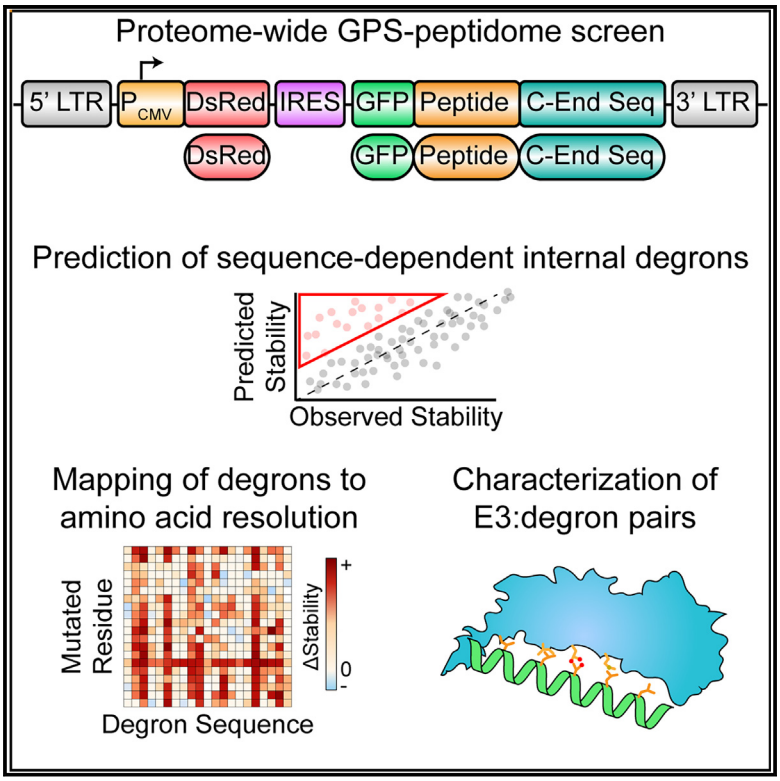
Molecular Cell

Resource

Elucidation of E3 ubiquitin ligase specificity through proteome-wide internal degron mapping

Graphical abstract



Highlights

- Proteome-wide global protein stability (GPS) assay identifies 15,800 degron peptides
- Critical degron residues were mapped by scanning and saturation mutagenesis
- CRISPR screening reveals cognate E3 ligases for 16 distinct degrons
- E3-degron pairs were docked by Alphafold2 and validated by co-immunoprecipitation

Authors

Zhiqian Zhang, Brandon Sie, Aiquan Chang, Yumei Leng, Christopher Nardone, Richard T. Timms, Stephen J. Elledge

Correspondence

selledge@genetics.med.harvard.edu

In brief

Zhang et al. combined a proteome-wide screen with machine learning to identify thousands of short linear motifs (degrons) in human proteins that are associated with sequence-dependent protein degradation. They went on to map critical residues, identify interacting ligases, and validate binding for a diverse collection of internal degrons.

Zhang et al., 2023, Molecular Cell *83*, 3377–3392
September 21, 2023
https://doi.org/10.1016/j.molcel.2023.08.022

CellPress

Molecular Cell

CellPress
OPEN ACCESS

Resource

Elucidation of E3 ubiquitin ligase specificity through proteome-wide internal degron mapping

Zhiqian Zhang,[1,2,4] Brandon Sie,[1,2,4] Aiquan Chang,[1,2] Yumei Leng,[1,2] Christopher Nardone,[1,2] Richard T. Timms,[1,2,3] and Stephen J. Elledge[1,2,5,*]

[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]Cambridge Institute of Therapeutic Immunology and Infectious Disease, Department of Medicine, University of Cambridge, Puddicombe Way, Cambridge CB2 0AW, UK
[4]These authors contributed equally
[5]Lead contact
*Correspondence: selledge@genetics.med.harvard.edu
https://doi.org/10.1016/j.molcel.2023.08.022

SUMMARY

The ubiquitin-proteasome system plays a critical role in biology by regulating protein degradation. Despite their importance, precise recognition specificity is known for a few of the 600 E3s. Here, we establish a two-pronged strategy for identifying and mapping critical residues of internal degrons on a proteome-scale in HEK-293T cells. We employ global protein stability profiling combined with machine learning to identify 15,800 peptides likely to contain sequence-dependent degrons. We combine this with scanning mutagenesis to define critical residues for over 5,000 predicted degrons. Focusing on Cullin-RING ligase degrons, we generated mutational fingerprints for 219 degrons and developed DegronID, a computational algorithm enabling the clustering of degron peptides with similar motifs. CRISPR analysis enabled the discovery of E3-degron pairs, of which we uncovered 16 pairs that revealed extensive degron variability and structural determinants. We provide the visualization of these data on the public DegronID data browser as a resource for future exploration.

INTRODUCTION

One of the primary mechanisms through which cells regulate signal transduction is through protein degradation. By selectively modulating protein degradation, cellular signaling pathways can quickly adapt to changing environmental conditions. Selective protein degradation is mainly mediated by proteases, selective autophagy, and the ubiquitin-proteasome system (UPS), a complex system involving approximately 600 E3 ubiquitin ligases in mammals that selectively interact and ubiquitinate their substrates.[1] It was estimated that at least 80% of protein degradation is mediated by the UPS.[2] UPS substrates play crucial roles in nearly all major cellular processes. Dysregulation of protein degradation is frequently associated with many pathological disorders, ranging from cancer and immune pathologies to neurodegeneration.[3]

In addition to responses to stimuli, constitutive regulation of protein stability is also important for quality control. Protein misfolding is a serious problem for which there are several dedicated pathways involved in quality control to direct misfolded proteins for degradation.[4] Likewise, cells need mechanisms that can recognize proteins that fail to assemble into proper complexes.[5,6] Other areas of quality control include addressing failures in the proper localization of proteins to subcellular organelles such as the mitochondria[7] or the endoplasmic reticulum (ER)[8] or other specific membranes[9,10] as seen for BAG6 recognition of mislocalized signal sequences.[8] Although some show regulation (e.g., the heat shock response), others act constitutively. Constitutively active E3 ligases may recognize masked degrons when proteins are misfolded or misassembled in complexes, act upon conformational change or post-translational modification of their substrate, or tie the abundance of a protein strictly to its mRNA level, which is regulated in response to some stimuli.

Why are so many different E3s needed for the UPS? Presumably, they have diverse modes of regulation and recognize distinct substrates.[11] However, since the discovery of the first degron in 1986,[12] the number of known degrons has remained sparse.[13] Furthermore, the motifs of many degrons were poorly characterized, making it challenging to predict E3 substrates using these sequences. We previously established the global protein stability (GPS)-peptidome technology and applied it to the analysis of C-degrons and N-degrons.[10,14,15] The simplicity and position of those degrons allowed them to be defined computationally. However, such an approach is unable to define more complex degrons, and therefore, a systematic attempt to characterize naturally existing internal degrons (position-independent degrons) is still lacking.

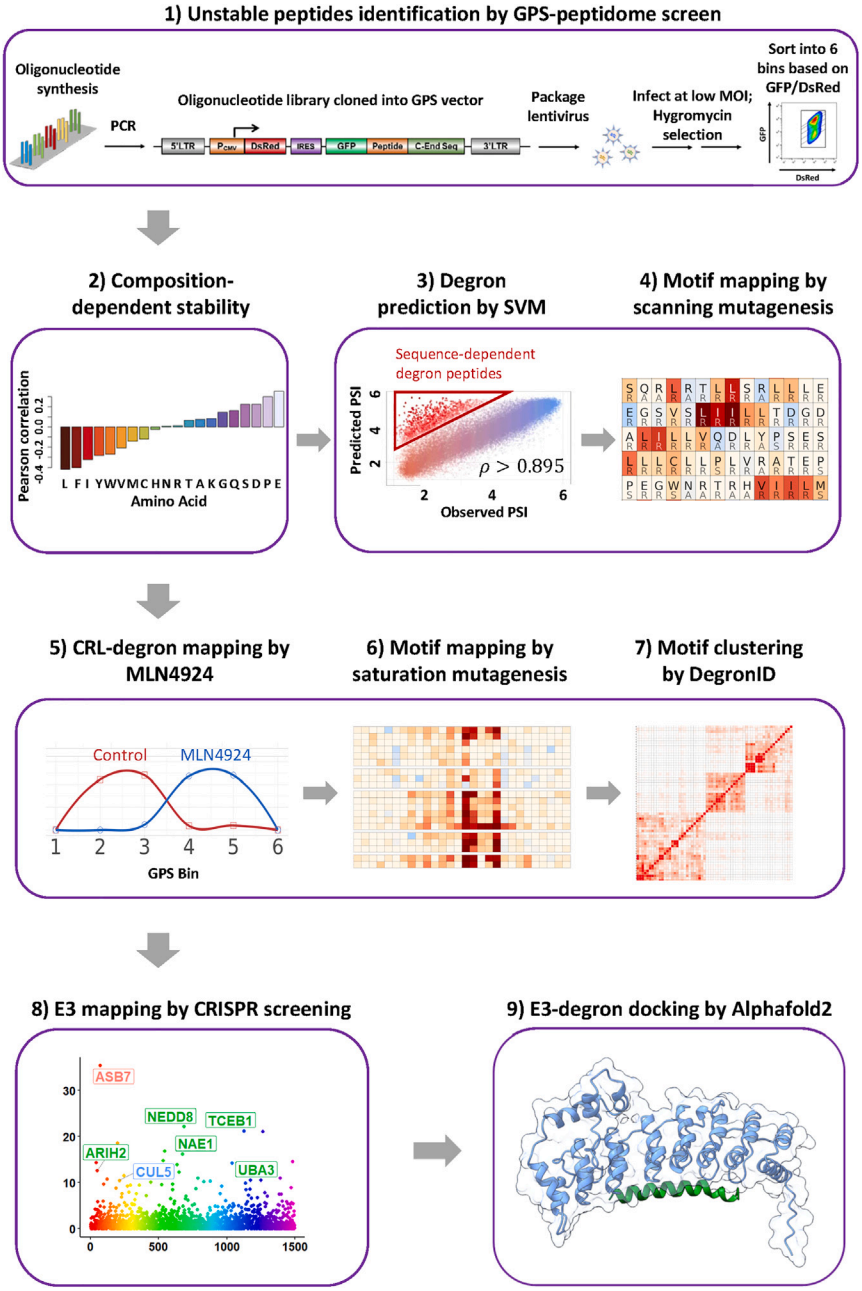


Figure 1. Schematic diagram illustrating the workflow of mapping E3-degron pairs

Overview of the study workflow.

RESULTS

GPS-peptidome screen for internal degrons

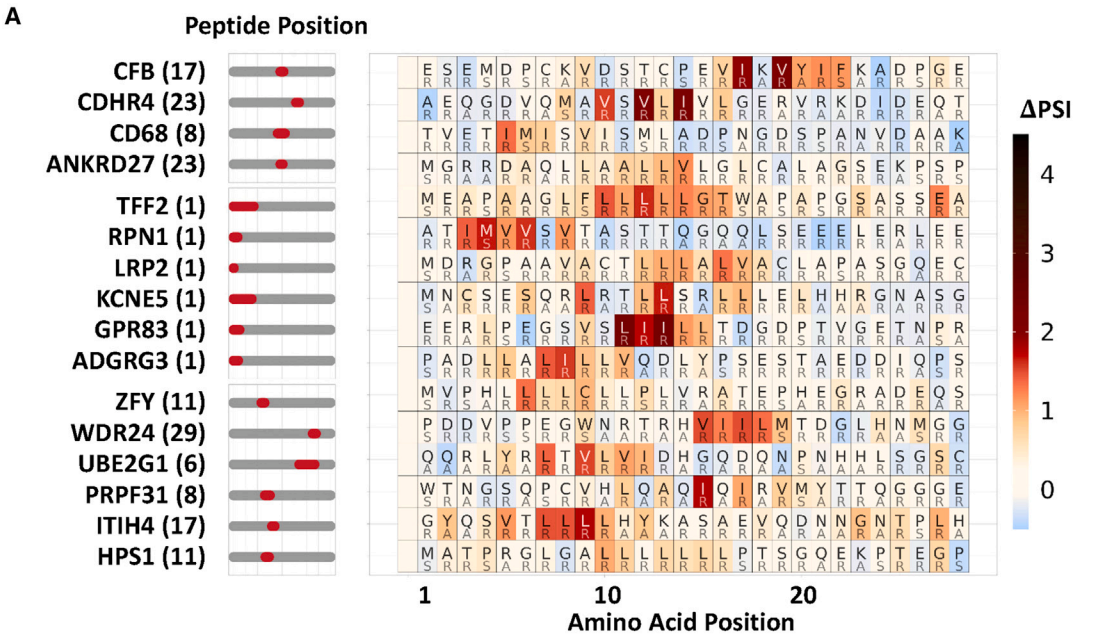
To develop a systematic method to identify internal degrons, we followed the strategy depicted in Figure 1. First, we designed[10,14] an oligonucleotide library encoding a peptide tiled across the entire human proteome. The library has ~470,000 peptides of 28 residues, overlapped by 5 residues. The peptidome library was cloned in frame with the C terminus of the GFP reporter. To avoid the unwanted artificial C-degrons when fused to the C terminus of GFP, a constant 20 amino acids lacking known C-degrons was positioned at the C terminus of the library (step 1 of Figure 1; STAR Methods).

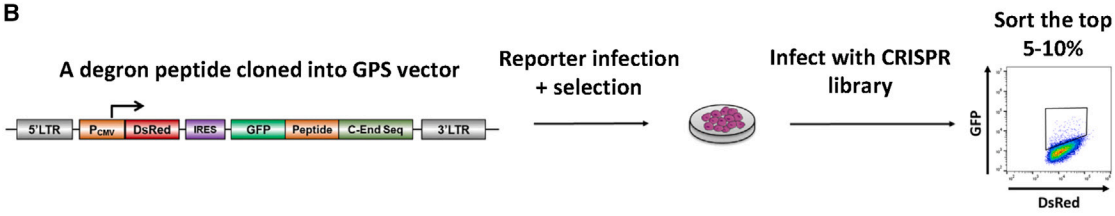
For authentic C-terminal degrons, we left in the natural stop codon to serve as previously studied controls.[14] The cloned GPS-peptidome library was packaged into lentivirus and used to infect HEK293T (American Type Culture Collection [ATCC] cat# CRL-3216) cells at a low multiplicity of infection (MOI) and selected using hygromycin resistance (step 1 of Figure 1). The resulting population was sorted into 6 bins by fluorescence-activated cell sorting (FACS) according to the stability of individual GFP-peptide fusions relative to the DsRed protein, followed by DNA sequencing to quantify the relative abundance of each GFP-peptide sequence in each bin. To determine the stability of each GFP-peptide fusion, we employed the protein stability index (PSI), which is calculated as a weighted average score of each peptide in the 6 bins, with PSI = 1 being the least stable, and PSI = 6 being the most stable[10,14] (Figure S1A). We focused our analysis on ~260,000 peptides (referred to as the 260,000 peptidome library) with sufficient reads for high-confidence analysis (Table S2; STAR Methods).

