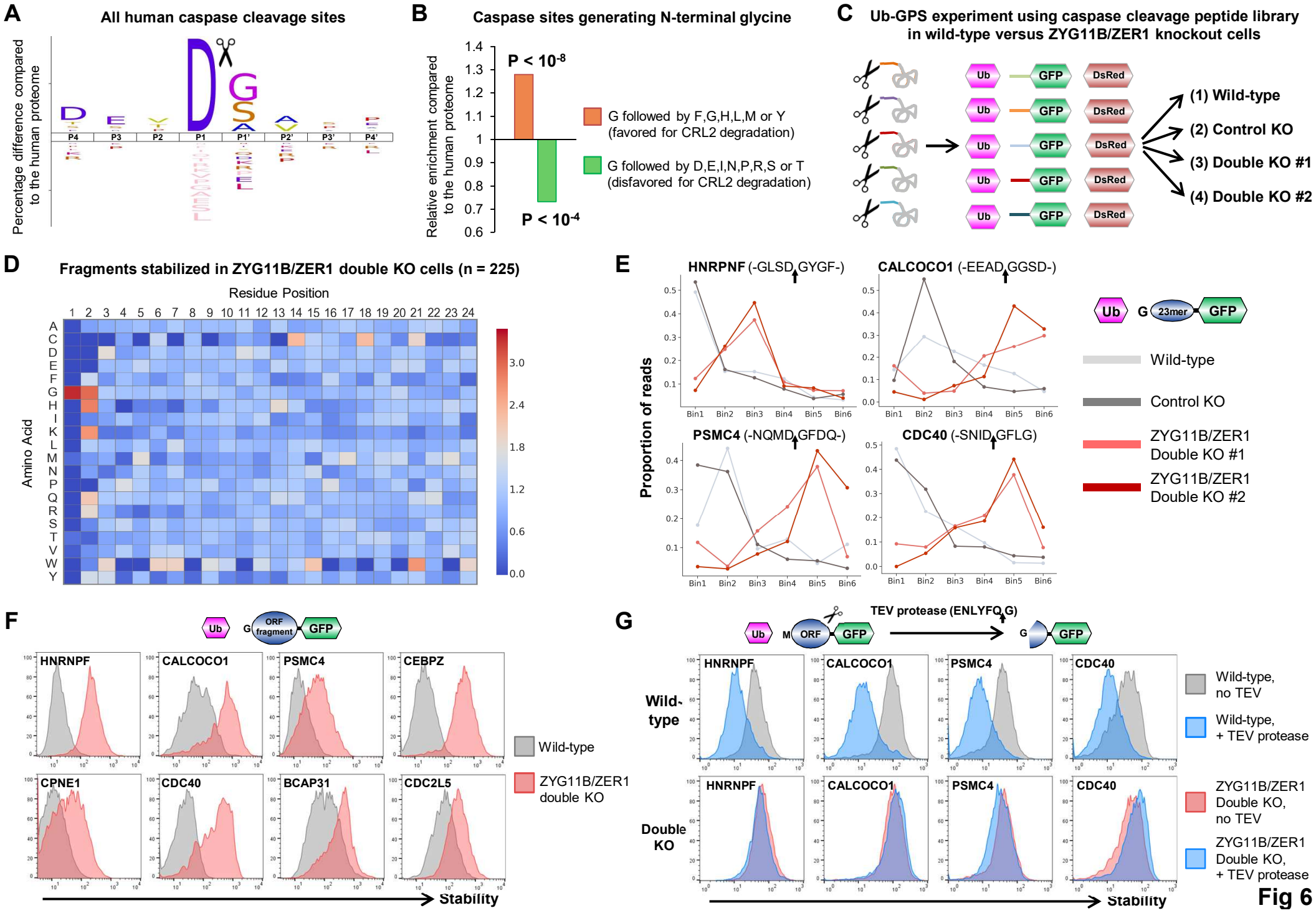
A
All human caspase cleavage sites
Percentage difference compared to the human proteome
P4
P3
P2
P1
P1'
P2'
P3'
P4'
B
Caspase sites generating N-terminal glycine
Relative enrichment compared to the human proteome
P < 10-8
P < 10-4
G followed by F,G,H,L,M or Y (favored for CRL2 degradation)
G followed by D,E,I,N,P,R,S or T (disfavored for CRL2 degradation)
C
Ub-GPS experiment using caspase cleavage peptide library in wild-type versus ZYG11B/ZER1 knockout cells
Ub
GFP
DsRed
(1) Wild-type
(2) Control KO
(3) Double KO #1
(4) Double KO #2
D
Fragments stabilized in ZYG11B/ZER1 double KO cells (n = 225)
Residue Position
Amino Acid
E
HNRPNF (-GLSD GYGF-)
CALCOCO1 (-EEAD GGSD-)
PSMC4 (-NQMD GFDQ-)
CDC40 (-SNID GFLG)
Proportion of reads
Bin1
Bin2
Bin3
Bin4
Bin5
Bin6
23mer
Wild-type
Control KO
ZYG11B/ZER1 Double KO #1
ZYG11B/ZER1 Double KO #2
F
ORF fragment
HNRPNF
CALCOCO1
PSMC4
CEBPZ
CPNE1
CDC40
BCAP31
CDC2L5
Wild-type
ZYG11B/ZER1 double KO
Stability
G
TEV protease (ENLYFQ G)
M
ORF
Wild-type
Double KO
Wild-type, no TEV
Wild-type, + TEV protease
ZYG11B/ZER1 Double KO, no TEV
ZYG11B/ZER1 Double KO, + TEV protease
Stability

Fig 6

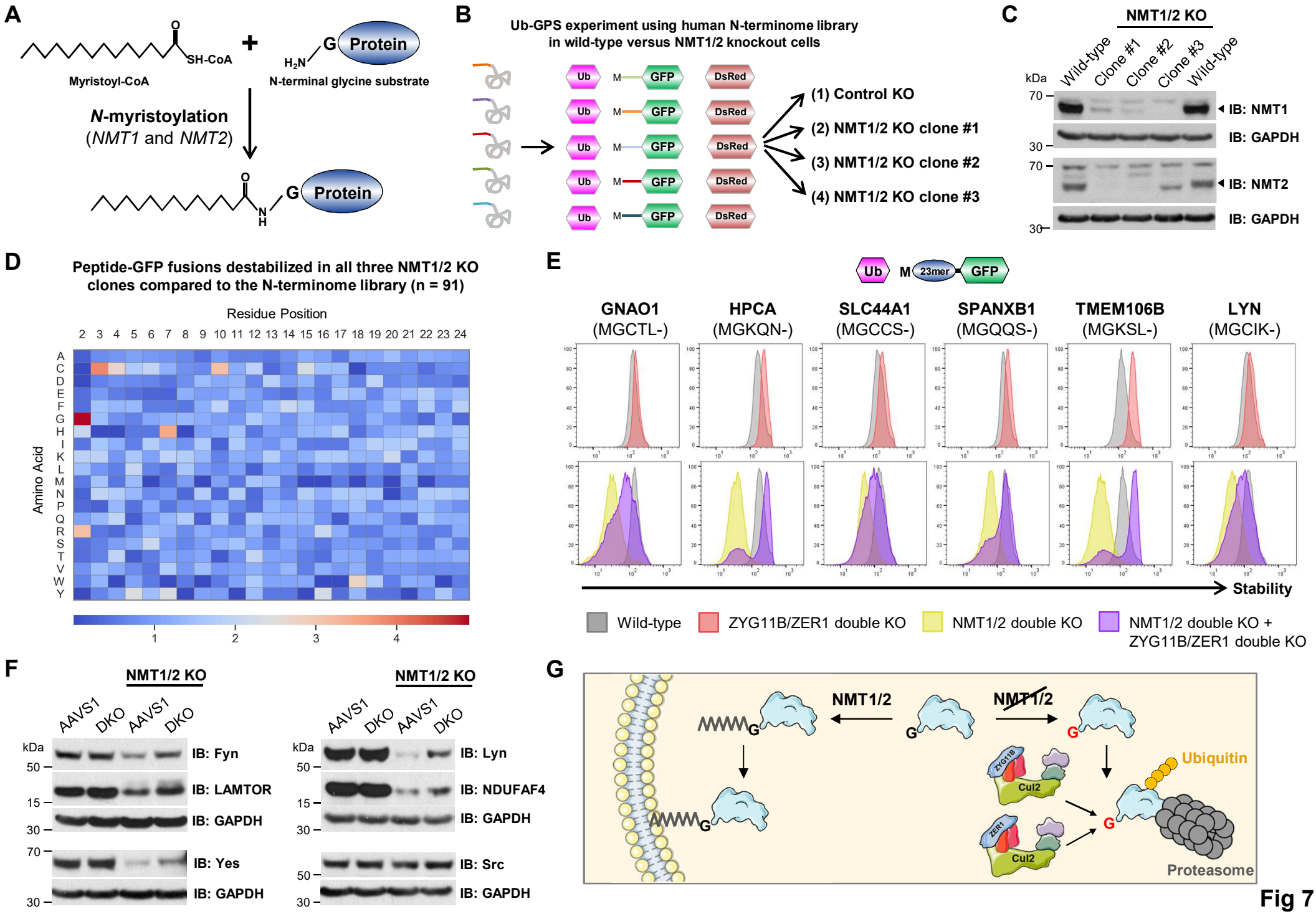
A
Myristoyl-CoA
SH-CoA
Protein
N-terminal glycine substrate
N-myristoylation
(NMT1 and NMT2)
B
Ub-GPS experiment using human N-terminome library
in wild-type versus NMT1/2 knockout cells
Ub
M
GFP
DsRed
(1) Control KO
(2) NMT1/2 KO clone #1
(3) NMT1/2 KO clone #2
(4) NMT1/2 KO clone #3
C
NMT1/2 KO
Wild-type
Clone #1
Clone #2
Clone #3
Wild-type
kDa
IB: NMT1
IB: GAPDH
IB: NMT2
IB: GAPDH
D
Peptide-GFP fusions destabilized in all three NMT1/2 KO
clones compared to the N-terminome library (n = 91)
Residue Position
Amino Acid
E
Ub
M
23mer
GFP
GNAO1
(MGCTL-)
HPCA
(MGKQN-)
SLC44A1
(MGCCS-)
SPANXB1
(MGQQS-)
TMEM106B
(MGKSL-)
LYN
(MGCIK-)
Stability
Wild-type
ZYG11B/ZER1 double KO
NMT1/2 double KO
NMT1/2 double KO +
ZYG11B/ZER1 double KO
F
NMT1/2 KO
AAVS1
DKO
IB: Fyn
IB: LAMTOR
IB: GAPDH
IB: Yes
IB: GAPDH
IB: Lyn
IB: NDUFAF4
IB: GAPDH
IB: Src
IB: GAPDH
G
NMT1/2
NMT1/2
Ubiquitin
ZYG11B
ZER1
Cul2
Proteasome