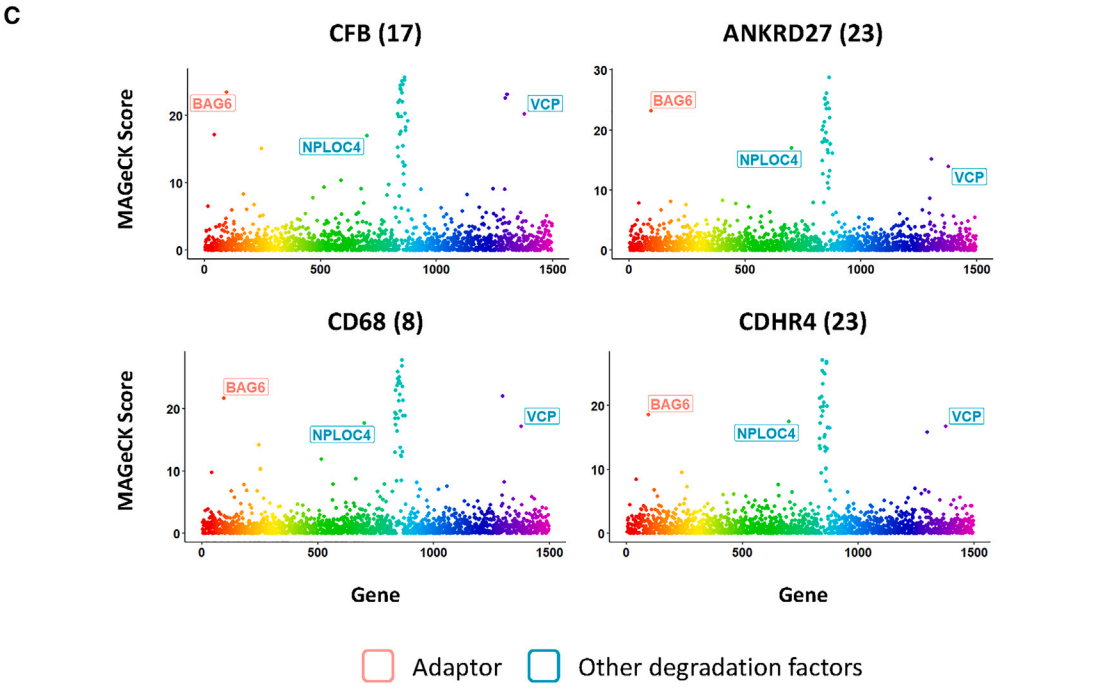
Correlation between amino acid composition and peptide stability

Peptide instability can be sequence-dependent, i.e., based on a defined sequential pattern of specific residues, or composition-dependent, i.e., based on the sum of the individual properties of amino acids, regardless of the order. We are primarily interested in the former class and sought to first identify the class of unstable peptides explained by composition-dependent mechanisms so as to minimize their contribution to our subsequent analysis. We observed a strong correlation between peptide amino acid composition and peptide stability (step 2 of Figure 1; Figure S1B). The number of amino acids with hydrophobic side chains, except alanine, negatively correlated with peptide stability. Notably, leucine showed a striking Pearson correlation coefficient of −0.42 with peptide PSI. The frequency of amino acids with acidic side chains (glutamic and aspartic acid) and proline positively correlated with PSI. Glutamic acid demonstrated the most significant correlation with a Pearson correlation coefficient of 0.35 (p value < 2.2e−16) followed by proline (r = 0.30, p value < 2.2e−16), aspartic acid (r = 0.23, p value < 2.2e−16),

CellPress
OPEN ACCESS







(legend on next page)

serine (r = 0.22, p value < 2.2e−16), glutamine (r = 0.16, p value < 2.2e−16), and glycine (r = 0.14, p value < 2.2e−16) (step 2 of Figure 1). Alanine, threonine, asparagine, arginine, histidine, and lysine were relatively neutral, whereas the remaining larger hydrophobics showed a negative correlation with stability. Strong correlations with instability indicate that particular amino acids play central roles in general peptide turnover, consistent with previous qualitative observations.[10,14]

Combining machine learning predictions and genetic analysis to identify sequence-dependent degrons

Next, we sought to use this information for the identification of complex degrons. We previously found neural networks could be used to predict sequence-dependent degron motifs in small datasets containing C-terminal degrons of 2 amino acids[16] but were unable to easily identify rarer and more complex degrons in larger datasets. We took a two-pronged approach to identifying sequence-specific degrons. First, we predicted peptides whose stabilities were composition-dependent. Second, we focused on those peptides whose stabilities could not be predicted by composition alone and thus were candidates for containing a sequence-dependent degron and carried out genetic analysis to identify sequences within them responsible for instability.

To predict composition-dependent peptides, we trained a support vector machine (SVM) model on 10% of the peptide data, using the counts of each amino acid within the peptide as features to predict PSI (Figure S1B; STAR Methods). We then used this trained model to predict PSI based on the composition of the remaining peptides. The predicted PSI and observed PSI showed a Pearson correlation coefficient of approximately 0.9, with more than 92% of peptides exhibiting less than 1 PSI unit of difference between predicted PSI and observed PSI (step 3 of Figure 1). The difference between the predicted PSI and the experimentally observed PSI is termed as degron index (DI). The larger the DI, the stronger the sequence-specific degron activity of the peptide. For example, C-end peptides with known C-degron motifs served as positive controls and showed significantly larger DIs compared with other C-end peptides (Figure S1C). We took ~15,800 peptides with DIs greater than +1 as our pool of peptides that potentially encode a large repertoire of degron peptides.

Many types of E3-regulated peptides should be present in the DI analysis, and we were particularly interested in those regulated by Cullin-RING ligases (CRLs). Thus, we also measured PSI for cells treated with MLN4924 to block CRL-mediated destruction of CRL-regulated peptides. We calculated the MLN4924 ΔPSI to determine which peptides were likely to be strong CRL substrates.

Next, we sought to identify the degron motifs within unstable peptides using a genetic approach. We generated a scanning mutagenesis library containing 283,880 oligonucleotides covering 9,817 peptides. This includes 1,782 peptides of ΔPSI > 1 for MLN2924 treatment for CRL substrates and 5,790 peptides of DI > 1 for non-CRL substrates, together with other peptides to facilitate even stability distribution of the library and effective sorting (Table S3; STAR Methods). To maximize the chance that each mutation causes a significant change in the amino acid's chemical properties, we employed an "opposite" scanning mutagenesis scheme where each amino acid was mutated to an amino acid with different chemical properties. "A, G, V, L, I, S, T, C, D, and E" were mutated to "R"; "M, W, F, and P" were mutated to "S"; and "Y, N, Q, K, R, and H" were mutated to "A" to generate a library of singly mutated peptides (STAR Methods). This GPS peptide library was screened as described for the original proteome-wide library to determine the PSI for each peptide.

As the CRL-degron peptides will be characterized later in greater detail below, we will momentarily focus here on the discussion of putative non-CRL degron peptides: degron peptides with large DIs but no significant stabilization by MLN4924 (MLN4924 ΔPSI < 0.2). Interestingly, we observed a pervasive motif commonly found in many non-CRL degron peptides (Figure 2A), often characterized by a series of hydrophobic residues, sometimes interrupted by one or two hydrophilic residues. For 4 representative degron peptides, we performed CRISPR screens to identify the E3 ligase(s) that target these degron peptides for degradation (Figure 2B; STAR Methods).

Strikingly, all 4 peptides were destabilized by BAG6 (Figure 2C). It had been previously established that BAG6 serves as a cytosolic protein chaperone that works with RNF126 to target proteins with an N-terminal signal peptide that fail to translocate into ER for destruction as a translocation quality control mechanism.[8,17] RNF126 did not score as a hit because its single guide RNAs (sgRNAs) dropped out during the screening process. It was proposed that BAG6 performs this function by recognizing the exposed signal peptide that should otherwise be buried upon proper translocation. However, the precise degron motif recognized by BAG6 was never resolved at the precise amino acid level. Signal peptides are often 16–30 amino acids in length and possess a tripartite structure with a hydrophobic core region.[18] Our results suggested that BAG6 recognizes a signal peptide through its hydrophobic core region, as many signal peptides were predicted to encode non-CRL degrons, and their scanning mutagenesis revealed BAG6 degron motifs within their hydrophobic regions (Figure 2A). Importantly, the motif recognized by BAG6 can be as short as 3–4 hydrophobic residues, as in CD68. As many proteins that do not have a signal peptide also encode a putative BAG6 motif that could be exposed upon misfolding (Figure 2A), our findings suggest that the traditional view that BAG6 specializes in translocation quality control of the secretory pathway might be incomplete.[4] Instead, BAG6 might work as a general quality control pathway to protect cells from

Figure 2. SVM machine learning-aided identification of a BAG6 degron motif

(A) Scanning mutagenesis of non-CRL-degron peptides. Peptides in the top group were subjected to CRISPR screens, and BAG6 was identified as a destabilizing gene. BAG6-like motifs were found not only in terminal peptides (middle group) deriving from N-terminal signalpeptide sequences but also in internal peptides (bottom group).

(B) Schematic diagram illustrating the workflow of the CRISPR screen designed to identify the genes required for the degradation of each degron peptide.

(C) CRISPR screens of 4 representative degron peptides with motifs described in (A) using a sgRNA lentiviral library of UPS-related genes identified BAG6 as a gene required for the degron activity.

stresses caused by aberrant protein folding and potential aggregation. As hydrophobicity is a major driver for protein folding, most protein misfolding will likely expose a short stretch of hydrophobic sequences that could be recognized by BAG6 for protein quality control through proteasomal degradation.

In addition to the BAG6 degron motif, our scanning mutagenesis also identified degron motif patterns in other putative non-CRL degron peptides (Figure S1D). Although this study primarily focused on characterizing CRL-degron peptides, we anticipate that future research that studies these degron peptides will identify a large set of non-CRL degron pathways.

GPS-peptidome screen with MLN4924 to identify degrons recognized by CRLs

We next focused on characterizing the CRL-regulated degrons. To identify CRL-regulated degron-containing peptides, we performed a GPS-peptidome screen in the presence of MLN4924 (Figures 3A and 3B; STAR Methods). A total of 4,245 peptides showed a MLN4924-ΔPSI score greater than 0.8. We selected a subset of 180 high-confidence peptides and cloned the DNA encoding them into lentiviral GPS reporters using one of the two forms of C-end sequences for downstream analyses (STAR Methods). A total of 101 validated peptides were validated and classified into different Cullin scaffolds using dominant-negative Cullins (Figures 3C and S1E; Table S5). We identified multiple peptides whose stability was regulated by five different Cullins. We observed that all peptides stabilized by DN CUL2 were stabilized equally by DN CUL5, but not all CUL5-regulated peptides were stabilized by DN CUL2, indicating a difference in inhibitory function between dominant-negative CUL2 and CUL5. Altogether, these confirmed that our approach sampled CRL degrons without biasing toward a specific degradation pathway.

DegronID enabled degron peptide clustering based on degron motif similarities

To systematically define degron motifs, we performed a GPS-peptidome screen with a saturation mutagenesis peptidome library containing 133,250 oligonucleotides covering 250 selected peptides stabilized by MLN4924, where every residue of each peptide was mutated to the other 19 amino acids (Table S6). We achieved high-resolution mapping of 219 degron footprints. Thirty-one peptides were missed either because of low representation or because they possessed multiple degrons that could not be easily disambiguated. The saturation mutagenesis revealed distinct degron motifs. However, we also noticed similarities among several degron motifs, raising the possibility that multiple peptides may be recognized by the same cognate E3 ligase.

We then developed a motif clustering algorithm, DegronID, to classify peptides containing similar motifs (step 7 of Figure 1). Briefly, for a degron footprint with saturation mutagenesis data, DegronID scores the human peptidome library for similarity to the degron footprint. DegronID then computes a similarity score between pairs of footprints and performs hierarchical clustering to group similar motifs (STAR Methods). The peptides from our reference database that were predicted by DegronID to be most similar to our finely mapped degron motifs tend to be less stable than our library as a whole (Figure 4A) and less stable than expected based on our composition-dependent SVM model (Figure S2B). Furthermore, they are more likely to contain more peptides than statistically expected that are stabilized by MLN4924 (Figure 4B). From DegronID's hierarchical clustering of degron footprints, we limit the dendrogram to 40 groups (DegronID group) for practicality and observe multiple meta-clusters with similarity among multiple adjacent groups (Figure 4C). We additionally annotated α-helix secondary structures using JPred4[19] and identified Cullin and E3 ligase adaptors identified from subsequent validation (Figure 4C). As expected, we find multiple peptide substrates with predicted motif similarity that share an E3 ligase. As computational validation of DegronID, we find that peptides with shared E3 tend to cluster together and score similarly to each other relative to unrelated peptides (Figures 4C and S2D).

We also validated that the DegronID-scoring algorithm chooses unstable peptides by examining sequences that most closely resemble other degrons that have been previously characterized. D-box motif degrons from the APC/C-degron repository[20] and C-terminal degron motifs containing either a GG* or RG* terminus from our previous study[14] also predict unstable peptides (Figure S2A; STAR Methods). Furthermore, we checked that DegronID can properly predict known degrons among the top hits for a particular ligase. For example, for the BTRCP-recognized degron motif, 2 known substrates, CDC25A and CDC25B, rank in the top 0.02% of predictions, ranking 40th and 41st, respectively. In addition, known substrates degraded by KLHL15 score among the top 0.2% for similarity to a strong FRY domain from GLB1L(2) that we characterized by saturation mutagenesis (Figure S2C, top left). Additionally, of the 52 proteins identified as interacting with KLHL15 in BioPlex 3.0, DegronID identifies 16 proteins (31%) containing an [FL]R[FY] motif at the sequence that most closely matches the GLB1L(2) FRYV motif, with 13 additional proteins containing a weaker version of the [FL]R[FY] degron. By contrast, only 8.8% of the proteins in our library contained the [FL]R[FY] motif (Figure S2C). By Fisher's exact test, the enrichment of [FL]R[FY] in KLHL15-interacting proteins compared with all human proteins in our library has a significance of $p < 6.7e{-}06$. This suggests that DegronID may be useful to prune high-throughput immunoprecipitation data for substrates that may be degraded by a particular E3.

To examine the possibility of using DegronID to characterize endogenous substrates of a particular E3, we chose six KLHL15-interacting proteins from BioPlex 3.0 that contained a FRY-like motif and exhibited sensitivity to MLN4924 in our peptidome screen. The full-length open reading frames (ORFs) for these proteins were cloned into the GPS 6.0 destination (DEST) vector and then stably expressed in either wild-type (WT) or KLHL15 knockout (KO) HEK293T cells. ZNF511 was strongly stabilized in KLHL15 KO cells, whereas GPS-ORF constructs containing other KLHL15-interacting proteins exhibited a more subtle increase in stability in KLHL15 KO cells, and control GPS-ORF constructs containing randomly selected proteins that do not contain a FRY-like motif and are not known to interact with KLHL15 show no difference in stability in WT or KLHL15 KO cells (Figure S5D).

CRISPR screen identified the cognate E3 ligases of CRL-degron peptides

DegronID identifies clusters of motifs that look similar by saturation mutagenesis (Figure S3). To identify the cognate E3 ligases

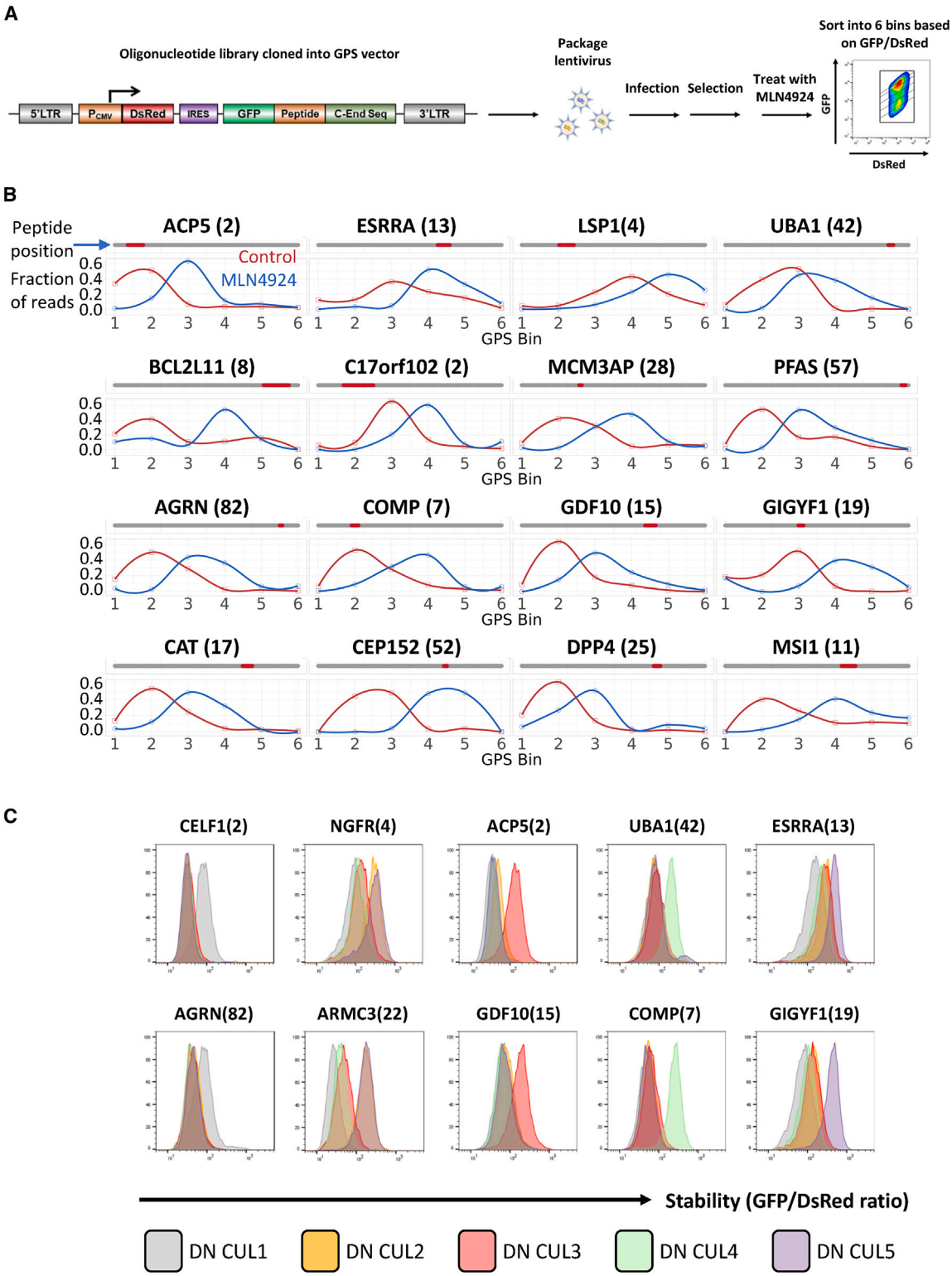
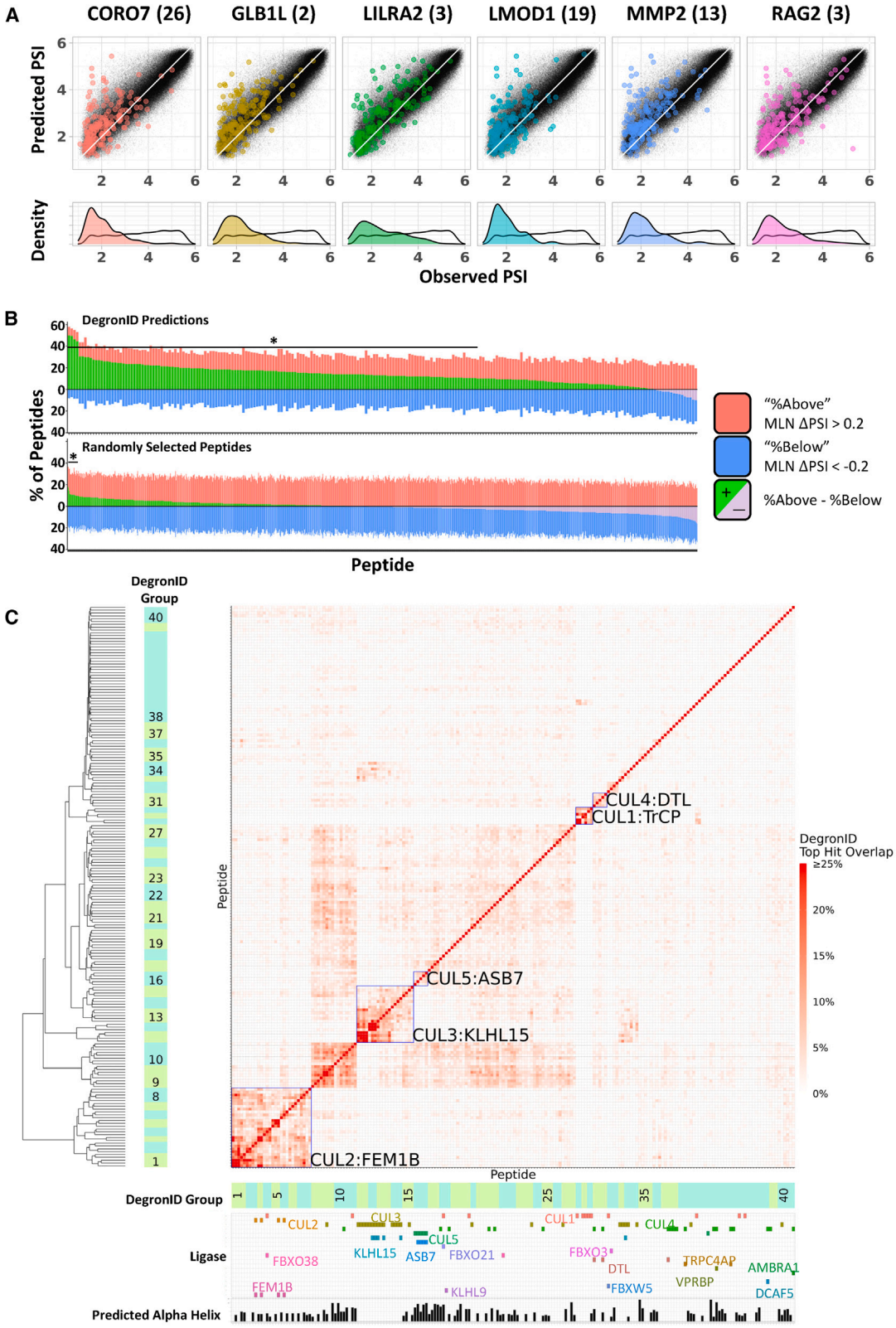


Figure 3. The GPS-peptidome screen using MLN4924 identified CRL-dependent degron peptides

(A) Schematic diagram illustrating the workflow of the GPS-peptidome screen in the presence of MLN4924.

(B) Representative peptides identified in the screen in (A) as responsive to MLN4924 are shown. For each peptide, their relative distribution across the 6 bins in the control condition was compared with that of the MLN4924-treated condition. See also Figure S1.

(C) Representative GPS measurements for peptide stability with dominant-negative (DN) CUL expression. Two peptides selectively stabilized by DN CUL1, DN CUL2/5, DN CUL3, DN CUL4A, and DN CUL5 are shown.



(legend on next page)

responsible for the degron activity of peptides with apparently distinct motifs, we performed CRISPR screens for degron peptides in different clusters. Fifteen CRL adaptors were identified to be responsible for the degron activity of the peptides, including $CUL1^{FBXO21}$, $CUL1^{FBXO3}$, $CUL1^{FBXO38}$, $CUL1^{FBXW5}$, $CUL1^{\beta\text{-}TrCP}$, $CUL2^{FEM1B}$, $CUL3^{KLHL9}$, $CUL3^{KLHL15}$, $CUL4^{DTL}$, $CUL4^{DCAF5}$, $CUL4^{TRPC4AP}$, $CUL4^{AMBRA1}$, $CUL4^{VPRBP}$, $CUL5^{ASB7}$, and $CUL5^{ASB3}$ (Figures 6 and S4; STAR Methods). One peptide was mapped to CUL1, but with no scoring CUL1 adaptor, raising the possibility that the sgRNAs targeting the corresponding adaptor dropped out during the screening process or were redundant. Subsequent examination of the degron motif showed that it matches a canonical β-TrCP degron motif[21,22] of which there are two paralogs, β-TrCP1 and β-TrCP2. Although the canonical β-TrCP degron motifs require one or two phosphoserines, our degrons have acidic residues that act as phosphomimetics, as seen in CDC25B. We also performed the CRISPR screen for peptide DLST(7), one of the 19 peptides for which we did not achieve a high-resolution mapping of its footprint, and mapped it to $CUL5^{ASB3}$. In addition to the substrate receptors, each of these screens identified commonly known components of CRL pathways, including ARIH1, ARIH2, NEDD8, and NAE1, and other core UPS components, including VCP and proteasome subunits.

Among the 15 CRL adaptors, $CUL1^{\beta\text{-}TrCP}$ and $CUL4^{DTL}$ have been studied extensively, and their cognate degron motifs have been accurately defined.[21,23] For $CUL2^{FEM1B}$, $CUL3^{KLHL15}$, and $CUL4^{TRPC4AP}$, although instances of their degron motifs have been identified, our results here include many new degrons that are distinct variants of the previously known degron motifs. Some new motif variants are related to previously established degron motifs, as in the case of $CUL3^{KLHL15}$, or could be completely unrelated sequences that bind to the same E3 ligase via a different binding site, as predicted to be the case of $CUL2^{FEM1B}$ (see below). To our knowledge, for the other 10 E3 ligases that we identified, including $CUL1^{FBXO21}$, $CUL1^{FBXO3}$, $CUL1^{FBXO38}$, $CUL1^{FBXW5}$, $CUL3^{KLHL9}$, $CUL4^{DCAF5}$, $CUL4^{AMBRA1}$, $CUL4^{VPRBP}$, $CUL5^{ASB7}$, and $CUL5^{ASB3}$, no instances of precisely defined degron motifs have been identified. Degrons detected by FBXO21 and VPRBP were studied previously but were not precisely defined.[24–26]

FEM1B degron

The largest degron meta-cluster (clusters 1–8) was characterized by a degron motif resembling WxxYL and more generally W[VAC]x[YRT][ILT] (Figure S3B). A CRISPR screen identified $CUL2^{FEM1B}$ as a destabilizing factor of DPP4(25), a representative peptide of this meta-cluster (Figures 5F and 6F). Several other members of meta-cluster sharing similar motifs were also tested and confirmed to be FEM1B substrates by CRISPR-mediated FEM1B KO (Figures 4C and S5C). This connection was unexpected because previous results from our group identified a R* C-end degron recognized by FEM1B. The presence of 2 structurally distinct classes of degrons recognized by the same E3 ligase suggests that the same E3 ligase could possess different degron-binding domains.[27,28]

KLHL15 degron

The second largest degron meta-cluster (clusters 11–15) was characterized by a degron motif resembling LRF (Figure S3C). CRISPR screen identified $CUL3^{KLHL15}$ as a destabilizing factor of the peptide SH3BP2(2) and a representative member of this cluster (Figures 5H and 6H). Several other members of cluster sharing similar motifs were also tested and confirmed to be KLHL15 substrates by CRISPR-mediated KLHL15 KO (Figure S5B). It was previously established that KLHL15 recognizes a tripeptide FRY degron, with 3 identified substrates containing the FRY degron to date.[29–31] However, our results here indicate that the actual motif is much more nuanced, as KLHL15 recognizes not only the motif but also other variants, notably FRF, LRF, and LRY (Figure S3C). Moreover, there are also residues adjacent to the core tripeptide motif that appear to be required in some contexts, pointing to flexibility in degron E3 recognition. These underscored how a precise understanding of degron motifs could enable more specific motif prediction while at the same time expanding the potential substrate repertoire recognized by the E3 ligase.

FBXO21 degron

CRISPR screening identified $CUL1^{FBXO21}$ as a regulator of IFNA8(5) stability. Interestingly, our saturation mutagenesis revealed that the degron-containing IFNA8(5) peptide, whose stability is regulated by FBXO21, has discontinuous stretches of motif residues (Figures 5A and 6A). The spacing was reminiscent of an amphipathic α-helix, and analysis by PROTEUS2[32] determined that it was highly likely to form an α-helical structure (Figure S6B). Additionally, although most substitutions were tolerated in non-essential residues within the motif region of IFNA8(5), mutating any residue within this region to proline, which is known to disrupt α helices, completely abolished the degron activity throughout the stretch of amino acids, supporting the need for an α-helix for degron activity (Figure 5A).

ASB7 degron

Peptides of cluster 16 are among several clusters that showed regularly spaced critical residues in their degrons (Figures 5N and S3E). CRISPR screening identified $CUL5^{ASB7}$ as a regulator of CCDC17(9) stability and a representative member of this cluster (Figure 5). Several additional peptides of this cluster sharing similar motifs, GIGYF1(19), LZTS1(16), and CEP152(52), were

Figure 4. Degron ID classified saturation mutagenesis motifs into clusters based on their sequence similarities

(A) Distribution of peptides in terms of PSI observed in GPS screen (x) and composition-based PSI prediction (y) from the 260,000 library (black) and the top 200 scoring hits by DegronID (color).

(B) Summary of top 200 hits by DegronID by MLN ΔPSI for (top) DegronID predictions for 198 CRL peptides; (bottom) iterations of random selections of 200 peptides from our 260,000 library. The bracket and asterisk indicate the instances for which the value of the green bar is greater than or equal to that which would be expected with an false discovery rate (FDR) of 0.1.

(C) Hierarchical clustering of saturation mutagenesis degron footprints. α-helical structure predictions and paired ligases from validation experiments are indicated below the clustergram. Meta-clusters that correspond to groups of degrons with multiple members sharing the same CRL are boxed and labeled (see Figure S3). See also Figure S2.