Fig 7

Supplementary Materials for

A glycine-specific N-degron 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: itay.koren@biu.ac.il, selledge@genetics.med.harvard.edu

This PDF file includes:

Figs. S1 to S15
Captions for databases S1 to S6

Other Supplementary Materials for this manuscript includes the following:

Databases S1 to S6 as zipped archives

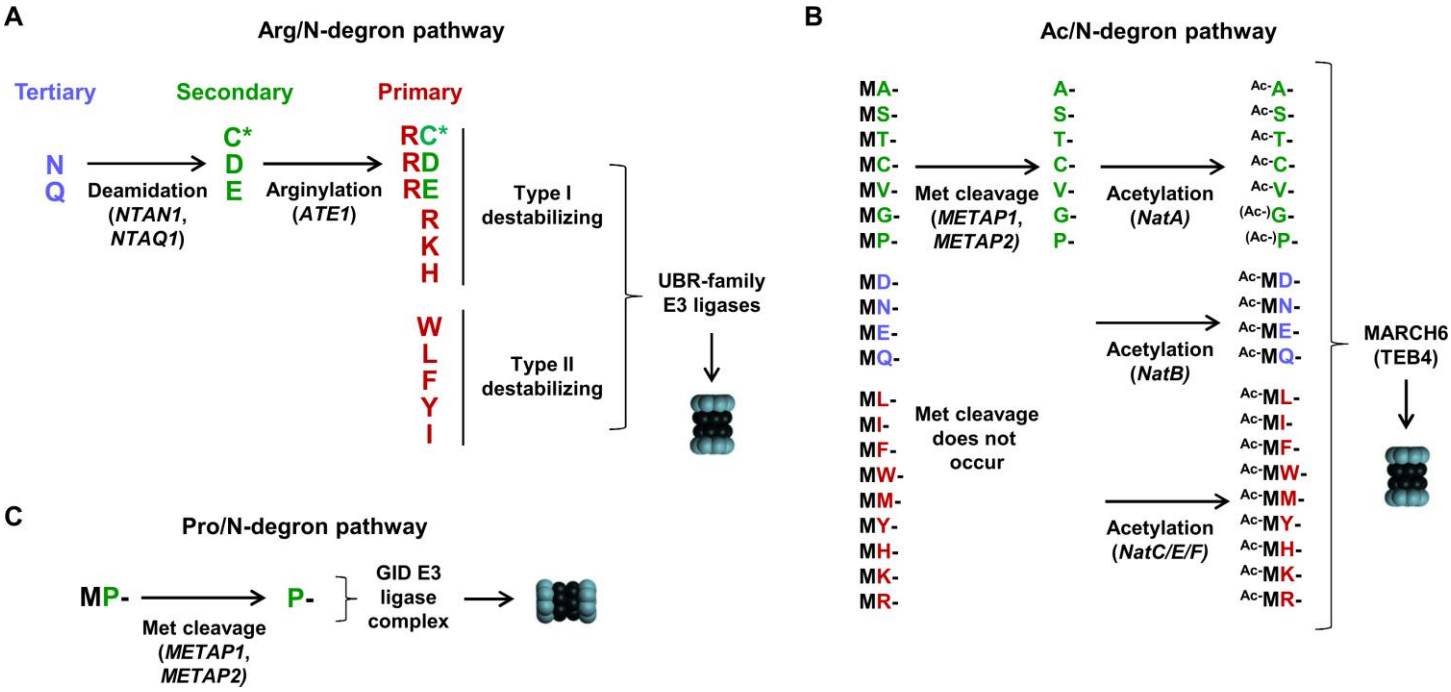

Fig. S1. Summary of known N-degron pathways.

(A) The Arg/N-degron pathway. 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-degron 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-degron pathway. 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-degron pathway. 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*).

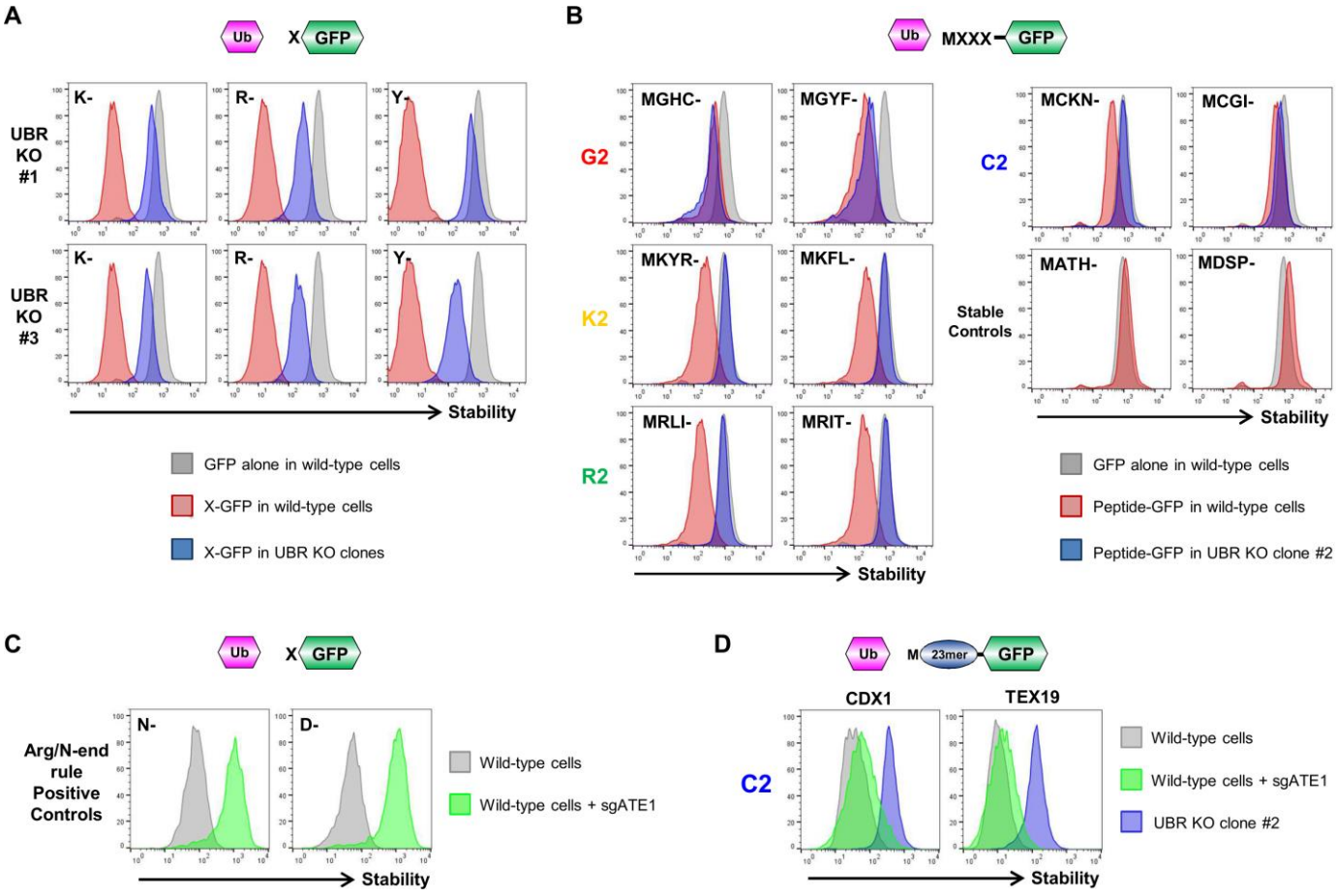

Fig. S2. Assessing the role of UBR family proteins in the degradation of proteins bearing N-terminal degron motifs.

(A) Functional validation of UBR KO clone #1 and clone #3. Optimal Arg/N-degron 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-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.

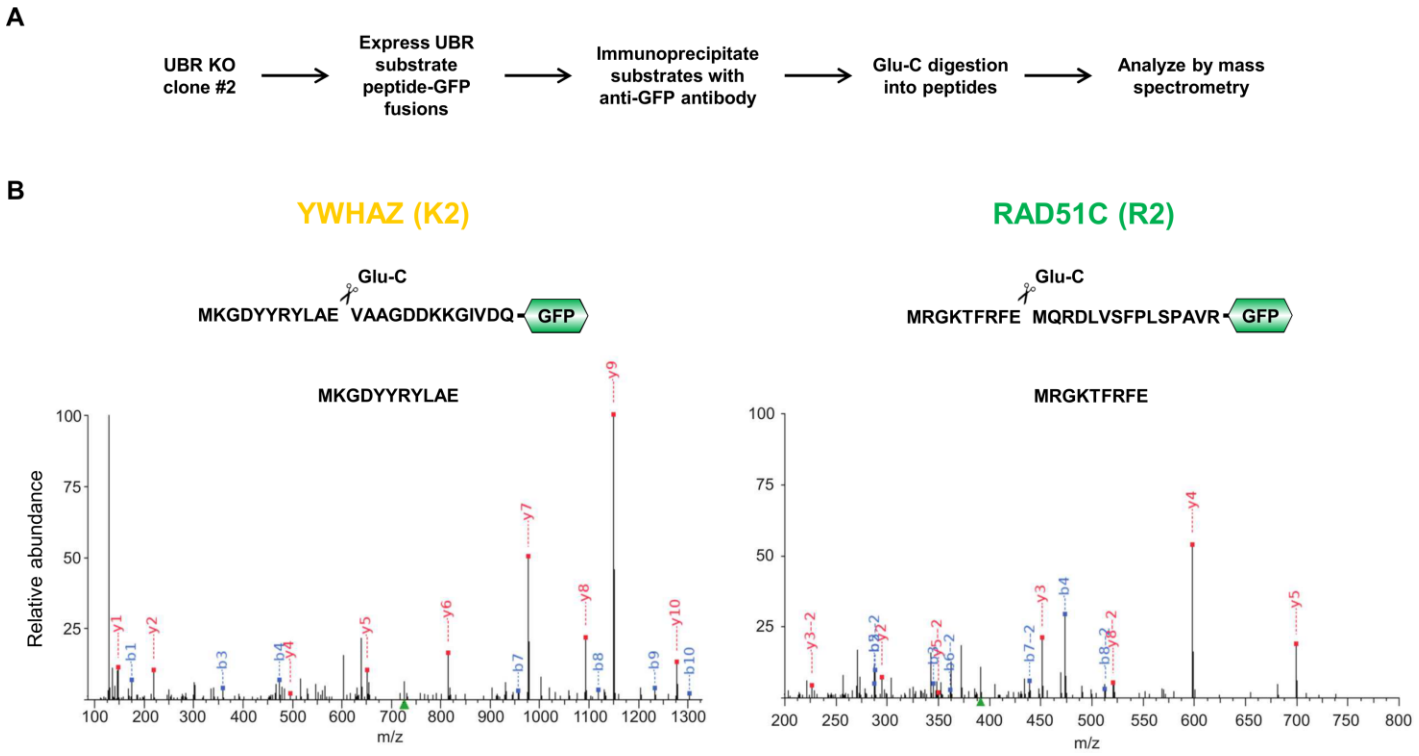

Fig. S3. UBR substrates commencing MK- and MR- retain their initiator methionine.

(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.

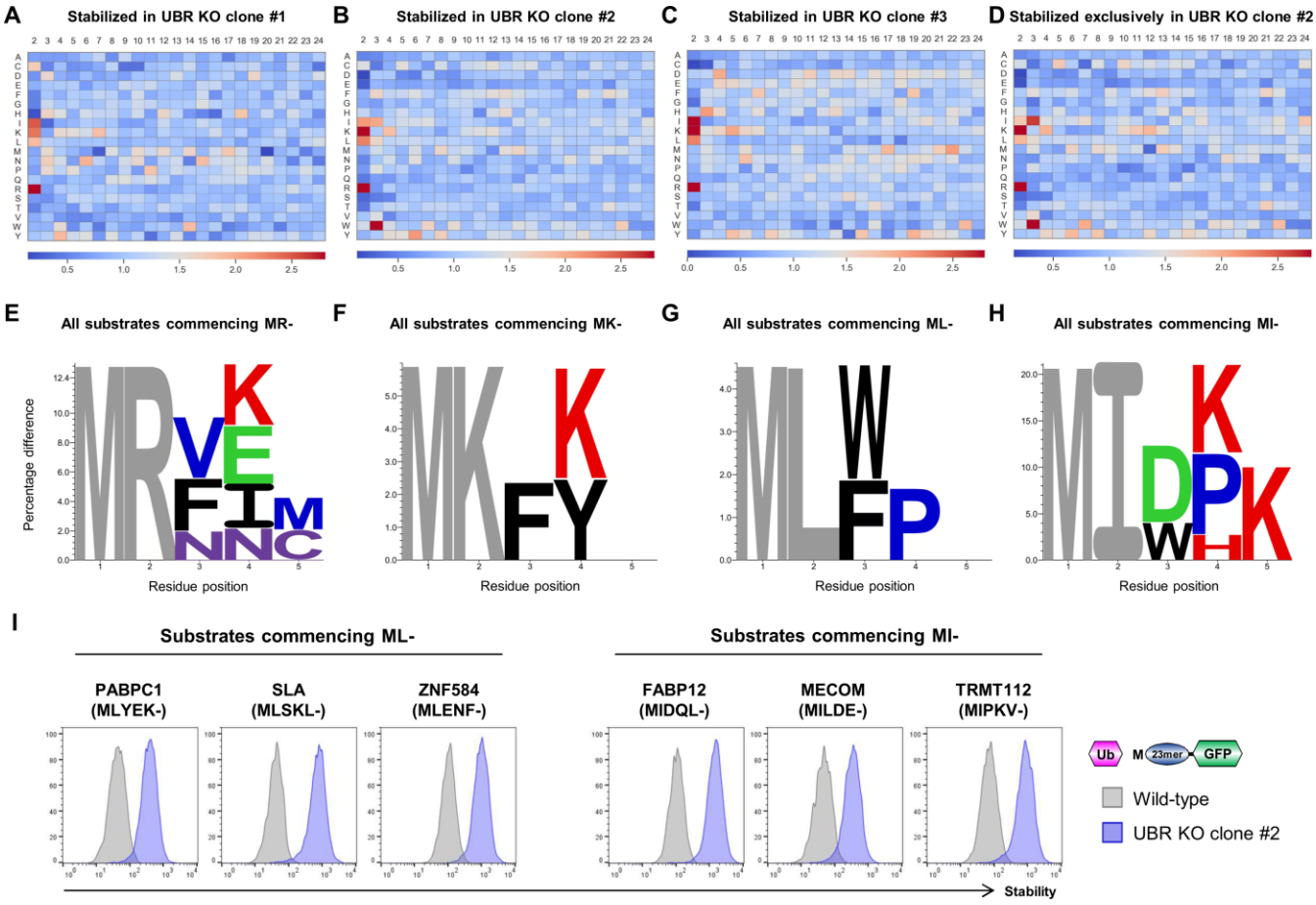

Fig. S4. Global identification of UBR substrates across the human N-terminome.

(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.