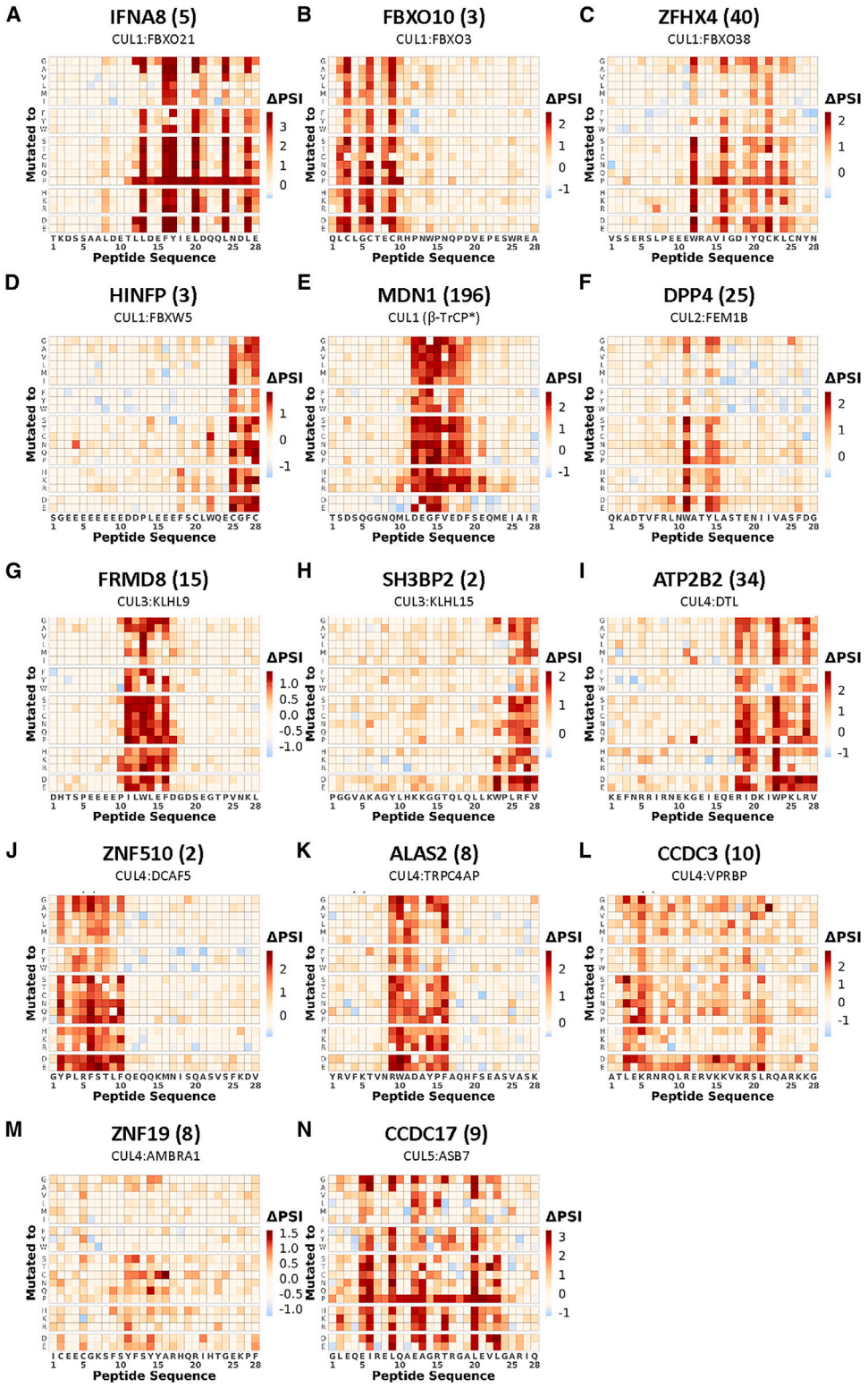


Figure 5. Saturation mutagenesis identified degron motifs in peptides stabilized by MLN4924

Saturation mutagenesis degron footprints for selected CRL-degron peptides: (A) IFNA8 (5), (B) FBXO10 (3), (C) ZFHX4 (40), (D) HINFP (3), (E) MDN1 (196), (F) DPP4 (25), (G) FRMD8 (15), (H) SH3BP2 (2), (I) ATP2B2 (34), (J) ZNF510 (2), (K) ALAS2 (8), (L) CCDC3 (10), (M) ZNF19 (8), and (N) CCDC17 (9). The cognate E3 identified from subsequent CRISPR screening is also indicated. See also Figures 6 and S3.

also tested and confirmed to be regulated by ASB7 (Figure S5A). These peptides are predicted to form an α-helical structure whose degron activity is clearly disrupted by prolines at any position within the predicted helical region (Figure S6). Interestingly, the 4 peptide substrates of $CUL5^{ASB7}$ motifs show substantial flexibility over what are clearly related extended motifs (Figure S3E). The typical ASB7-regulated degron is spread out over 16–19 amino acids with 6 or 7 critical amino acids all along the same surface of an α-helix. At each position, it appears that one or two amino acids can be important, although there are certain positions where pairs of amino acids seem to be dominant. Going from left to right, at position 1 (residue +1), there is an acidic residue, usually E, often followed by a hydrophobic residue such as L or I. At position 2, (residue +5) is frequently an L residue but can tolerate an I or M. Position 3 (residue +8) is another acidic acid residue, usually E, which can sometimes tolerate a T or a hydrophobic. Position 4 (residue +12) is often an L but is often flanked with charged residues. Positions 5 (residue +16) and 6 (residue +19) are often leucine that can tolerate I or M substitutions. As the surface interaction is extensive, suboptimal amino acids in one position might be compensated for by stronger interactions at one of the other positions to generate sufficient binding efficiency to interact with the E3 ligase. This makes a prediction about this degron complex.

AlphaFold2-assisted global docking

E3 ubiquitin ligases typically recognize key residues in the degron of their substrate, as has been shown by several crystal structures.[28,33–35] Given recent advances in structural prediction algorithms, we sought to explore predicted complexes between our degrons and their cognate E3 ligases. Thus, we leveraged the AlphaFold2-multimer algorithm[36,37] (STAR Methods). We found that docking for FEM1B, FBXO21, and ASB7 showed predicted interaction interfaces that were consistent with our degron saturation mutagenesis results.

The ASB7 adaptor is an ankyrin repeat protein that encodes seven ankyrin repeats, with each repeat forming a helix-turn-helix structure.[38] Multiple ankyrin repeats thus have the potential to recognize dispersed interacting residues on target proteins and could possibly recognize the distributed residues in the α-helix degron. Strikingly, docking of CCDC17(9) and LZTS1(16) onto ASB7 predicted an interaction surface that mapped directly onto the critical residues we previously determined by saturation mutagenesis, with each point of interaction with the critical residues on the degron peptide being recognized by a distinct ankyrin repeat (Figure 7A). The docking of these peptide-critical residues to ASB7 was stable across the top three ranked AlphaFold structure prediction models. The notion that dispersed residues may be able to align to E3 substrate adaptors that are composed of multiple repeats, as many CRL adaptors are, may be a general property that allows the evolution of many distinct substrate specificities.

Although many F-box protein adaptors contain repeats such as WD40 and LRR repeats, a substantial fraction lack such repeats and belong to the FBXO class. One of the successful docking predictions was for FBXO21. Motif residues found by saturation mutagenesis in peptide IFNA8(5) were predicted as interacting residues in the docking structures (Figure S6C).

We also applied AlphaFold2 to predict degron docking for FEM1B substrates and found that the prediction successfully identified the critical W residue in the FEM1B motif as crucial for FEM1B binding (Figure S7A). The structure of FEM1B binding to a R* degron peptide was solved earlier. However, docking of internal FEM1B degron peptides onto FEM1B showed that they bind to FEM1B through a binding region distinct from that of R*. We postulate that the presence of 2 degron-binding sites on the same E3 ligase could enable selective degradation regulation through binding distinct surfaces, thereby allowing this E3 ligase to be very flexible in substrate recognition.[27,28]

Characterization of E3-degron binding

Based on AlphaFold2 docking of E3-degron multimers, we sought to test whether mutagenesis of putative critical residues would affect E3-degron binding. For ASB7, FEM1B, and FBXO21, we performed co-immunoprecipitation (coIP) experiments to measure binding in HEK293T cells stably expressing GFP-peptide degrons and transiently transfected with N-FLAG-tagged E3 ligase. We tested WT E3 and also E3 and degron mutants, as shown (Figures 7B, S6D, and S7B). We observed effective immunoprecipitation of WT E3 with paired degron peptide and that mutagenesis of certain E3 and degron mutants that we identified as potentially critical for binding based on AlphaFold2 docking structures disrupted the ability of the E3 to coIP the degron peptide with the exception of ASB7 H126A. We also tested the effect of E3 mutants on degron stability by GPS, and we found that the E3 mutants that disrupt binding by coIP also disrupt the efficacy of the ligase to destabilize the degron peptide by flow (Figures 7C and S7C).

Measurement of degron peptide half-life

We also tested the effect of ASB7, FEM1B, and FBXO21 mutants on the stability of their cognate degron peptides by cycloheximide chase (Figure S6A). In all three cases, we find that the presence of E3 ligase results in a shorter half-life of the corresponding GPS reporter peptide.

DegronID data browser

To allow easy access to this large mutagenesis dataset, we created a web-hosted application to share data visualizations. The website provides visualizations of the effect of Cullin inhibitor MLN4924 on stability measured by GPS (Figure 3B) and degron footprints from scanning and saturation mutagenesis (Figures 2A and 5). Additionally, the website incorporates the degron clustering result of DegronID and enables the user to easily explore groups of related degron footprints. The application is freely accessible at https://elledge.hms.harvard.edu/?page_id=2960.

DISCUSSION

Systematic elucidation of the cellular degradome is critical to understanding the protein stability landscape and its role in shaping both homeostatic and regulatory biology. Here, we identify peptides with the potential to be degrons in proteins, dependent on their accessibility within a protein, which can be affected by allosteric actions, complex formation, or misfolding. The vast majority of the peptides in this library are likely recognized by general quality control mechanisms responsive to disordered sequences with

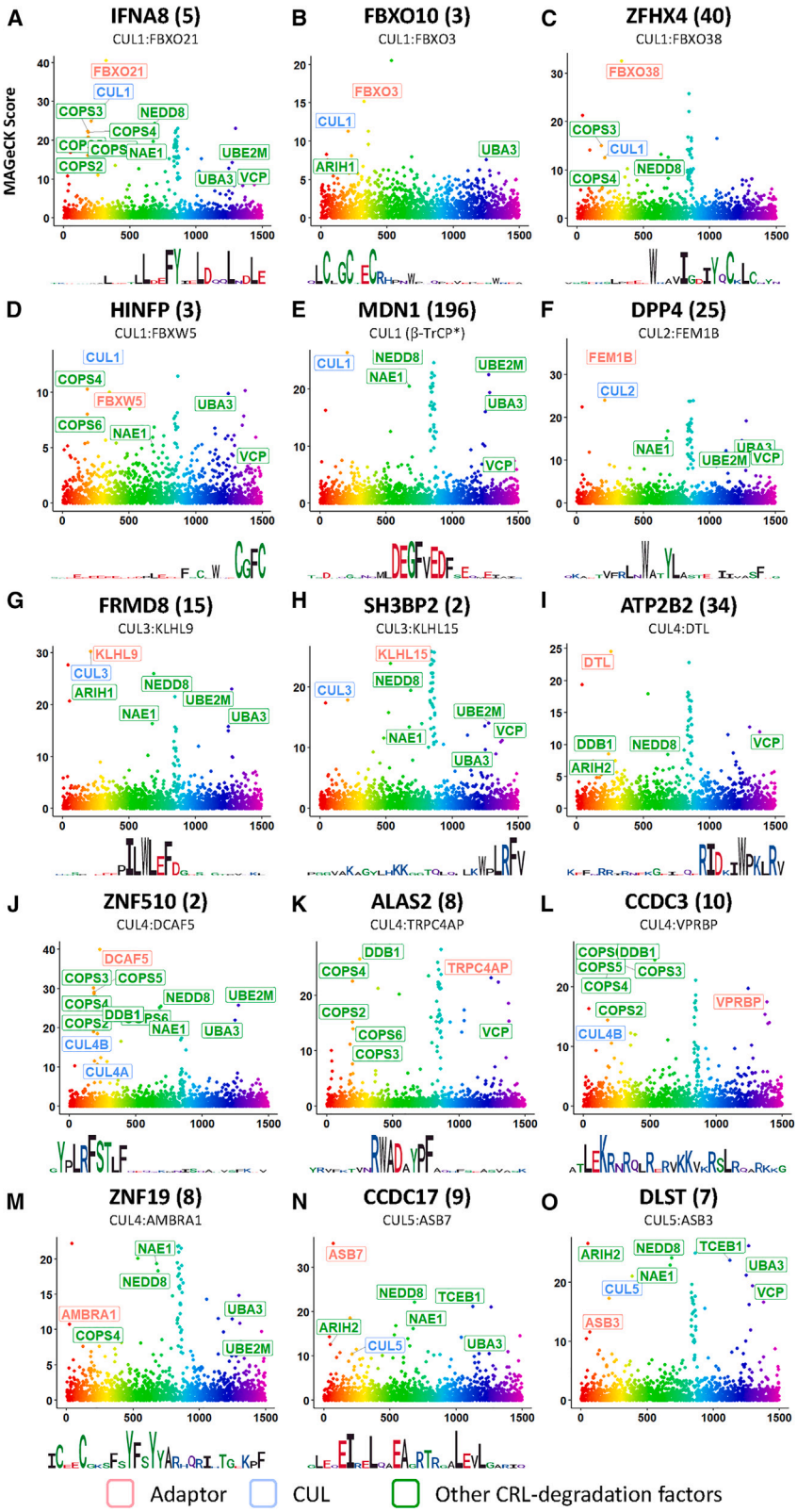


Figure 6. CRISPR screens identified the cognate E3 ligases for CRL-degron peptides

MAGeCK scores from CRISPR screens to identify cognate E3 ligase of each shown peptide: (A) IFNA8 (5), (B) FBXO10 (3), (C) ZFHX4 (40), (D) HINFP (3), (E) MDN1 (196), (F) DPP4 (25), (G) FRMD8 (15), (H) SH3BP2 (2), (I) ATP2B2 (34), (J) ZNF510 (2), (K) ALAS2 (8), (L) CCDC3 (10), (M) ZNF19 (8), (N) CCDC17 (9), and (O) DLST (7) (STAR Methods). Logo plots derived from the each peptide's mutagenesis footprint are also shown when available. See also Table S7 and Figures S4 and S5.