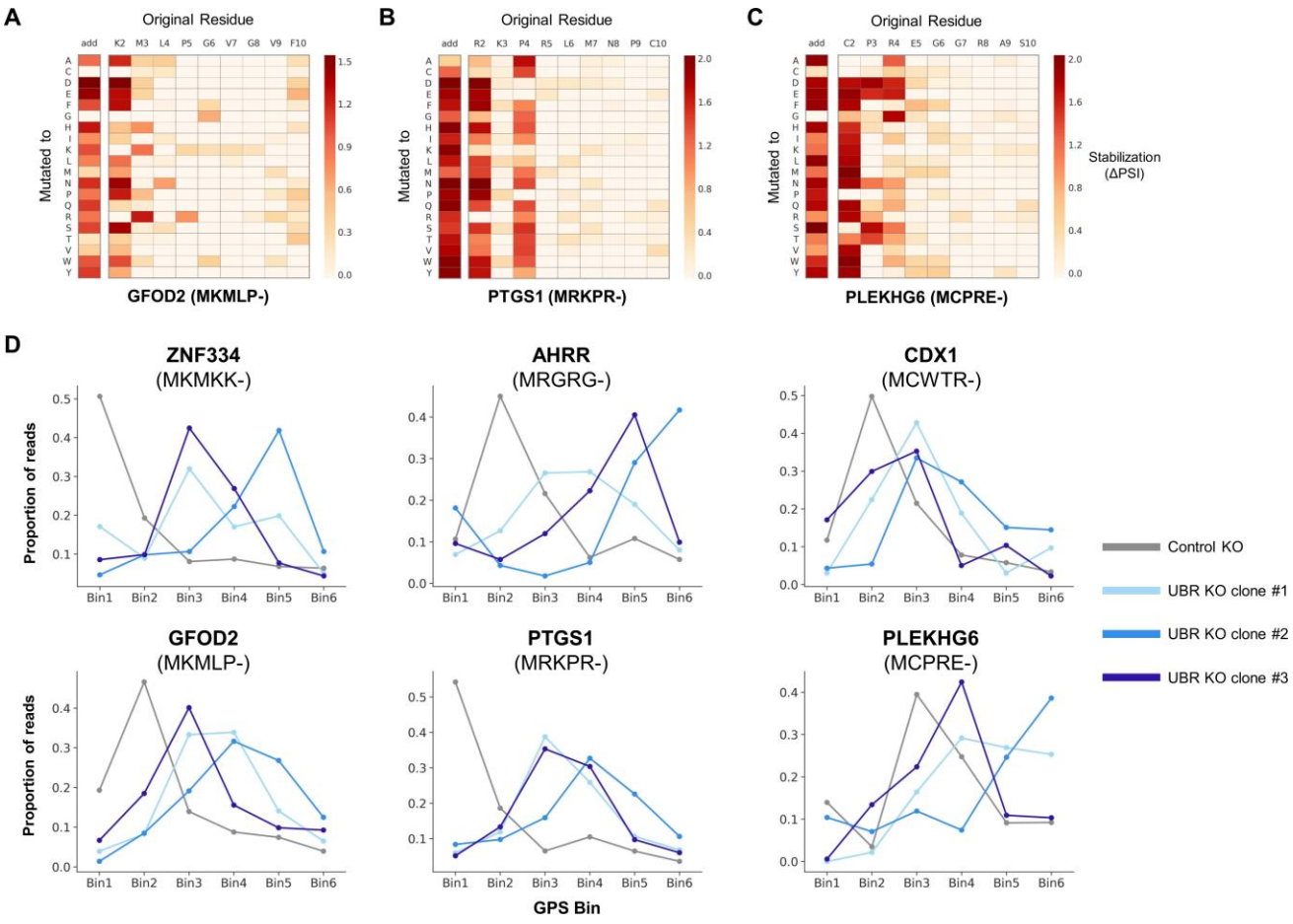

Fig. S5. Characterizing N-terminal UBR degrons through synthetic mutagenesis.

(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).

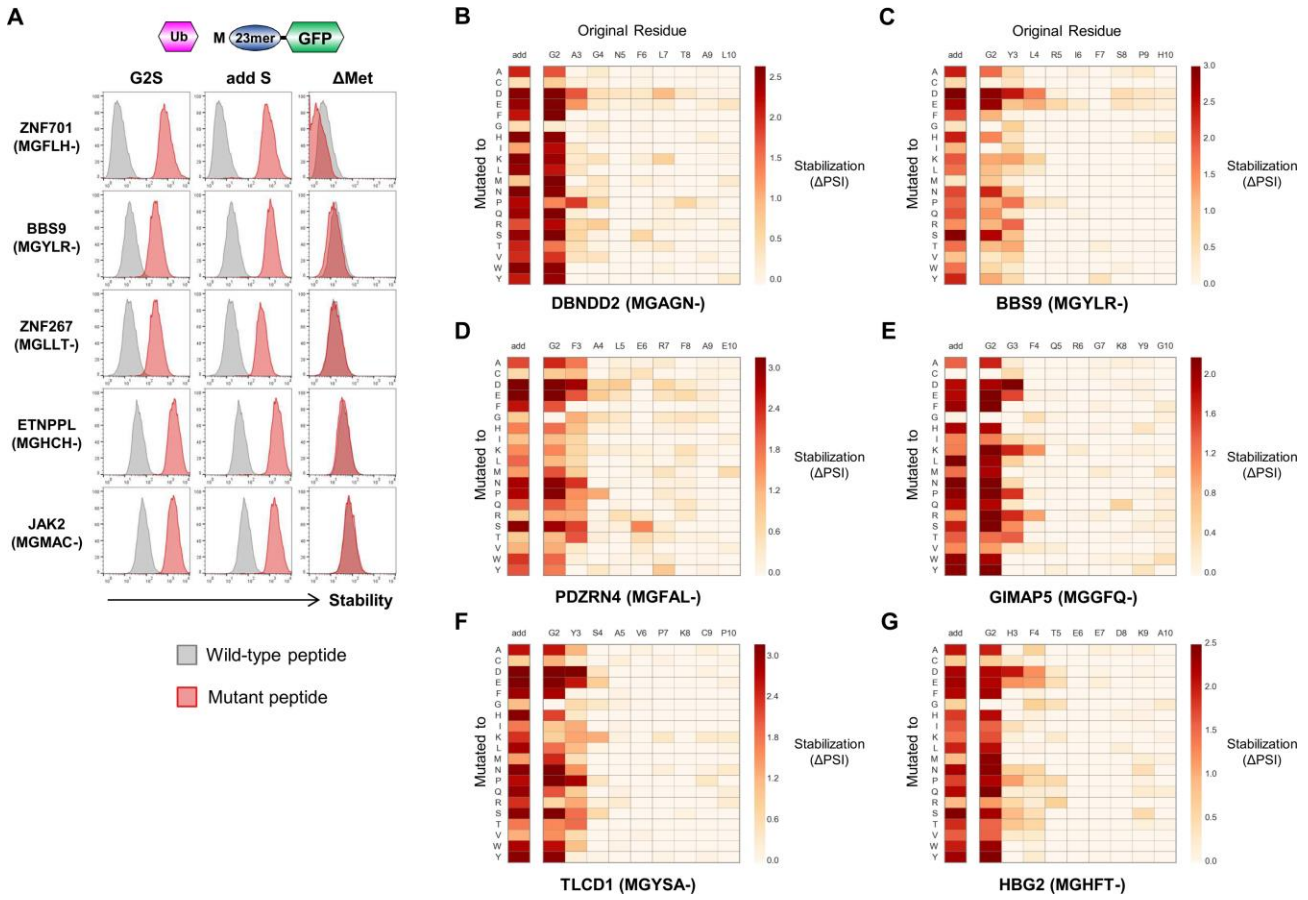

Fig. S6. Synthetic mutagenesis defines the composition of N-terminal glycine degrons.

(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.

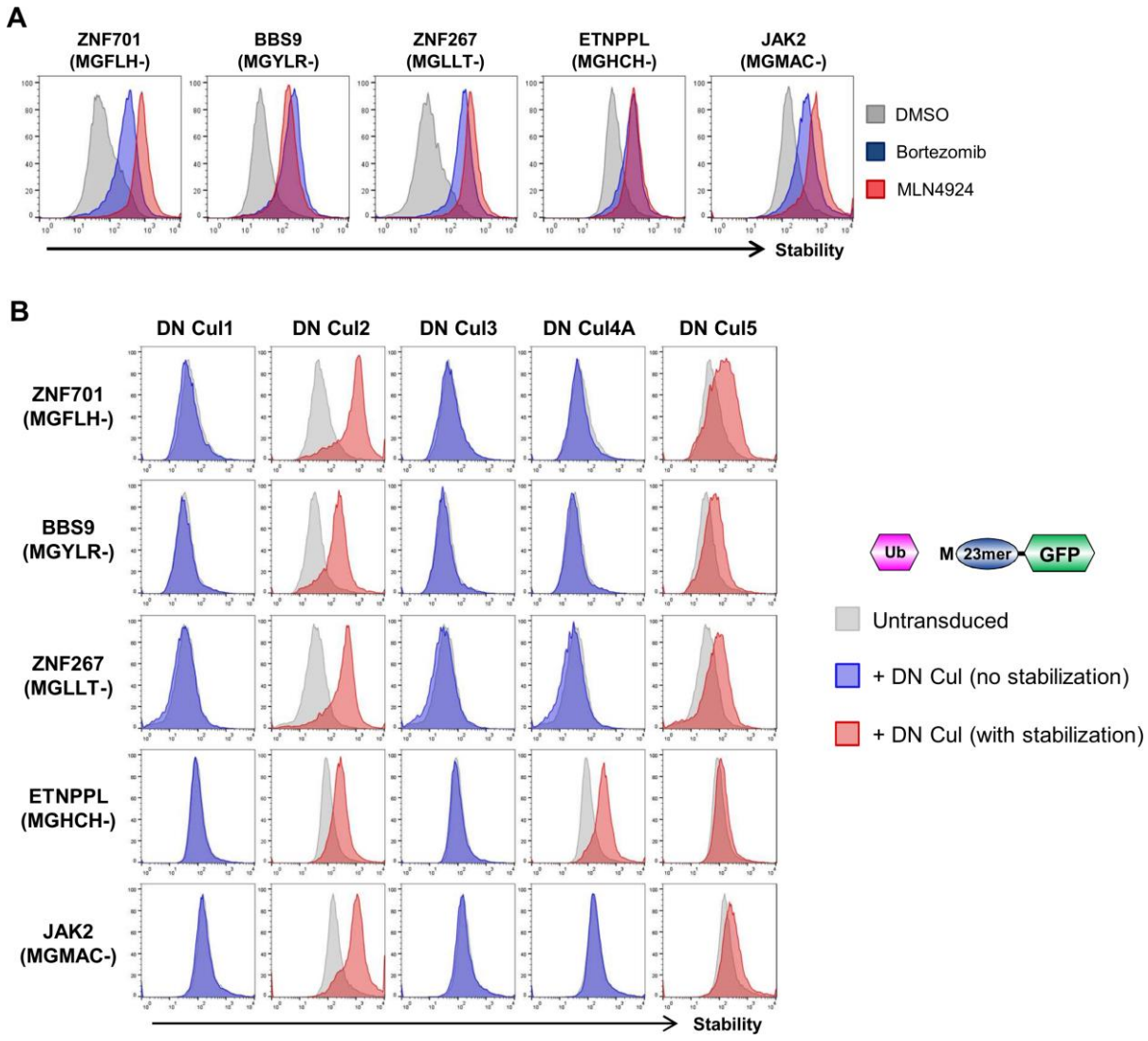

Fig. S7. CRL2 complexes target N-terminal glycine.

(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) 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.

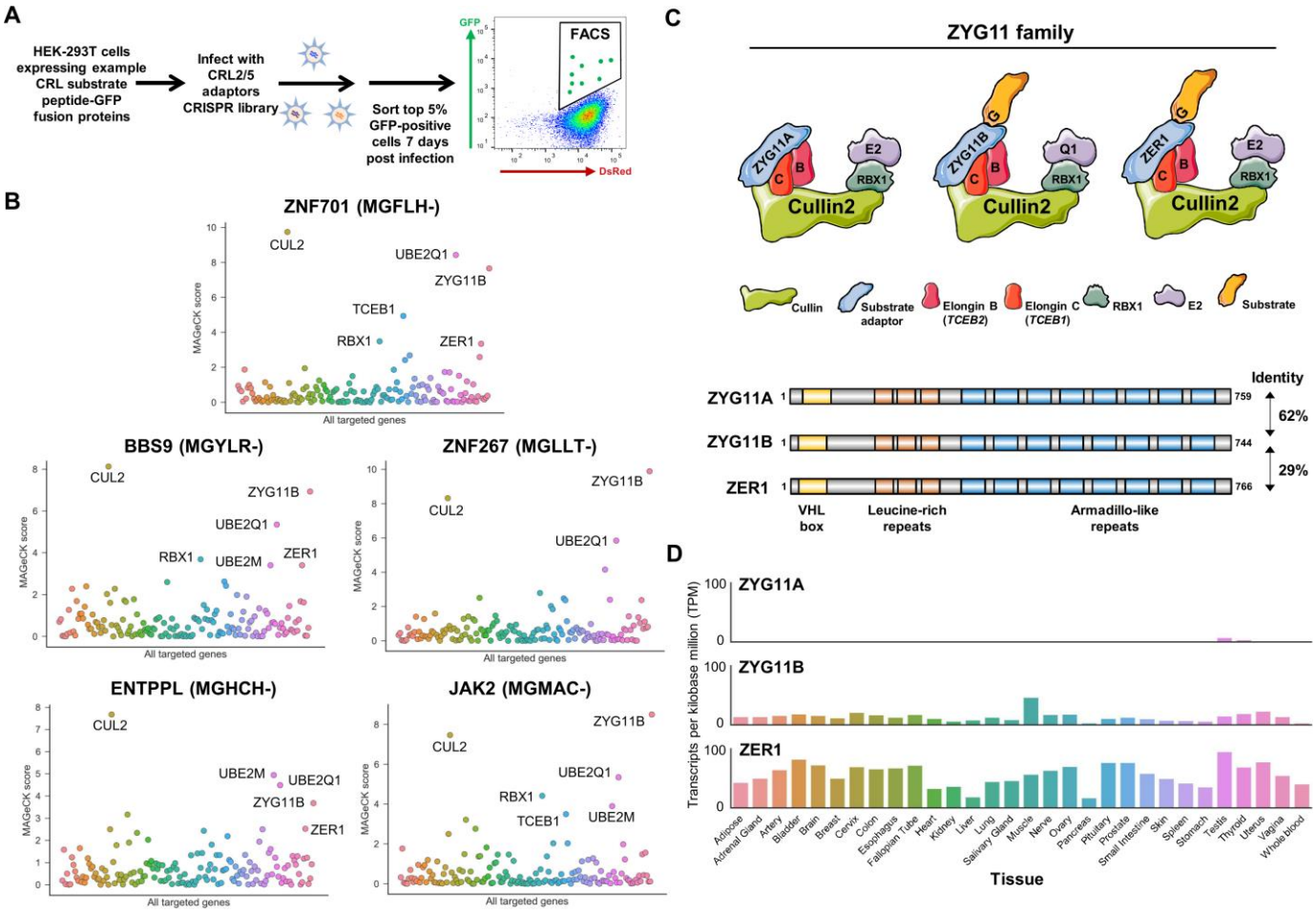

Fig. S8. The ZYG11 family of Cul2 adaptors target N-terminal glycine.

(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.