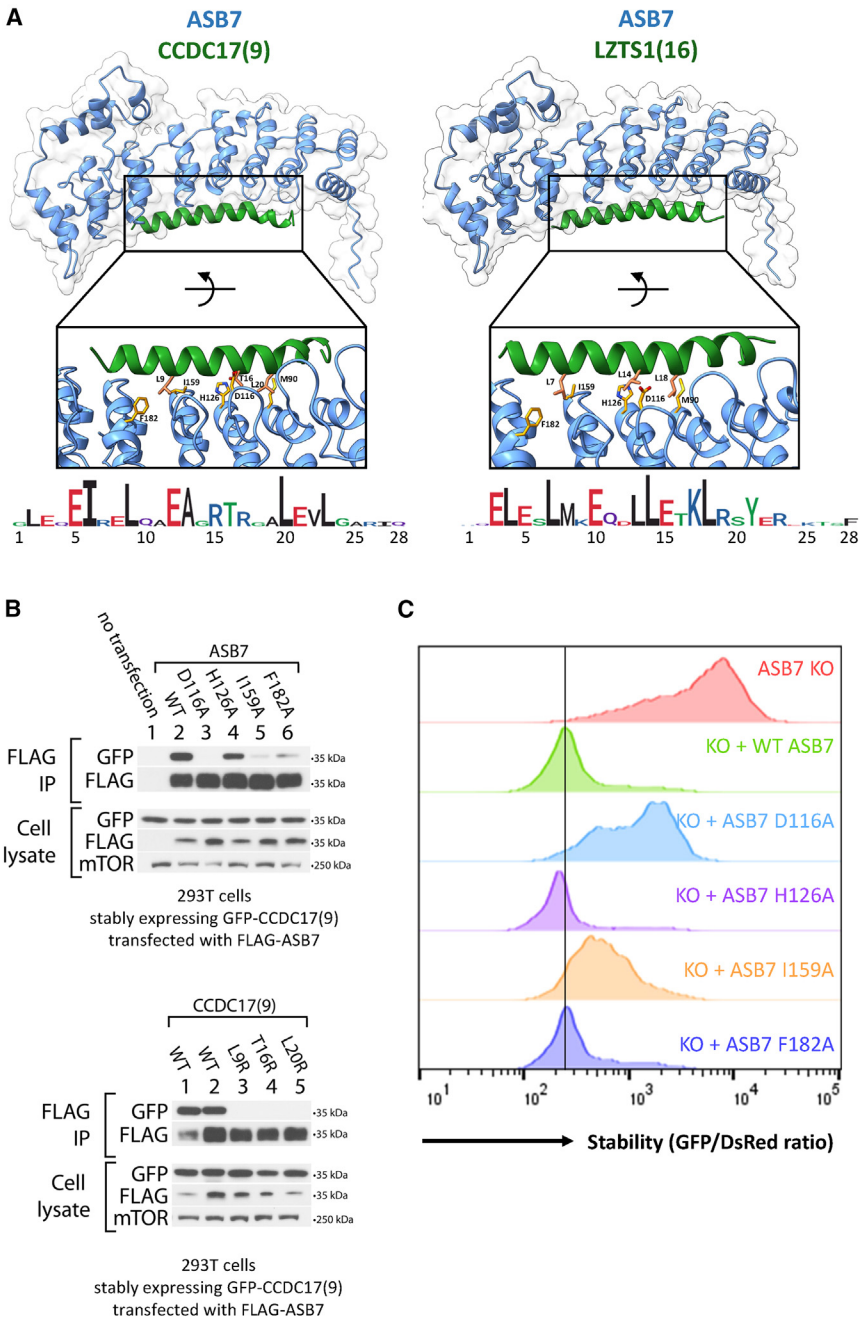


Figure 7. E3-degron docking by AlphaFold2 identified critical degron residues consistent with those revealed by saturation mutagenesis, co-immunoprecipitation, and GPS

(A) Alphafold2 multimer docking of ASB7 degron peptides onto ASB7. Selected residues at the interaction interface are shown and labeled.

(B) Immunoblots for ASB7:CCDC17(9) binding in 293T ASB7 KO cells. (Top) Co-immunoprecipitation of FLAG-ASB7 (WT or the indicated mutant) in cells stably expressing GFP-CCDC17(9) (WT). (Bottom) Co-immunoprecipitation of FLAG-ASB7 (WT) in cells stably expressing GFP-CCDC17(9) (WT or the indicated mutant). Cells from lanes 1, 3, 4, and 5 stably express the specified GFP-degrons fusions flanked with C-end sequence A (QGRARP NQEVQIGEMENQLS), whereas cells from lane 2 stably express GFP-CCDC17(9) flanked with C-end sequence B (QGRARPNQEVQIGEMENQLD).

(C) Flow stability data for CCDC17(9) GPS reporter peptide with KO, stably expressed WT, or stably expressed mutant ASB7. See also Figures S6 and S7.

simple amino acid compositions, as their stabilities can be predicted by summing the contributions due to composition alone. The amino acid correlations to stability were similar to those we previously observed with N- and C-end degrons, such as hydrophobicity correlating with instability and acidic residues correlating with stability; however, a few important differences stand out. Most prominently, we found that proline, glutamine, glycine, and serine correlated with stability, and threonine trended in that direction. Unexpectedly, all amino acid constituents of the so-called "PEST" (proline, glutamic acid, serine, and threonine) sequences[39] correlated with stability. PEST sequences have long been considered to be regions that confer instability on proteins.[39,40] Our findings suggest that, statistically speaking, these amino acids alone are very unlikely to cause instability. Either the PEST hypothesis is incorrect or it is a particular arrangement of those residues or phosphorylation that generates the degrons within these low-complexity regions. How these sequences confer stability on proteins remains to be determined.

From our scanning mutagenesis screen, we found that the most frequent class of non-CRL degrons was regulated by BAG6, which is an E3-associated chaperone that is known to recognize hydrophobic signal sequences on mislocalized membrane proteins. We found that BAG6 recognizes short hydrophobic degrons of 5–7 residues in length of a defined composition that is not relegated to stretches of hydrophobic residues but can accommodate internal neutral or even charged residues. There are clearly many additional sequence-specific degrons in the non-CRL class that are not BAG6 regulated, as determined by our scanning mutagenesis patterns, that should expand this degron class as well.

Mapping degron peptides to their cognate E3 ligases revealed 15 distinct sets of cognate CRLs, including $CUL1^{FBXO21}$, $CUL1^{FBXO3}$, $CUL1^{FBXO38}$, $CUL1^{FBXW5}$, $CUL1^{\beta\text{-}TrCP}$, $CUL2^{FEM1B}$, $CUL3^{KLHL9}$, $CUL3^{KLHL15}$, $CUL4^{DTL}$, $CUL4^{DCAF5}$, $CUL4^{TRPC4AP}$, $CUL4^{AMBRA1}$, $CUL4^{VPRBP}$, $CUL5^{ASB7}$, and $CUL5^{ASB3}$. Notably, except for $CUL1^{\beta\text{-}TrCP}$ and $CUL4^{DTL}$, our study identified previously undefined sets of precisely defined degron motifs

recognized by the other 13 CRLs, and based on the motif separation by DegronID, many additional degrons remain unexplored, and we expect many more orphan E3-degron pairs to emerge from future analysis of these degron motifs. The same strategy developed in this study could be applied to studying the non-CRL degron. Furthermore, cellular degron under different cellular conditions or in different cell types with differential E3 ligase activities could also be surveyed using this strategy that we built here.

We found that the degron motifs grouped together by DegronID tend to have visually noticeable similarities in their degron footprints. Degrons that cluster together by DegronID do not necessarily have the same E3. Multiple E3 ligases may recognize similar degron recognition sequences. However, by exploring degron motif clusters generated by DegronID, we did find sets of degron motifs that cluster together and do share the same E3 ligase, which has provided deeper insights into degron and E3 plasticity. A case in point is $CUL3^{KLHL15}$, an E3 ligase known to recognize the FRY degron. However, our saturation mutagenesis detected a much more nuanced degron that allowed many substitutions and expanded the number of additional residues involved in recognition from 3 to between 4 and 7 residues, with both required and avoided residues at each position. Presumably, degrons that deviate from the central FRY motif can take advantage of distal residues to improve their degrative potentials by gaining additional interactions. For example, many FRY degrons have a hydrophobic residue, often V, at the +4 position. Those having L at +1 position, e.g., CILP (51) and SYT2 (11), often have a hydrophobic at −2 and occasionally at −3 or −4.

Second, we identified a distinct FEM1B degron motif lacking the R* motif, W[VC]xxL, which makes use of a different degron-binding pocket present on FEM1B as predicted by AlphaFold2 modeling. In the case of R* degron, the C-terminal R binds to the D131 on the ANK3 repeat.[28] However, for the W[VC]xxL motif, it was predicted that the W of the degron coordinates with W367 located near the TPR domain. In the case of WxxYL, the Y of the degron coordinates with F501 on FEM1B (Figure S7A). We have validated that FEM1B W367 and F501 are critical to binding between FEM1B and a WVTYL-containing peptide by coIP and critical to destabilization of a WVTYL-containing peptide by GPS (Figures S7B and S7C). By contrast, FEM1B mutations H345A and Y84A, which are not predicted to contact critical residues of the WVTYL degron (Figure S7A), do not impair FEM1B:WVTYL binding by coIP.

Third, although many degrons reside in unstructured regions of proteins,[13] we identified peptides that stood out as clearly involving stable secondary structures, notably the CCDC17(9) peptide regulated by $CUL5^{ASB7}$ and the IFNA8(5) peptide regulated by $CUL1^{FBXO21}$ (Figures 7A and S6C). Both peptides were predicted to be α helical, and the hydrophobic residues required for degradation were primarily located on one face of the helix. Consistent with this structural prediction, placement of the α-helix-breaking proline and, to a lesser extent, glycine at any position along the degron region caused stabilization (Figures 5A and 5N), supporting a role for the helical structure in degron function. Importantly, docking these degron peptides with their E3 showed that the residues on one helical face that were predicted to be critical for degron activity were precisely the same residues that made contact with the E3. In the case of the ankyrin repeat protein ASB7, loops at the ends of the ankyrin repeats were spaced with the same pitch as the α-helix to identify equivalently spaced critical residues on the helical face. We have validated that mutations D116A or I159A on ASB7 perturb the ability of ASB7 to bind or destabilize the CCDC17(9) α-helix (Figures 7B and 7C). We also find for both ASB7:CCDC17(9) and FBXO21:IFNA8(5) that mutation of the putative critical residues along one helical face of the degron peptide impairs E3:degron binding (Figures 7B and S6D).

The large number of degrons mapped in this study have the potential to expand the proteolysis targeting chimeras (PROTACs) toolbox. PROTACs are an increasingly important concept in therapeutics.[41] Despite the profound excitement generated by PROTACs in the past decade, their full potential has not been fully realized. This is in part limited by the relatively few E3 ligases with precise definitions of their cognate degrons.[42] PROTACS often mimic endogenous substrates.[43] It was demonstrated that a chemical mimicking a known FEM1B degron enabled the discovery of a new PROTACs therapeutic.[44] This ligand binds to FEM1B by targeting a previously identified degron-binding pocket on FEM1B involving ANK3-ANK6.[27] This illustrates the feasibility and importance of developing new PROTACs therapeutics through understanding natural degron motifs.

Through the 40-year history of degron analysis, most degrons were discovered once per study, and the reported motifs were often incomplete. We began developing a high-throughput system for characterizing degrons but were previously only able to easily identify patterns and the N and C termini of peptides that were simple and gained specificity from being positioned at protein ends. However, the vast majority of degrons are internal. Here, we extended our earlier work on N/C-degrons and applied GPS to systematically study internal degrons. To gauge the degree of our progress relative to previous progress in the field of internal degrons over the last 40 years, the eukaryotic linear motif (ELM) database currently presents a collection of internal mammalian degrons recognized by 14 ubiquitin ligases.[45] Our proof-of-concept work involved mutagenic fingerprinting of a large set of internal degrons recognized by 16 ubiquitin ligases and highlighted many additional peptides with potential degron activity. Of these 16, only two E3s overlapped with the ELM collection, β-TrCP and CDT2 (DTL). Consequently, anticipate that many more degrons await to be uncovered in this dataset, and their knowledge will fuel the discovery of portions of full-length proteins regulated by cognate E3s and possibly coupled with structural prediction programs like AlphaFold2. Predictions of full-length substrates may become possible and may be facilitated by filtering high-throughput immunoprecipitation data by motif search similar to DegronID. Future studies would be aided by methods for high-throughput screening of E3:degron pairs, which have recently been developed.[46] We foresee that future research into these degrons will help unravel the complex but yet under-explored cellular degradome and their physiological roles in biological systems.

Limitations of the study

We chose a 28-mer tile peptidome approach to represent human internal degrons, and thus, we are likely to inherently miss some structural or conformational motifs. However, our peptidome

screen was able to effectively capture many α-helical degron motifs. The E3-degron pairs in this study were characterized genetically, and therefore, in some cases, the destabilization of the degron peptide by the E3 could in theory be due to indirect effects. However, the results from our substrate binding assays for FBXO21, ASB7, and FEM1B demonstrate that these effects are direct, as are the impacts of the stabilizing mutations in the peptide that disrupt E3 binding. This concurs with the analysis from our previous studies on N- and C-terminal degrons, in which the effects have all been direct when tested by ourselves and others.[10,14,28,33–35]

Some peptides with an E3-interacting motif may not score as a degron peptide for several reasons: (1) activation of degron activity requires a spatially accessible lysine for ubiquitination. Although there are 18 surface-accessible lysine residues in GFP, we cannot rule out the possibility that, in some cases, the E3s recruited by degrons cannot access a lysine for ubiquitination. (2) The cognate E3s for some degrons may not be expressed in HEK293T cells. (3) The CRL degrons require activation by post-translation modifications such as phosphorylation or acetylation. This peptide system likely lacks most protein modifications and will miss many of them. However, although CUL1$^{\beta\text{-TrCP}}$ required phosphorylation of its degrons, we identified 6 degrons that lack phosphosites but contain clusters of acidic residues that are likely to act as phosphomimetics (Figure S5A). Thus, we may be able to detect E3 ligases that require phosphates on some degrons.

Although we were able to use the AlphaFold2 prediction models to identify critical E3 residues for degron binding that disrupted coIPs when mutagenized, confirmed structures would certainly aid in the identification of such residues.

Despite these limitations, these systematic studies represent a substantial contribution to the available knowledge in this field and should provide the basis for the discovery of many additional degron-E3 pairs in the future.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.molcel.2023.08.022.

ACKNOWLEDGMENTS

We thank members of the Elledge lab for helpful suggestions and reagents. We thank the Harvard University Medical School Department of Immunology Flow Cytometry Core Facility for their invaluable help in cell sorting. The α-helix icon used in the graphical abstract was designed by DBCLS https://togotv.dbcls.jp/en/pics.html and is licensed under CC-BY 4.0 Unported https://creativecommons.org/licenses/by/4.0/. Z.Z. is a Croucher PhD Scholar. R.T.T. is a Sir Henry Wellcome Postdoctoral Fellow (201387/Z/16/Z) and a Pemberton-Trinity Fellow. This work was supported by the National Institutes of Health Aging grant AG11085 (S.J.E.) S.J.E. is a member of the Ludwig Center at Harvard and an Investigator with the Howard Hughes Medical Institute.

AUTHOR CONTRIBUTIONS

Conceptualization, Z.Z., B.S., and S.J.E.; methodology, Z.Z., B.S., and S.J.E.; software, Z.Z. and B.S.; visualization, Z.Z. and B.S.; investigation, Z.Z., B.S., A.C., Y.L., C.N., and S.J.E.; writing – original draft, Z.Z., B.S., and S.J.E.; writing – review & editing, Z.Z., B.S., and S.J.E.; funding acquisition: S.J.E.; resources, R.T.T.; supervision, S.J.E.

DECLARATION OF INTERESTS

S.J.E. is a founder of TSCAN Therapeutics, MAZE Therapeutics, ImmuneID and Mirimus; serves on the scientific advisory boards of Homology Medicines, ImmuneID, MAZE Therapeutics, and TSCAN Therapeutics; is an advisor for MPM Capital; and is on the Editorial Board for *Molecular Cell*. None of this affects this work.