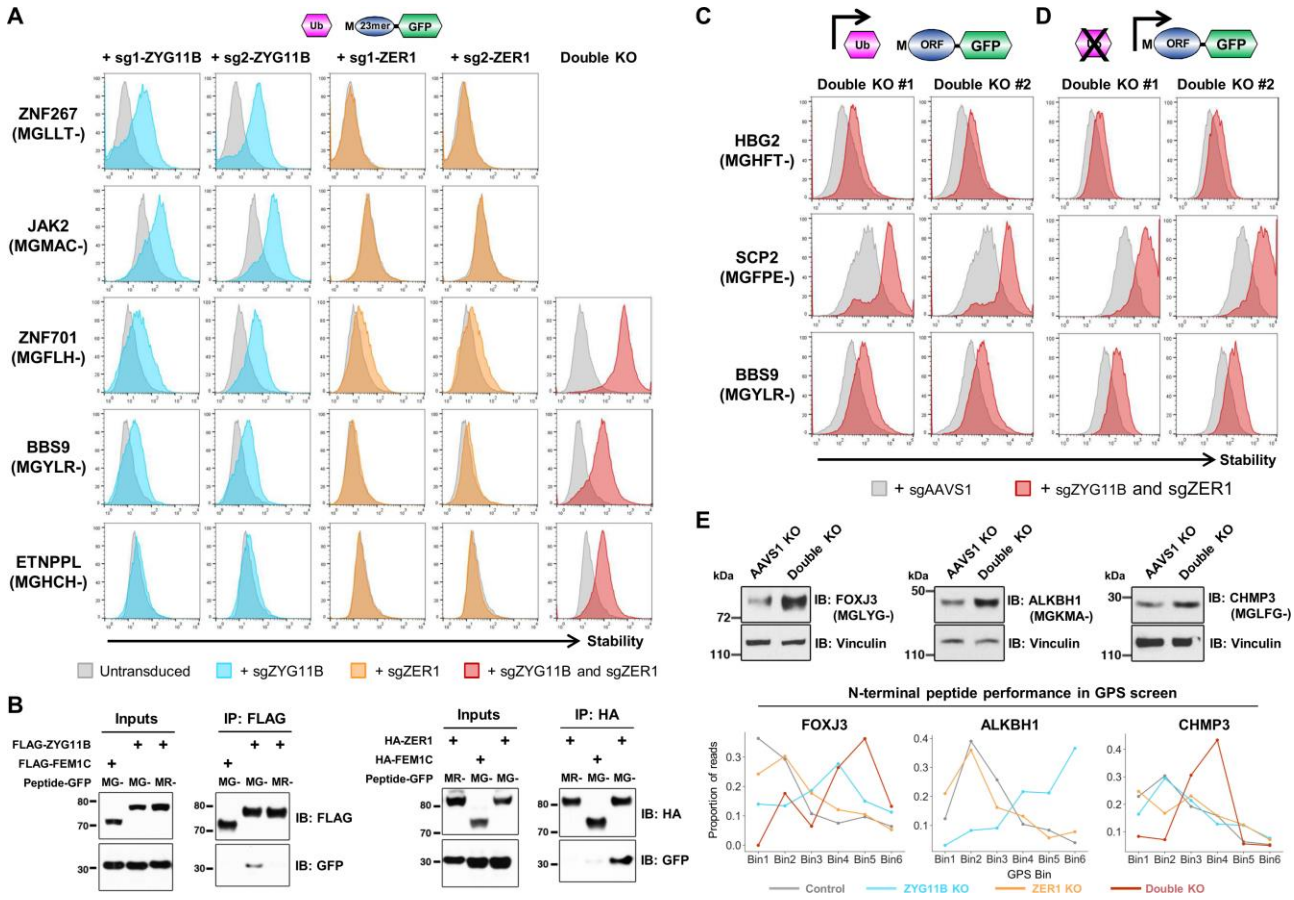

Fig. S9. Cul2^{ZYG11B} and Cul2^{ZER1} redundantly target N-terminal glycine degrons.

(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.

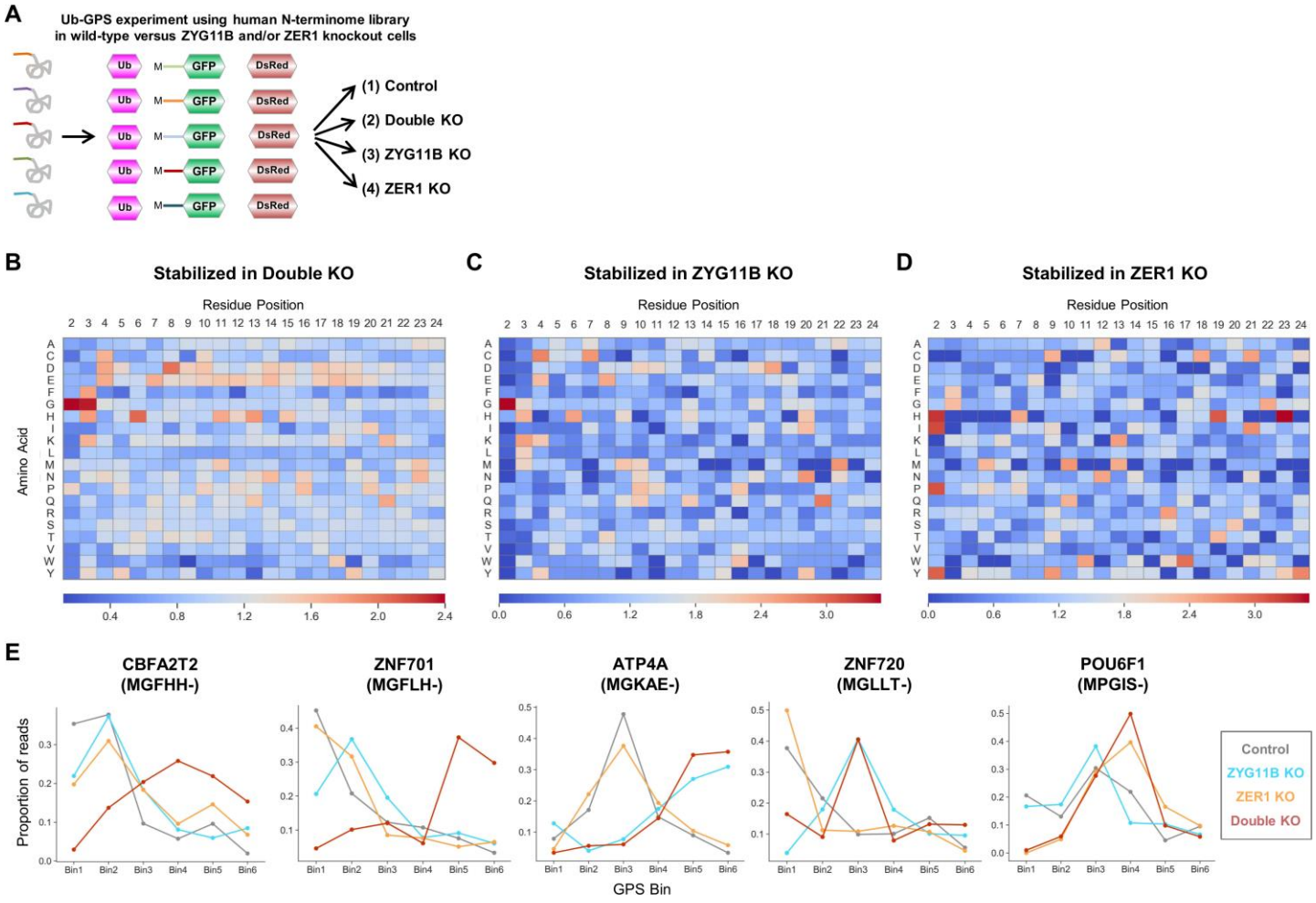

Fig. S10. Global identification of Cul2^{ZYG11B} and Cul2^{ZER1} substrates across the human N-terminome.

(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).