Received: February 17, 2023
Revised: June 7, 2023
Accepted: August 17, 2023
Published: September 21, 2023

REFERENCES

1. Li, W., Bengtson, M.H., Ulbrich, A., Matsuda, A., Reddy, V.A., Orth, A., Chanda, S.K., Batalov, S., and Joazeiro, C.A.P. (2008). Genome-wide and functional annotation of human E3 ubiquitin ligases identifies MULAN, a mitochondrial E3 that regulates the organelle's dynamics and signaling. PLoS One *3*, e1487. https://doi.org/10.1371/JOURNAL.PONE.0001487.
2. Collins, G.A., and Goldberg, A.L. (2017). The logic of the 26S proteasome. Cell *169*, 792–806. https://doi.org/10.1016/j.cell.2017.04.023.
3. Popovic, D., Vucic, D., and Dikic, I. (2014). Ubiquitination in disease pathogenesis and treatment. Nat. Med. *20*, 1242–1253. https://doi.org/10.1038/nm.3739.

4. Juszkiewicz, S., and Hegde, R.S. (2018). Quality control of orphaned proteins. Mol. Cell *71*, 443–457. https://doi.org/10.1016/J.MOLCEL.2018.07.001.

5. Yanagitani, K., Juszkiewicz, S., and Hegde, R.S. (2017). UBE2O is a quality control factor for orphans of multiprotein complexes. Science *357*, 472–475. https://doi.org/10.1126/science.aan0178.

6. Mena, E.L., Kjolby, R.A.S., Saxton, R.A., Werner, A., Lew, B.G., Boyle, J.M., Harland, R., and Rape, M. (2018). Dimerization quality control ensures neuronal development and survival. Science *362*, eaap8236. https://doi.org/10.1126/science.aap8236.

7. Mårtensson, C.U., Priesnitz, C., Song, J., Ellenrieder, L., Doan, K.N., Boos, F., Floerchinger, A., Zufall, N., Oeljeklaus, S., Warscheid, B., et al. (2019). Mitochondrial protein translocation-associated degradation. Nature *569*, 679–683. https://doi.org/10.1038/s41586-019-1227-y.

8. Hessa, T., Sharma, A., Mariappan, M., Eshleman, H.D., Gutierrez, E., and Hegde, R.S. (2011). Protein targeting and degradation are coupled for elimination of mislocalized proteins. Nature *475*, 394–397. https://doi.org/10.1038/nature10181.

9. Khmelinskii, A., Blaszczak, E., Pantazopoulou, M., Fischer, B., Omnus, D.J., Le Dez, G., Brossard, A., Gunnarsson, A., Barry, J.D., Meurer, M., et al. (2014). Protein quality control at the inner nuclear membrane. Nature *516*, 410–413. https://doi.org/10.1038/nature14096.

10. Timms, R.T., Zhang, Z., Rhee, D.Y., Harper, J.W., Koren, I., and Elledge, S.J. (2019). A glycine-specific N-degron pathway mediates the quality control of protein N-myristoylation. Science *365*, eaaw4912. https://doi.org/10.1126/science.aaw4912.

11. Lucas, X., and Ciulli, A. (2017). Recognition of substrate degrons by E3 ubiquitin ligases and modulation by small-molecule mimicry strategies. Curr. Opin. Struct. Biol. *44*, 101–110. https://doi.org/10.1016/j.sbi.2016.12.015.

12. Bachmair, A., Finley, D., and Varshavsky, A. (1986). In vivo half-life of a protein is a function of its amino-terminal residue. Science *234*, 179–186. https://doi.org/10.1126/science.3018930.

13. Mészáros, B., Kumar, M., Gibson, T.J., Uyar, B., and Dosztányi, Z. (2017). Degrons in cancer. Sci. Signal. *10*, eaak9982. https://doi.org/10.1126/scisignal.aak9982.

14. Koren, I., Timms, R.T., Kula, T., Xu, Q., Li, M.Z., and Elledge, S.J. (2018). The eukaryotic proteome is shaped by E3 ubiquitin ligases targeting C-terminal degrons. Cell *173*, 1622–1635.e14. https://doi.org/10.1016/j.cell.2018.04.028.

15. Sherpa, D., Chrustowicz, J., and Schulman, B.A. (2022). How the ends signal the end: regulation by E3 ubiquitin ligases recognizing protein termini. Mol. Cell *82*, 1424–1438. https://doi.org/10.1016/j.molcel.2022.02.004.

16. Tokheim, C., Wang, X., Timms, R.T., Zhang, B., Mena, E.L., Wang, B., Chen, C., Ge, J., Chu, J., Zhang, W., et al. (2021). Systematic characterization of mutations altering protein degradation in human cancers. Mol. Cell *81*, 1292–1308.e11. https://doi.org/10.1016/j.molcel.2021.01.020.

17. Rodrigo-Brenni, M.C., Gutierrez, E., and Hegde, R.S. (2014). Cytosolic quality control of mislocalized proteins requires RNF126 recruitment to Bag6. Mol. Cell *55*, 227–237. https://doi.org/10.1016/j.molcel.2014.05.025.

18. Jarjanazi, H., Savas, S., Pabalan, N., Dennis, J.W., and Ozcelik, H. (2008). Biological implications of SNPs in signal peptide domains of human proteins. Proteins *70*, 394–403. https://doi.org/10.1002/prot.21548.

19. Drozdetskiy, A., Cole, C., Procter, J., and Barton, G.J. (2015). JPred4: a protein secondary structure prediction server. Nucleic Acids Res. *43*, W389–W394. https://doi.org/10.1093/nar/gkv332.

20. Davey, N.E., and Morgan, D.O. (2016). Building a regulatory network with short linear sequence motifs: lessons from the degrons of the anaphase-promoting complex. Mol. Cell *64*, 12–23. https://doi.org/10.1016/j.molcel.2016.09.006.

21. Low, T.Y., Peng, M., Magliozzi, R., Mohammed, S., Guardavaccaro, D., and Heck, A.J. (2014). A systems-wide screen identifies substrates of the SCFβTrCP ubiquitin ligase. Sci. Signal. *7*, rs8. https://doi.org/10.1126/scisignal.2005882.

22. Loveless, T.B., Topacio, B.R., Vashisht, A.A., Galaang, S., Ulrich, K.M., Young, B.D., Wohlschlegel, J.A., and Toczyski, D.P. (2015). DNA damage regulates translation through β-TRCP targeting of CReP. PLoS Genet. *11*, e1005292. https://doi.org/10.1371/journal.pgen.1005292.

23. Havens, C.G., and Walter, J.C. (2009). Docking of a specialized PIP Box onto chromatin-bound PCNA creates a degron for the ubiquitin ligase CRL4Cdt2. Mol. Cell *35*, 93–104. https://doi.org/10.1016/j.molcel.2009.05.012.

24. Lee, J.M., Lee, J.S., Kim, H., Kim, K., Park, H., Kim, J.Y., Lee, S.H., Kim, I.S., Kim, J., Lee, M., et al. (2012). EZH2 generates a methyl degron that is recognized by the DCAF1/DDB1/CUL4 E3 ubiquitin ligase complex. Mol. Cell *48*, 572–586. https://doi.org/10.1016/j.molcel.2012.09.004.

25. Wang, X., Arceci, A., Bird, K., Mills, C.A., Choudhury, R., Kernan, J.L., Zhou, C., Bae-Jump, V., Bowers, A., and Emanuele, M.J. (2017). VprBP/DCAF1 regulates the degradation and nonproteolytic activation of the cell cycle transcription factor FoxM1. Mol. Cell. Biol. *37*, e00609-16. https://doi.org/10.1128/MCB.00609-16.

26. Zhang, C., Li, X., Adelmant, G., Dobbins, J., Geisen, C., Oser, M.G., Wucherpfenning, K.W., Marto, J.A., and Kaelin, W.G. (2015). Peptidic degron in EID1 is recognized by an SCF E3 ligase complex containing the orphan F-box protein FBXO21. Proc. Natl. Acad. Sci. USA *112*, 15372–15377. https://doi.org/10.1073/pnas.1522006112.

27. Manford, A.G., Mena, E.L., Shih, K.Y., Gee, C.L., McMinimy, R., Martínez-González, B., Sherriff, R., Lew, B., Zoltek, M., Rodríguez-Pérez, F., et al. (2021). Structural basis and regulation of the reductive stress response. Cell *184*, 5375–5390.e16. https://doi.org/10.1016/j.cell.2021.09.002.

28. Chen, X., Liao, S., Makaros, Y., Guo, Q., Zhu, Z., Krizelman, R., Dahan, K., Tu, X., Yao, X., Koren, I., et al. (2021). Molecular basis for arginine C-terminal degron recognition by Cul2FEM1 E3 ligase. Nat. Chem. Biol. *17*, 254–262. https://doi.org/10.1038/s41589-020-00704-3.

29. Ferretti, L.P., Himmels, S.F., Trenner, A., Walker, C., von Aesch, C., Eggenschwiler, A., Murina, O., Enchev, R.I., Peter, M., Freire, R., et al. (2016). Cullin3-KLHL15 ubiquitin ligase mediates CtIP protein turnover to fine-tune DNA-end resection. Nat. Commun. *7*, 12628. https://doi.org/10.1038/ncomms12628.

30. Song, J., Merrill, R.A., Usachev, A.Y., and Strack, S. (2021). The X-linked intellectual disability gene product and E3 ubiquitin ligase KLHL15 degrades doublecortin proteins to constrain neuronal dendritogenesis. J. Biol. Chem. *296*, 100082. https://doi.org/10.1074/jbc.RA120.016210.

31. Oberg, E.A., Nifoussi, S.K., Gingras, A.C., and Strack, S. (2012). Selective proteasomal degradation of the B′β subunit of protein phosphatase 2A by the E3 ubiquitin ligase adaptor Kelch-like 15. J. Biol. Chem. *287*, 43378–43389. https://doi.org/10.1074/jbc.M112.420281.

32. Montgomerie, S., Cruz, J.A., Shrivastava, S., Arndt, D., Berjanskii, M., and Wishart, D.S. (2008). PROTEUS2: a web server for comprehensive protein structure prediction and structure-based annotation. Nucleic Acids Res. *36*, W202–W209. https://doi.org/10.1093/nar/gkn255.

33. Yan, X., Li, Y., Wang, G., Zhou, Z., Song, G., Feng, Q., Zhao, Y., Mi, W., Ma, Z., and Dong, C. (2021). Molecular basis for recognition of Gly/N-degrons by CRL2ZYG11B and CRL2ZER1. Mol. Cell *81*, 3262–3274.e3. https://doi.org/10.1016/j.molcel.2021.06.010.

34. Yan, X., Wang, X., Li, Y., Zhou, M., Li, Y., Song, L., Mi, W., Min, J., and Dong, C. (2021). Molecular basis for ubiquitin ligase CRL2FEM1C-mediated recognition of C-degron. Nat. Chem. Biol. *17*, 263–271. https://doi.org/10.1038/s41589-020-00703-4.

35. Rusnac, D.V., Lin, H.C., Canzani, D., Tien, K.X., Hinds, T.R., Tsue, A.F., Bush, M.F., Yen, H.-C.S., and Zheng, N. (2018). Recognition of the diglycine C-end Degron by CRL2KLHDC2 ubiquitin ligase. Mol. Cell *72*, 813–822.e4. https://doi.org/10.1016/j.molcel.2018.10.021.

36. Tsaban, T., Varga, J.K., Avraham, O., Ben-Aharon, Z., Khramushin, A., and Schueler-Furman, O. (2022). Harnessing protein folding neural networks for peptide–protein docking. Nat. Commun. *13*, 176. https://doi.org/10.1038/s41467-021-27838-9.

37. Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022). ColabFold: making protein folding accessible to all. Nat. Methods *19*, 679–682. https://doi.org/10.1038/s41592-022-01488-1.

38. Li, J., Mahajan, A., and Tsai, M.D. (2006). Ankyrin repeat: a unique motif mediating protein−protein interactions. Biochemistry *45*, 15168–15178. https://doi.org/10.1021/bi062188q.

39. Rogers, S., Wells, R., and Rechsteiner, M. (1986). Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. Science *234*, 364–368. https://doi.org/10.1126/science.2876518.

40. Rechsteiner, M., and Rogers, S.W. (1996). PEST sequences and regulation by proteolysis. Trends Biochem. Sci. *21*, 267–271. https://doi.org/10.1016/S0968-0004(96)10031-1.

41. Burslem, G.M., and Crews, C.M. (2020). Proteolysis-targeting chimeras as therapeutics and tools for biological discovery. Cell *181*, 102–114. https://doi.org/10.1016/j.cell.2019.11.031.

42. Paiva, S.L., and Crews, C.M. (2019). Targeted protein degradation: elements of PROTAC design. Curr. Opin. Chem. Biol. *50*, 111–119. https://doi.org/10.1016/j.cbpa.2019.02.022.

43. Ichikawa, S., Flaxman, H.A., Xu, W., Vallavoju, N., Lloyd, H.C., Wang, B., Shen, D., Pratt, M.R., and Woo, C.M. (2022). The E3 ligase adapter cereblon targets the C-terminal cyclic imide degron. Nature *610*, 775–782. https://doi.org/10.1038/s41586-022-05333-5.

44. Henning, N.J., Manford, A.G., Spradlin, J.N., Brittain, S.M., Zhang, E., McKenna, J.M., Tallarico, J.A., Schirle, M., Rape, M., and Nomura, D.K. (2022). Discovery of a covalent FEM1B recruiter for targeted protein degradation applications. J. Am. Chem. Soc. *144*, 701–708. https://doi.org/10.1021/jacs.1c03980.

45. Kumar, M., Michael, S., Alvarado-Valverde, J., Mészáros, B., Sámano-Sánchez, H., Zeke, A., Dobson, L., Lazar, T., Örd, M., Nagpal, A., et al. (2022). The Eukaryotic Linear Motif resource: 2022 release. Nucleic Acids Res. *50*, D497–D508. https://doi.org/10.1093/nar/gkab975.

46. Timms, R.T., Mena, E.L., Leng, Y., Li, M.Z., Tchasovnikarova, I.A., Koren, I., and Elledge, S.J. (2023). Defining E3 ligase-substrate relationships through multiplex CRISPR screening. Nat. Cell Biol. https://doi.org/10.1038/s41556-023-01229-2.

47. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359. https://doi.org/10.1038/nmeth.1923.

48. Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25. https://doi.org/10.1186/gb-2009-10-3-r25.

49. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. *17*, 10–12. https://doi.org/10.14806/ej.17.1.200.

50. Li, W., Xu, H., Xiao, T., Cong, L., Love, M.I., Zhang, F., Irizarry, R.A., Liu, J.S., Brown, M., and Liu, X.S. (2014). MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. Genome Biol. *15*, 554. https://doi.org/10.1186/s13059-014-0554-4.

51. Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.H., and Ferrin, T.E. (2021). UCSF ChimeraX: structure visualization for researchers, educators, and developers. Protein Sci. *30*, 70–82. https://doi.org/10.1002/pro.3943.

52. Kuhn, M. (2008). Building predictive models in R using the caret package. J. Stat. Softw. *28*, 1–26. https://doi.org/10.18637/jss.v028.i05.

53. Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., et al. (2019). Welcome to the tidyverse. J. Open Source Softw. *4*, 1686. https://doi.org/10.21105/joss.01686.