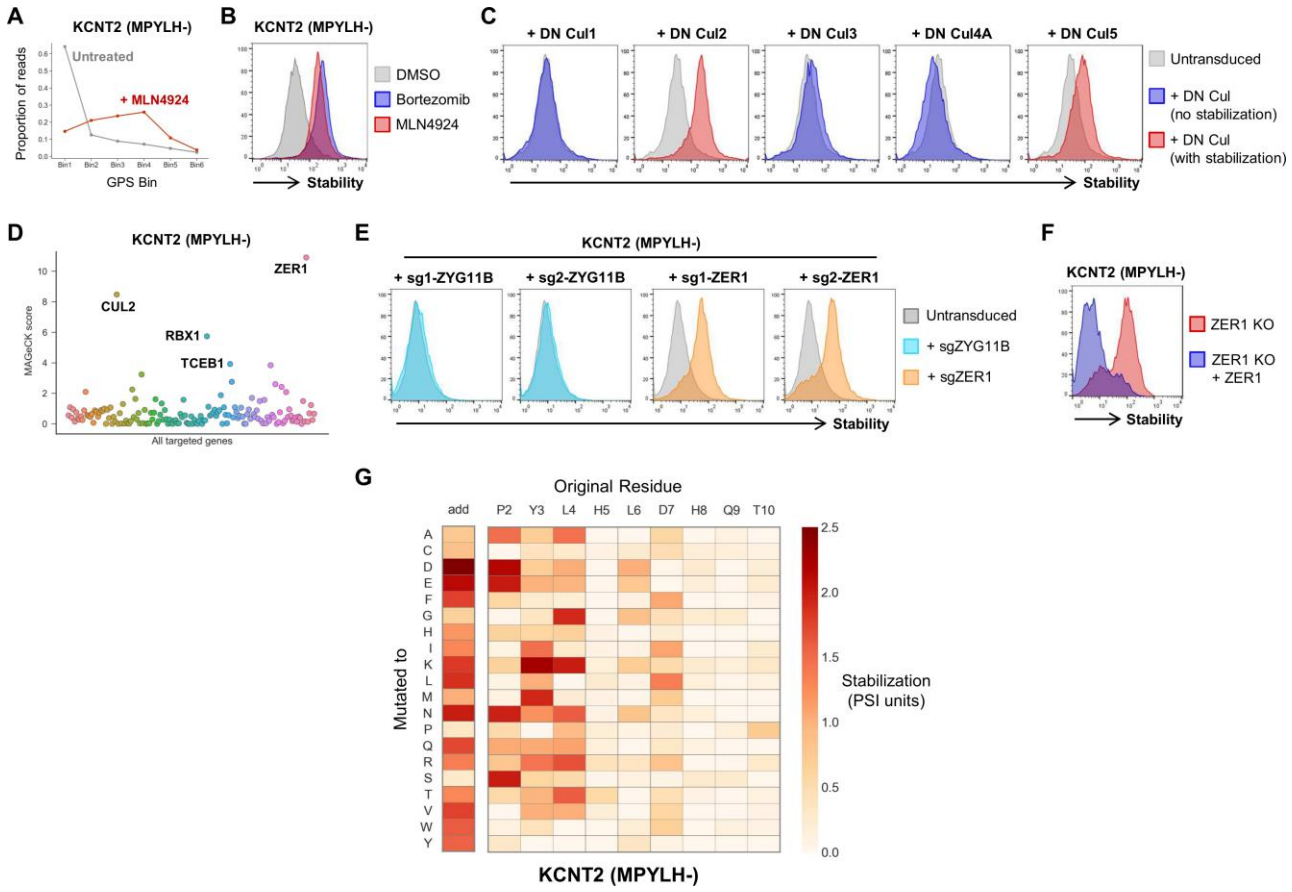

Fig. S11. An N-terminal glycine residue is not absolutely critical for substrate recognition by Cul2^{ZER1}.

(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.

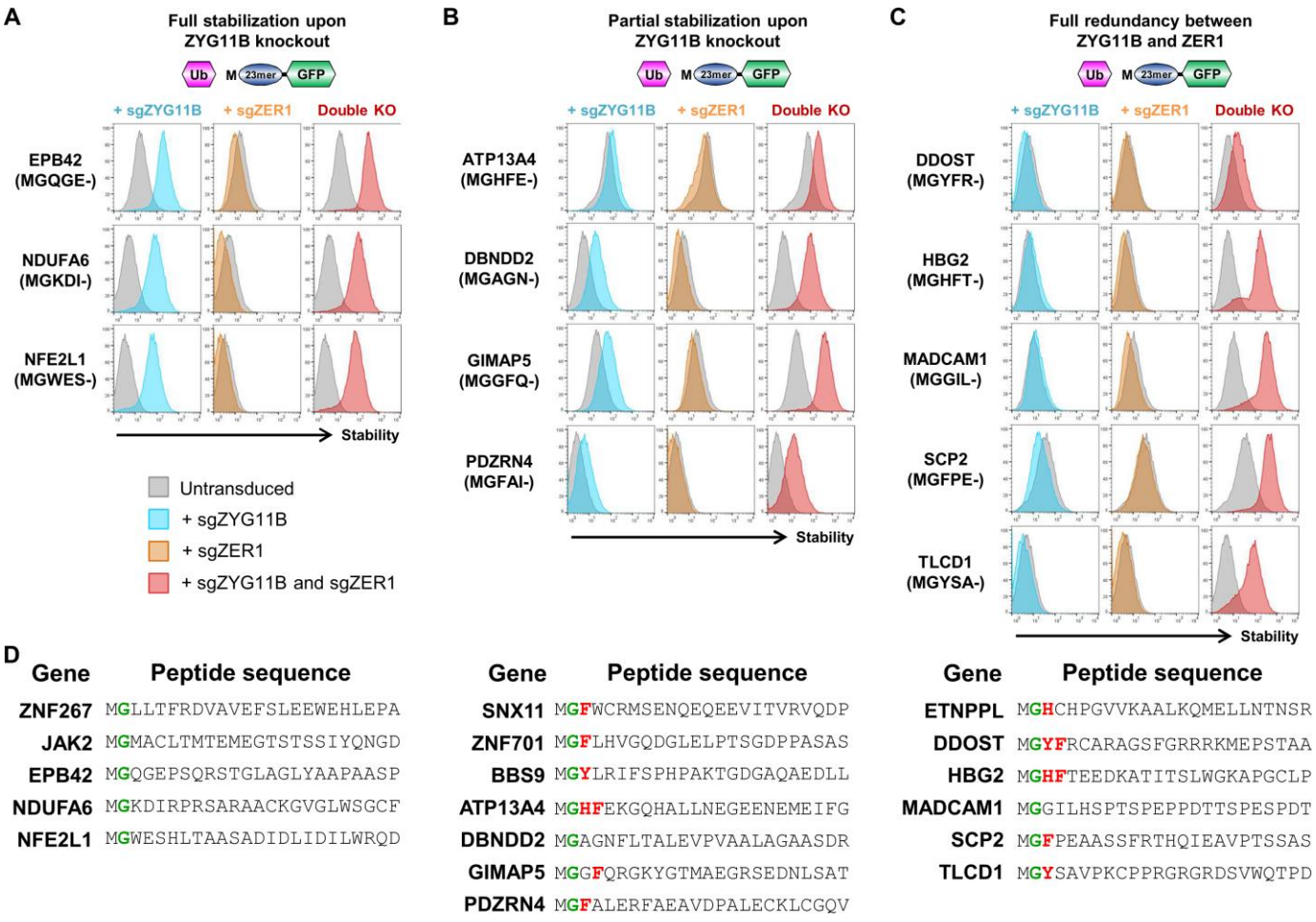

Fig. S12. Exploring the degron requirements for Cul2^{ZYG11B} versus Cul2^{ZER1}.

(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.

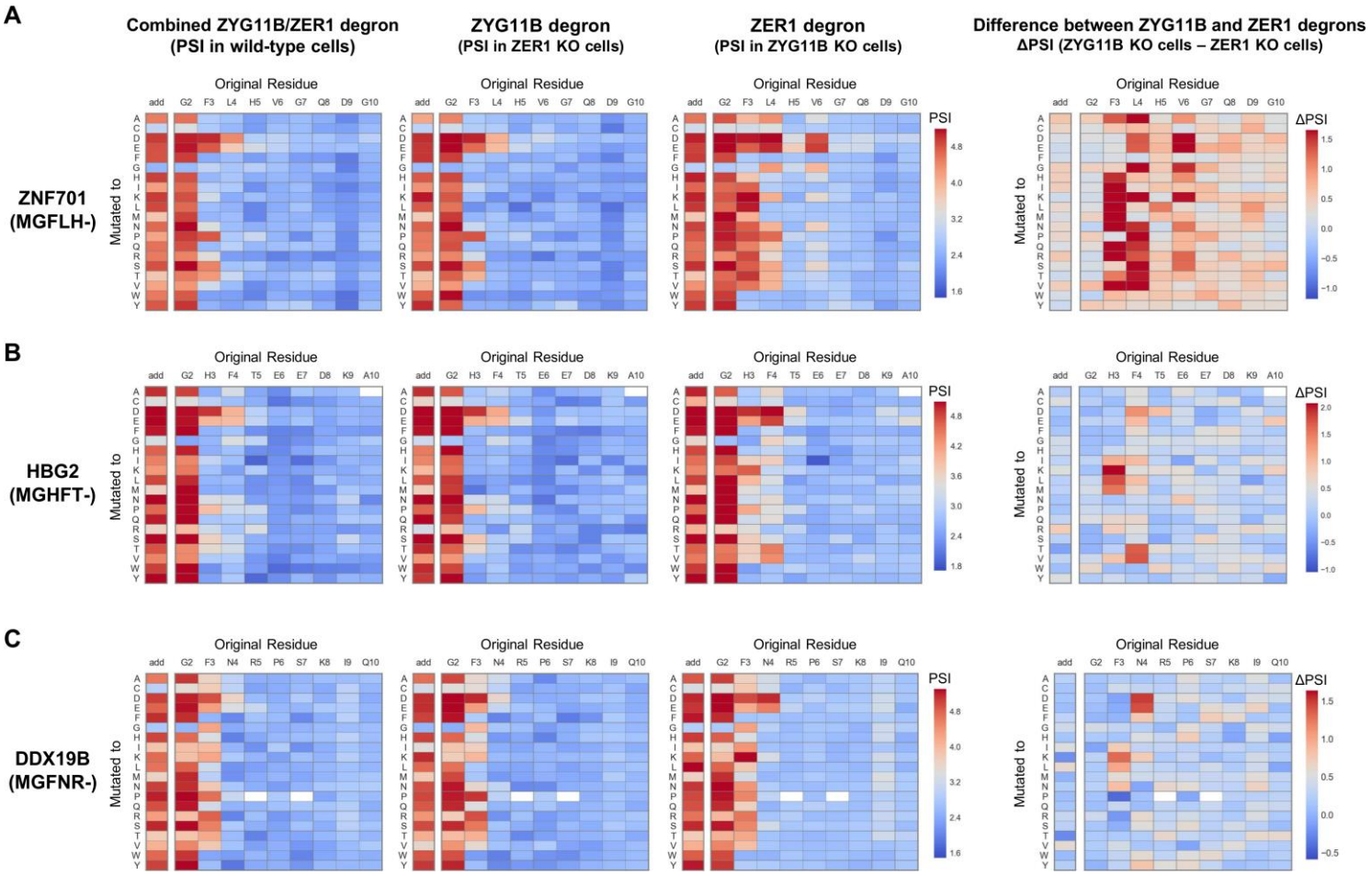

Fig. S13. Saturation mutagenesis defines the degrons recognized by Cul2^{ZYG11B} and Cul2^{ZER1}.

(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.

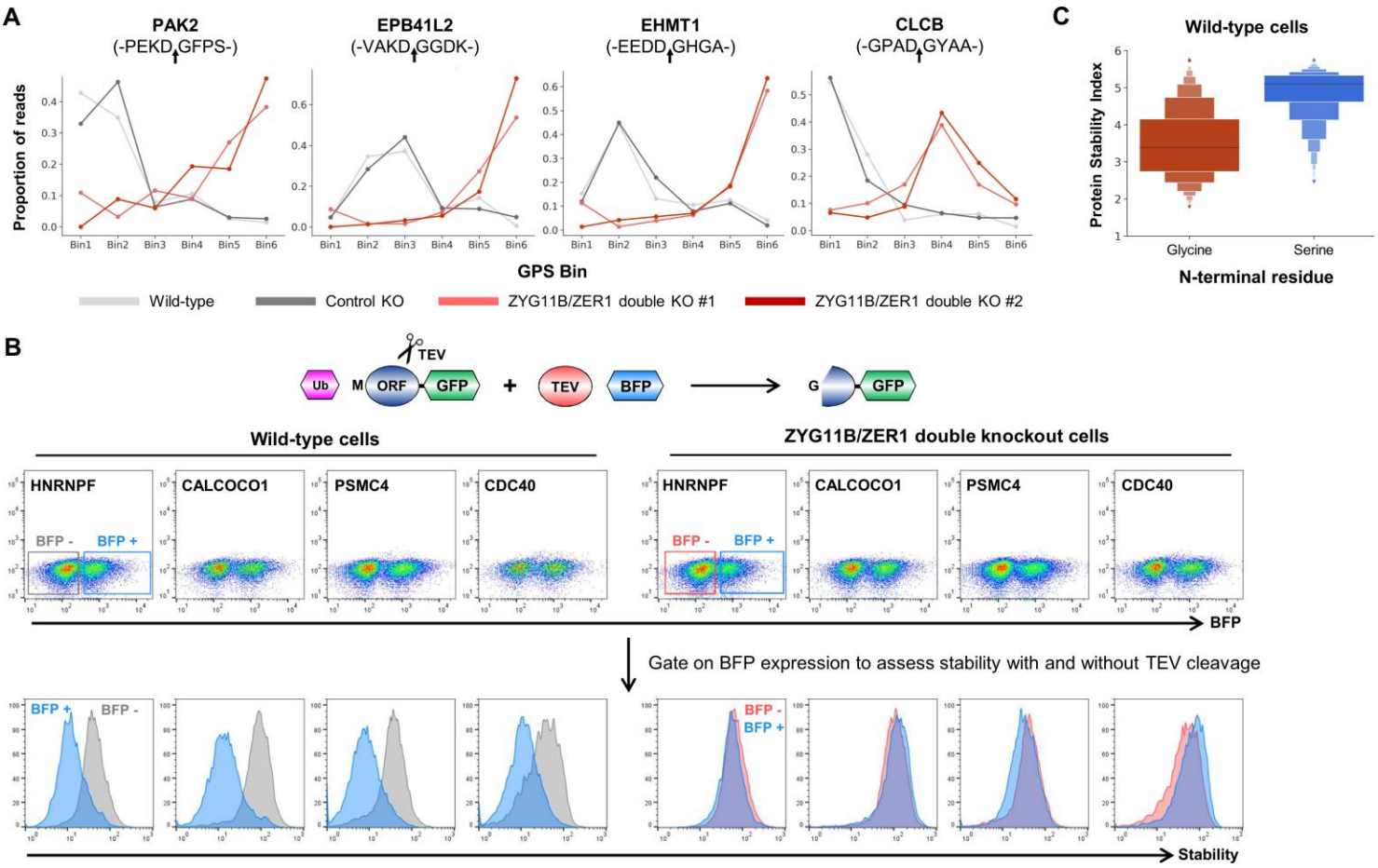

Fig. S14. Caspase cleavage generates N-terminal glycine degrons targeted by ZYG11B and ZER1.

(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.

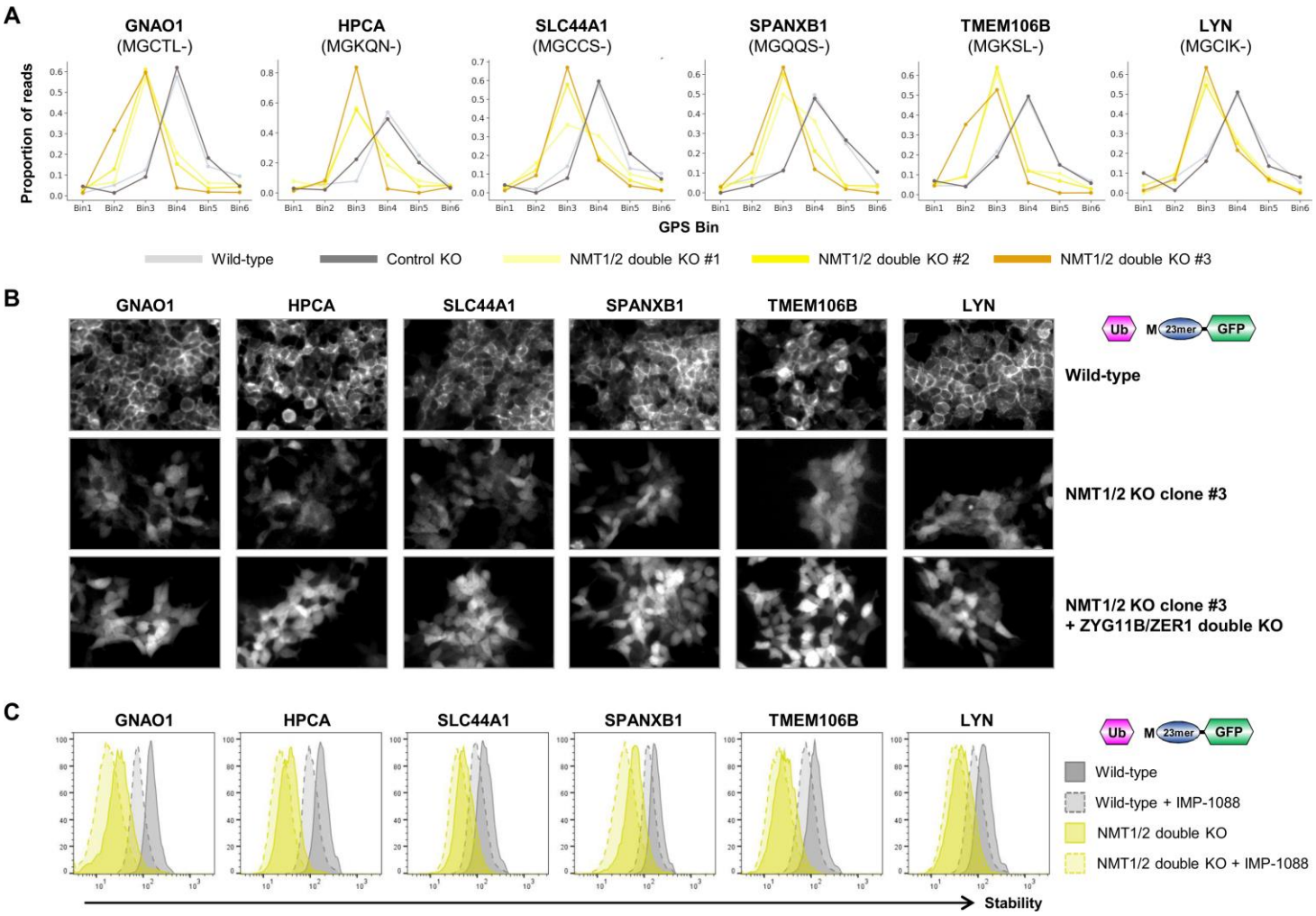

Fig. S15. A failure of N-myristoylation exposes N-terminal glycine degrons targeted by ZYG11B and ZER1.

(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)
CRISPR screen data.

Additional Data table S6 (separate file)
Caspase cleavage product GPS screen data.