54. Sanjana, N.E., Shalem, O., and Zhang, F. (2014). Improved vectors and genome-wide libraries for CRISPR screening. Nat. Methods *11*, 783–784. https://doi.org/10.1038/nmeth.3047.

55. Meyer, M., and Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb. Protoc. *2010*, pdb.prot5448. https://doi.org/10.1101/pdb.prot5448.

56. Ginestet, C. (2011). ggplot2: elegant graphics for data analysis. J. R. Stat. Soc. Ser. A Stat. Soc. *174*, 245–246. https://doi.org/10.1111/j.1467-985X.2010.00676_9.x.

57. Zhao, Y., Gran, B., Pinilla, C., Markovic-Plese, S., Hemmer, B., Tzou, A., Whitney, L.W., Biddison, W.E., Martin, R., and Simon, R. (2001). Combinatorial peptide libraries and biometric score matrices permit the quantitative analysis of specific and degenerate interactions between clonotypic TCR and MHC peptide ligands. J. Immunol. *167*, 2130–2141. https://doi.org/10.4049/jimmunol.167.4.2130.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit monoclonal anti-mTOR	Cell Signaling Technology	Cat# 2983; RRID:AB_2105622
Rabbit monoclonal anti-FLAG	Cell Signaling Technology	Cat# 14793; RRID:AB_2572291
Rabbit monoclonal anti-GFP	Cell Signaling Technology	Cat# 2956; RRID:AB_1196615.
Goat anti-Rabbit secondary antibody, HRP conjugate	Cell Signaling Technology	Cat# 7074; RRID:AB_2099233
Chemicals, peptides, and recombinant proteins		
MLN4924	Active Biochem	Cat# A-1139
Polybrene	Santa Cruz	Cat# sc-134220
Critical commercial assays		
QIAEX II Gel Extraction Kit	QIAGEN	Cat# 20051
PCR purification kit	QIAGEN	Cat# 28106
Gentra Puregene Cell Kit	QIAGEN	Cat# 158767
Deposited data		
GPS peptidome processed screen data and raw images	This paper	Mendeley Data: https://doi.org/10.17632/32h94v2yv2.2
GPS peptidome raw and processed screen data	This paper	Gene Expression Omnibus: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?&acc=GSE240610
Experimental models: Cell lines		
HEK293T	ATCC	Cat# CRL-3216
Oligonucleotides		
sg-FEM1B: GTGACATAGCCAAGCAGATAG	Integrated DNA Technologies	N/A
sg-KLHL15: GATTTCGGCGTAAACATCGA	Integrated DNA Technologies	N/A
sg-ASB7: GCCAACATCGACATTCAGAA	Integrated DNA Technologies	N/A
Recombinant DNA		
pHAGE-GPS6.0-DEST	This paper	N/A
pHAGE-GPS6.0-peptide libraries	This paper	N/A
lentiCRISPR v2	Addgene	52961
Ultimate ORF Collection	Thermo Fisher Scientific	https://www.thermofisher.com/us/en/home/life-science/cloning/clone-collections/ultimate-orf-clone-collection.html
Software and algorithms		
Bowtie 2	Langmead and Salzberg[47]; Langmead et al.[48]	http://bowtie-bio.sourceforge.net/index.shtml
Cutadapt	Martin[49]	http://cutadapt.readthedocs.io/en/stable/index.html
MAGeCK	Li et al.[50]	https://sourceforge.net/projects/mageck/
Flowjo	Flowjo	https://www.flowjo.com
ColabFold	Mirdita et al.[37]	https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb
ChimeraX	Pettersen et al.[51]	https://www.rbvi.ucsf.edu/chimerax/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Proteus2	Montgomerie et al.[32]	http://www.proteus2.ca/proteus2/
JPred4	Drozdetskiy et al.[19]	https://www.compbio.dundee.ac.uk/jpred/
DegronID	This paper	Zenodo: https://doi.org/10.5281/zenodo.8103915
R 4.1.2	R Core Team	https://www.r-project.org/
Caret	Kuhn[52]	https://doi.org/10.18637/jss.v028.i05
Tidyverse	Wickham et al.[53]	https://doi.org/10.21105/joss.01686
Biostrings	Bioconductor	https://doi.org/10.18129/B9.bioc.Biostrings
aplot	CRAN	https://cran.r-project.org/web/packages/aplot
ggdendro	CRAN	https://cran.r-project.org/web/packages/ggdendro

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Stephen J. Elledge (selledge@genetics.med.harvard.edu).

Materials availability

Any reagents that are unique to this study will be made available upon request.

Data and code availability

- Original GPS peptidome screen data and raw western blot images were deposited to Mendeley Data and Gene Expression Omnibus (GEO) and are publicly available as of the date of publication. The DOIs are listed in the key resources table under "Deposited Data" as "GPS peptidome processed screen data and raw images" (Mendeley Data) and "GPS peptidome raw and processed screen data" (GEO). Detailed GPS peptidome screen data, mutagenesis screen data, CRISPR screen data, and DN CUL analyses are available in Tables S1, S2, S3, S4, S5, S6, and S7.
- All original code has been deposited to Zenodo and is publicly available as of the date of publication. The DOI is listed in the key resources table under "Software and algorithms" as "DegronID".
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Cell lines

We cultured HEK293T (ATCC CRL-3216) cells at 37°C and 5% CO_2 in DMEM, high glucose, GlutaMAX™ Supplement, pyruvate (ThermoFisher Scientific) added with 10% fetal bovine serum (HyClone), 1% of 10,000 U/mL Penicillin-Streptomycin (ThermoFisher Scientific). For MLN4924 experiment, we incubated cells overnight with 1 mM MLN4924 and then analyzed the GFP/DsRed ratio by flow cytometry. For DN CULs experiment, we infected cells with the DN CULs lentivirus and analyzed the GFP/DsRed ratio by flow cytometry 2 days post-infection.

METHOD DETAILS

Transfection and lentivirus production

We generated lentivirus by transfection of HEK293T cells using lentiviral vector together DNA with plasmids encoding Gag-Pol, Rev, Tat and VSV-G using PolyJet In Vitro DNA Transfection Reagent (SignaGen Laboratories) based on the recommendation from the manufacturer. We collected the supernatants containing lentivirus after 2 days and used them to infect target cells. We added polybrene at 8 mg/ml for lentiviral infections when performing screens where there is need to increase infectivity.

CellPress
OPEN ACCESS

Plasmids

Dominant-negative Cullin constructs were kindly provided as a gift from W. Harper. For the cloning of selected degron substrates into the GPS6.0 DEST vector for downstream analyses (DN CULs analysis, CRISPR screens, E3-degron pair validation using individual sgRNAs), the degron peptides were cloned into GPS6.0 DEST vector in different batches using either C-End sequence A (QGRARPNQEVQIGEMENQLS, same linker as the one in the peptidome libraries), or C-End sequence B (QGRARPNQEV QIGEMENQLD) (see Table S5 for details). Both C-End sequences have no known C-degrons and differ only by the C-terminal residue. For individual gene-disruption experiments by CRISPR/Cas9, sgRNAs were cloned into lentiCRISPR v2 (Addgene #52961, deposited by Feng Zhang) as described.[54] The sgRNA sequences involved in this study were listed in the key resources table. For co-immunoprecipitation, E3 ligase ORFs were cloned into the pHAGE-CMV-2xFlag-N DEST vector. For flow-based stability assays, E3 ligase ORFs were cloned into pHAGE EF1α BFP.

Flow cytometry

HEK293T cells were detached from plates using trypsin and then analyzed on a CytoFLEX (Beckman Coulter). We collected flow cytometry data using CytoFLEX (Beckman Coulter) and performed analyses using FlowJo. We performed the cell sorting via the MoFlo Astrios (Beckman Coulter) instrument.

Design and generation of GPS libraries

GPS-peptidome library

The human proteome sequences were truncated into 28 residues with 5 residues overlaps between neighboring peptides and with 15 bp flanking primers, encoding GIRSG-28-mer-QGRAR. Oligonucleotides were synthesized by Agilent and amplified by PCR. The amplicons were then cloned into the lentiviral GPS vector pHAGE-GPS6.0-DEST via NEBuilder® HiFi DNA Assembly Master Mix (New England Biolabs) at 6x coverage. The 28-mer peptides were followed by a 20 amino acid C-End sequence (QGRARPNQEVQIGEMENQLS).

Mutagenesis libraries

The scanning mutagenesis libraries included a total of 9817 peptides, including 5790 peptides of degron index greater than 1, 1782 selected peptides of MLN4924-ΔPSI greater than 1, together with 3023 other peptides for the purpose balancing the GFP/DsRed distribution of the entire library to assist sorting. For scanning mutagenesis oligo libraries, we mutated each residue from each selected peptide was in succession to an amino acid of different side-chain chemical property: "A, G, V, L, I, S, T, C, D, E" were mutated to "R"; "M, W, F, P" were mutated to "S"; "Y, N, Q, K, R, H" were mutated to "A". For saturation mutagenesis, we mutated each residue from each selected peptide was in succession to all of the other 19 possible amino acids. For scanning mutagenesis of C-terminal peptides, mutations to "R" at the third position from C terminus were substituted with mutations to "H", and mutations to "A" at the first and second positions from C terminus were substituted with mutations to "S", for the purpose of avoiding the generation of known C-degrons by the mutagenesis. We added two constant 15 bp flanking sequences at the two ends for amplification purposes. Oligonucleotides were synthesized by Agilent and amplified by PCR. The amplicons were then cloned into the lentiviral GPS vector pHAGE-GPS6.0-DEST by NEBuilder® HiFi DNA Assembly Master Mix (New England Biolabs) at >100x coverage. The 28-mer peptides were followed by a 20 amino acid C-End sequence (QGRARPNQEVQIGEMENQLS).

GPS screens

We packaged GPS-peptidome lentiviral libraries and transduced HEK293T at low multiplicity of infection (MOI). Two days after transduction, hygromycin was added to eliminate untransduced cells and 7-9 days after transduction, cells were sorted into 6 bins in approximately 1:1 ratio according to the GFP/DsRed ratio. Sorted cells were immediately subjected to genomic DNA extraction or were expanded by cell culture for 2-3 days before downstream genomic DNA extraction.

For each bin, we performed genomic DNA extraction (QIAGEN Gentra Puregene Cell Kit). We amplified the peptide sequences cloned into GFP by PCR with Q5 Hot Start Polymerase (NEB) using PCR primers annealing to constant sequences shared across the library. We then purified the amplicons (QIAGEN PCR purification kit). For each bin, we added >100 ng purified amplicons as template for another PCR reaction that adds an Illumina P5 sequence and nucleotides of 1-7 bp at the 5' end for the purpose of staggering, together with an Illumina index and a P7 sequence at the 3' end.[55] We then quantified samples by agarose gel electrophoresis and pooled them together. We then purified the resulting sample by agarose gel electrophoresis followed by gel extraction (QiaQuick Gel Extraction kit). The purified sample was then sequenced by an Illumina NextSeq instrument. We aimed to maintain >50-fold representation throughout the screening process.

CRISPR screens

We packaged the sgRNA lentiviral libraries targeting ~1500 genes related to ubiquitination and transduced HEK293T expressing specific GFP-peptide fusions at a low MOI. Two days after transduction, puromycin was added to eliminate untransduced cells. 7-9 d after transduction, the top 5-10% cells with the largest GFP/DsRed ratio were isolated by FACS. For sgRNA lentiviral libraries prepared from the same batch of CRISPR library DNA, one group of unsorted population was collected as a control. sgRNA sequences in sorted samples and unsorted controls were then amplified by 2-steps PCR as described for the GPS screens and quantified by Illumina sequencing. We aimed to maintain >100-fold representation throughout the screening process. However, for some samples, coverages were compromised at the genomic DNA extraction step.

Measurement of protein half-life

HEK293T cells were separately mutated for ASB7, FBXO21, or FEM1B using lentiCRISPR v2 as mentioned earlier in STAR Methods. E3 mutant and wild-type 293T cells were infected with lentivirus to induce stable expression of a GPS peptide construct that we identified from the GPS peptidome screen as regulated by that E3. For ASB7, we used CCDC17(9). For FBXO21, we used IFNA8(5). For FEM1B, we used UNC13D(26). The 28-mer peptides were each followed by C-end sequence A (QGRARPNQEVQIGEMENQLS). E3 knockout and WT cells stably expressing GPS peptide reporter constructs were seeded onto 6-well plates at 0.4M cells/well and incubated in a CO2 incubator at 37°C. After 48 h of incubation, the media was replaced with media containing 100 ug/mL cycloheximide (Millipore Sigma 239765) for inhibition of protein synthesis, or DMSO as a control. Cells were exposed to cycloheximide for 0, 0.5, 1, 2, 4, or 6 hours. Then, cells were rinsed once with PBS and lysed using RIPA buffer supplemented with 1x protease and phosphatase inhibitor cocktail (Thermo Fisher Scientific 78441). Cell lysates were centrifuged at 21,000xg for 10 min at 4°C. Protein concentration in the supernatant was normalized across samples using BCA assay (Thermo Fisher Scientific 23225) prior to immunoblotting.

Immunoblotting

Expression levels of GFP-peptide fusions were measured by western blot with anti-GFP antibody. mTOR levels were also measured by western blot as a control. Samples for immunoblotting were diluted in Tris-Glycine SDS sample buffer and loaded into 4-12% Tris-Glycine 15-well pre-cast gels (Thermo Fisher Scientific XP04125BOX) with Tris-Glycine SDS running buffer (Thermo Fisher Scientific LC2675-4). A 10-250 kDa prestained protein ladder was run alongside samples (Thermo Fisher Scientific 26619). Eletrophoresis was run at 165V until the ladder reached the bottom of the gel. The protein was then loaded into a Trans-Blot Turbo Transfer System (BioRad) and transferred from the gel to a 0.2 μm nitrocellulose membrane (BioRad 1704158). Membranes were blocked in 1x TBST (CST 9997S) with 5% milk for 30 min at room temperature. Then rabbit anti-mTOR (CST 2983S), rabbit anti-FLAG (CST 14793S), or rabbit anti-GFP (CST 2956S) primary antibodies were added to the blocking solution at 1:1000. Membranes were then incubated at 4°C overnight with gentle rocking. Membranes were then rinsed repeatedly with 1x TBST, and HRP-linked anti-rabbit secondary antibody (CST 7074S) were added at 1:2000 in 1x TBST with 5% milk. Blots were incubated with secondary antibody for 1 h at room temperature with gentle rocking, then washed again with 1x TBST. Blots were then exposed to Pierce ECL western blotting substrate (Thermo Fisher Scientific 32106) and exposed to high sensitivity autoradiography film (Denville Scientific E3218).

Co-immunoprecipitation

HEK 293T cells that had been previously mutated for ASB7, FBXO21, or FEM1B by lentiCRISPR v2 were made to stably express corresponding GPS reporter peptides by lentiviral infection (ASB7:CCDC17(9), FBXO21:IFNA8(5), FEM1B:WVTYL_28mer). The 28-mer peptides were each followed by C-end sequence A (QGRARPNQEVQIGEMENQLS). Cells were seeded onto 100 mm plates at 1.1M cells/plate. 48 h later, cells were transfected with 3 ug DNA/plate of WT or mutant E3 ligase in the pHAGE-CMV-2xFlag-N DEST vector. Approximately 24 h later, media was aspirated and replaced with media containing 1 μM MLN4924. After 16 h of incubation with MLN4924, cells were rinsed with ice-cold PBS and collected by scraping in 0.7 mL of lysis buffer containing 0.5% CHAPS, 40 mM HEPES pH 7.4, 100 mM NaCl, 4 mM EDTA, and 1x protease and phosphatase inhibitor cocktail. Cell lysates were incubated at 4°C with end-over-end rotation for 10 min, then centrifuged at 21,000xg for 15 min at 4°C. A 50 uL aliquot of the supernatant was collected as input, and the rest of the supernatant was transferred to a clean Eppendorf tube for immunoprecipitation. Anti-FLAG magnetic beads (Sigma M8823) were rinsed three times in lysis buffer, and the equivalent of 15 μL of beads was added to each sample. Sample and bead mixtures were incubated at 4°C with end-over-end rotation for 90 min, then washed three times with lysis buffer and resuspended in Tris-Glycine SDS sample buffer containing 10% 2-mercaptoethanol. Samples were then heated to 95°C for 3 min to elute proteins for subsequent immunoblotting.

QUANTIFICATION AND STATISTICAL ANALYSIS

CRISPR screens

For each Illumina read, constant sequences of lentiCRISPRv2 vector were trimmed away via Cutadapt.[49] The number of reads perfectly mapped to each sgRNA sequences was counted via Bowtie2.[47,48] Top performing genes targeted by multiple sgRNAs enriched in the sorted population were ranked using the model-based analysis of genome-wide CRISPR-Cas9 knockout (MAGeCK) method.[50] For some peptide samples belonging to certain batches of CRISPR screens, bench-top contamination of pre-cloned sgRNAs were observed. We did not annotate the MAGeCK hits considered to be due to contamination in the CRISPR screen MAGeCK plots. The full MAGeCK results for each screen are shown in Tables S4 and S7.

GPS-peptidome screens

For each Illumina read, constant sequences of the GPS6.0 vector were trimmed away via Cutadapt.[49] For each bin, the number of reads perfectly mapped to each peptide was counted via Bowtie2.[47] After normalizing the differences in sequencing depth for each bin, the PSI of each peptide was calculated as a weighted average of its distribution across the 6 bins, leading to a PSI score in between 1 (most unstable) and 6 (most stable):

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

where i denotes the bin (from 1 to 6) and Ri denotes the fraction of peptide reads in that particular bin.

Two replicates were performed for the control GPS-peptidome screen, and one was performed for the MLN4924 treated GPS-peptidome screen. We focused our analysis on ~260K peptides with >25 reads across all three samples (Table S2). Non-filtered data are available in Table S1.

An MLN4924-ΔPSI score was calculated by subtracting the peptide PSI in the absence of MLN4924 from the peptide PSI in the presence of MLN4924.

GPS-peptidome mutagenesis screens

Perfectly mapped peptide reads were used for the calculation of the PSI of each peptide. ΔPSI of each peptide mutant was generated by subtracting the PSI of the WT from the PSI of the mutant. Heatmaps were generated using ggplot2[56] for R and showed the ΔPSI of each mutant. The darker the color, the more stabilized the mutant. The mutagenesis data details were included in Table S3 for scanning mutagenesis and Table S6 for saturation mutagenesis.

Prediction of peptide PSI from composition

Ten percent of the ~260K peptidome data was used as the training set. For each peptide, the number of occurrences of each amino acid was counted as a feature. The training set, which included 20 features (count of 20 amino acids) and one output variable PSI, was then fed into the SVM-based machine learning algorithm for regression learning using the Caret package in R[54]. The learned algorithm was then used to predict the PSI of the entire ~260K peptidome data as the testing set. Degron index of each peptide was then calculated by subtracting observed PSI from predicted PSI.

Protein-peptide docking using AlphaFold2

For each protein-peptide docking, the full sequence of the cognate CRL adaptor and the 28-mer degron peptide were used as input for AlphaFold2 prediction. AlphaFold2-multimer prediction was performed on ColabFold v1.5.2 to generate 5 predicted unrelaxed docking structures with default parameters.[37] For each docking prediction, the highest scoring predicted structure is shown and illustrated using ChimeraX.[51] We examined all 5 predicted structure models and emphasize residues for which the intermolecular interactions are stable across models as described in the text.

DegronID: Scoring peptide:degron similarity

We performed saturation mutagenesis on 250 28 amino acid sequences from the human peptidome to characterize their degron motifs. Data for 19 motifs were ruled insufficient due to lack of data for multiple point mutations. Data for an additional 12 motifs were ruled insufficient due to poor degron mapping – for example, no individual mutation produced a substantial ΔPSI stabilization above 1. We proceeded to process data for the 219 remaining motifs. We then made further use of the mutagenesis ΔPSI values determined from these mutagenesis experiments to predict degron motif similarity to other sequences in the human peptidome. We use a custom implementation of a paradigm for scoring position specific scoring matrices.[57] More specifically, we use the position specific scoring matrices from our saturation matrices to generate a numerical prediction of the potential of any other sequence of interest to contain a closely related degron. Under the assumption that each amino acid position contributes independently to the degron, the basic structure of the predicted degron score for a given sequence is termed the ΔPSI summation score (S) and is calculated as

$$S = \sum_{i}^{20} \sum_{j}^{28} w_{ij} p_{ij} \Delta PSI_{ij}$$

Where i indicates one of the 20 standard amino acids, j indicates position along the mutagenized 28-mer, ΔPSI_{ij} is the experimentally determined stability measure for that amino acid substitution at that position, p_{ij} is equal to 1 if the queried amino acid sequence has amino acid i at position j and is equal to 0 otherwise, and w_{ij} is an additional weight described below.

Position weighting. Although the mutagenesis was performed over 28-mer peptide tiles, the sequence interval corresponding to a particular degron may by much shorter, often between three and ten amino acids in length. To quantify this, we devise a "position importance score" to represent how important each amino acid position is to the integrity of the degron. Each column of the mutagenesis scoring matrix has scores s for each amino acid $a \in \{a_1 \ldots, a_{20}\}$. "Position importance score" is then calculated for each column of the mutagenesis matrix as $c_j = \sum_{i=1}^{20} s_{a_i}$, which is then normalized to the maximum position importance score for any column of that matrix. A degron-relevant interval is determined (purple bracket) by the shortest continuous interval containing all amino acid positions with at least 0.5 conservation score, with two additional amino acids flanking on each side when possible.

Amino acid weighting. In addition to position-based weighting, we give weight to particularly disruptive amino acids even if substitution occurs at an otherwise lowly conserved position. For each matrix, a per-amino acid "amino acid importance score" is calculated for each row analogously to "position importance score" for each column above. Prior to calculating amino acid weighting, the 28-column saturation mutagenesis matrix is cropped to the degron-relevant interval described in the paragraph above.

A composite weight considering both position conservation and amino acid disruption w_{ij} is used in the final scoring and is defined as the maximum value of the separate position and amino acid weights. That is:

$$w_{ij} = \max(w_i, w_j)$$

We use this scoring paradigm to query the human peptidome for sequences most similar to the degrons represented in our 250 saturation mutagenesis experiments. For the analysis reported here, we focus on sequences represented in our 260k peptide library, but it is possible to query much larger databases without much added computational burden.

DegronID: Hierarchical clustering

Degron prediction scores were generated for all amino acid sequences of the appropriate length from the human peptidome. We mapped our 260k library to matching k-mer sequences from the human peptidome and assigned each 28mer the score of the best scoring k-mer contained within. For 198 high-confidence degron footprints, 200 top predicted hits were chosen as the 200 28mer peptides with the lowest ΔPSI summation score.

Hierarchical clustering of degron footprints from saturation mutagenesis was performed using ward agglomeration and Euclidean distance to group degrons based on the overlap in the top 200 DegronID hits between pairs of degron footprints. The dendrogram was cut into 40 small clusters, then manually reassembled into larger meta-clusters. DegronID analyses were performed in R version 4.1.2 with help of the O2 High Performance Compute Cluster, supported by the Research Computing Group, at Harvard Medical School. Analyses were conducted using the tidyverse[53] collection of R packages, along with Biostrings, aplot, and ggdendro (see key resources table).

DegronID: Benchmarking validation

Amino acid frequency at each position of the D box motif was calculated from aligned D box sequences from the APC/C degron repository.[20] This amino acid frequency matrix was used in place of a saturation mutagenesis matrix for input to DegronID. The 200 peptides from the human peptidome library that DegronID scores as most similar to the D-box motif are identified and their stability is compared to the library as a whole. For C-terminal motifs, we performed DegronID using our previously published saturation mutagenesis motifs as input.[14]

Secondary structure predictions

For the saturation mutagenesis motifs, we use JPred4[19] to predict the secondary structure of each peptide. For each of the 28-mer sequences, we concatenate the 25 N-terminal and 20 C-terminal flanking residues from the GPS construct, then run JPred4 predictions in bulk using the web server (https://www.compbio.dundee.ac.uk/jpred/). The consensus predictions were extracted, then cropped to the degron-relevant interval for each peptide as defined above. For each degron, we extract the length of degron sequence that is predicted to be an alpha helix or beta sheet by JPred4.

ADDITIONAL RESOURCES

Visualizations of degron footprints, peptide stability information, and DegronID clustering are publicly available via the DegronID Data Browser: https://brandonsie.shinyapps.io/DegronID/.

Molecular Cell, Volume *83*

Supplemental information

Elucidation of E3 ubiquitin ligase specificity through proteome-wide internal degron mapping

Zhiqian Zhang, Brandon Sie, Aiquan Chang, Yumei Leng, Christopher Nardone, Richard T. Timms, and Stephen J. Elledge

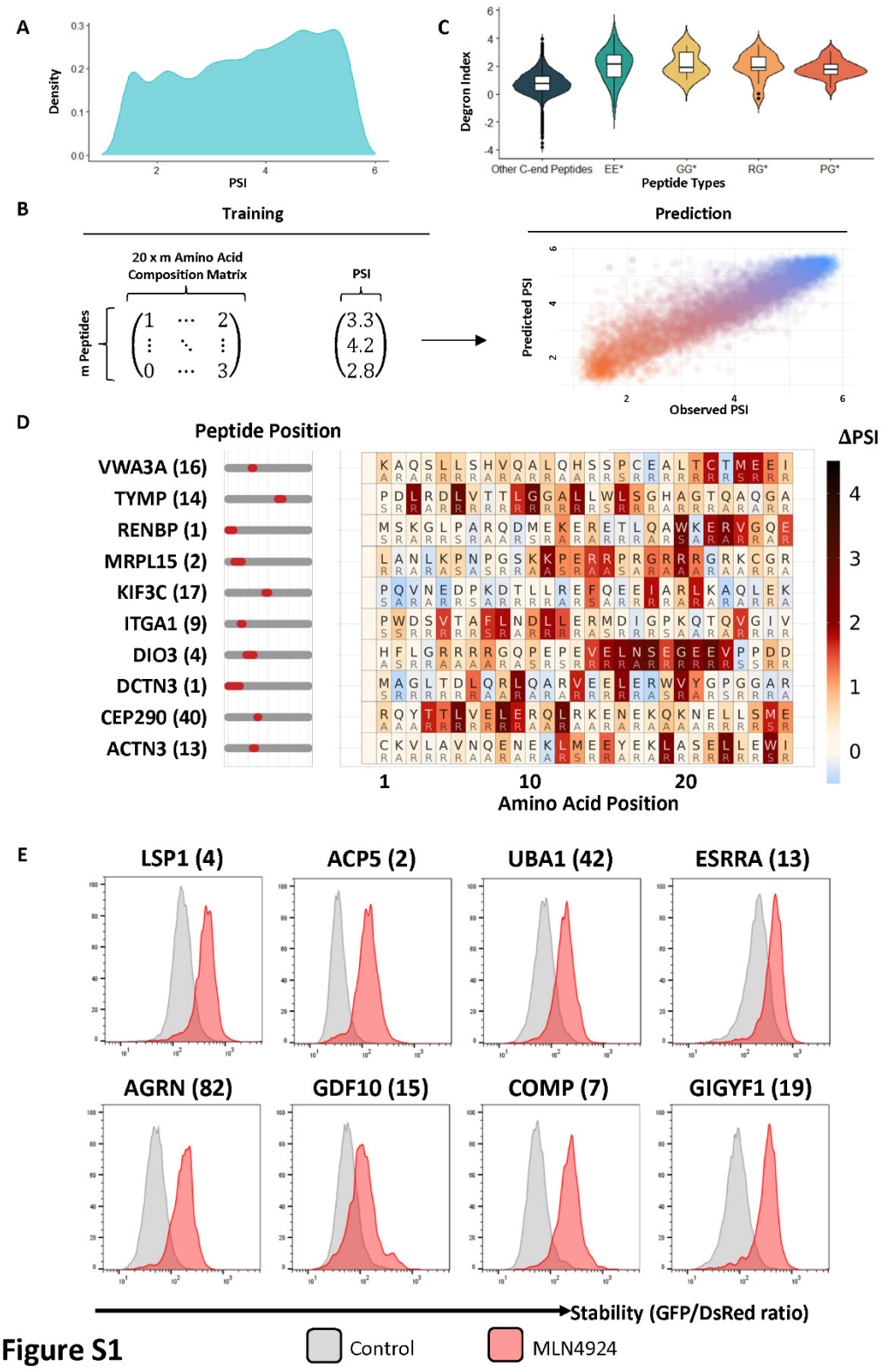

A
Density
PSI
C
Degron Index
Other C-end Peptides
EE*
GG*
RG*
PG*
Peptide Types
B
Training
20 x m Amino Acid Composition Matrix
m Peptides
PSI
Prediction
Predicted PSI
Observed PSI
D
Peptide Position
VWA3A (16)
TYMP (14)
RENBP (1)
MRPL15 (2)
KIF3C (17)
ITGA1 (9)
DIO3 (4)
DCTN3 (1)
CEP290 (40)
ACTN3 (13)
Amino Acid Position
ΔPSI
E
LSP1 (4)
ACP5 (2)
UBA1 (42)
ESRRA (13)
AGRN (82)
GDF10 (15)
COMP (7)
GIGYF1 (19)
Stability (GFP/DsRed ratio)
Control
MLN4924

Figure S1

Figure S1. GPS-peptidome screen identified many unstable peptides. Related to Figure 1, Figure 2, and Figure 3.

(A) Distribution of the PSI of individual peptides.

(B) Schematic diagram illustrating the workflow of SVM machine learning. Correlations between the experimentally observed PSIs and machine learning predict the PSI of each peptide is shown for 10K randomly selected peptides which are depicted as dots. Dots above the diagonal deviate from the predicted PSI and are candidates that may contain degrons.

(C) Population of peptides containing known degrons[14] (EE*, GG*, RG*, PG*) have significantly larger degron indexes than the distribution of the other C-end peptides from the 28-mer peptidome library. Degron index = Predicted PSI – Observed PSI.

(D) Scanning mutagenesis of representative non-CRL degron peptides with diverse degron motifs. The darker the color, the larger the degree of stabilization by the corresponding mutation

(E) Eight representative peptides identified as responsive to MLN4924 were individually examined by treatment with 1 uM MLN4924 for 24 h to assay stability as denoted by the GFP/DsRed ratio.

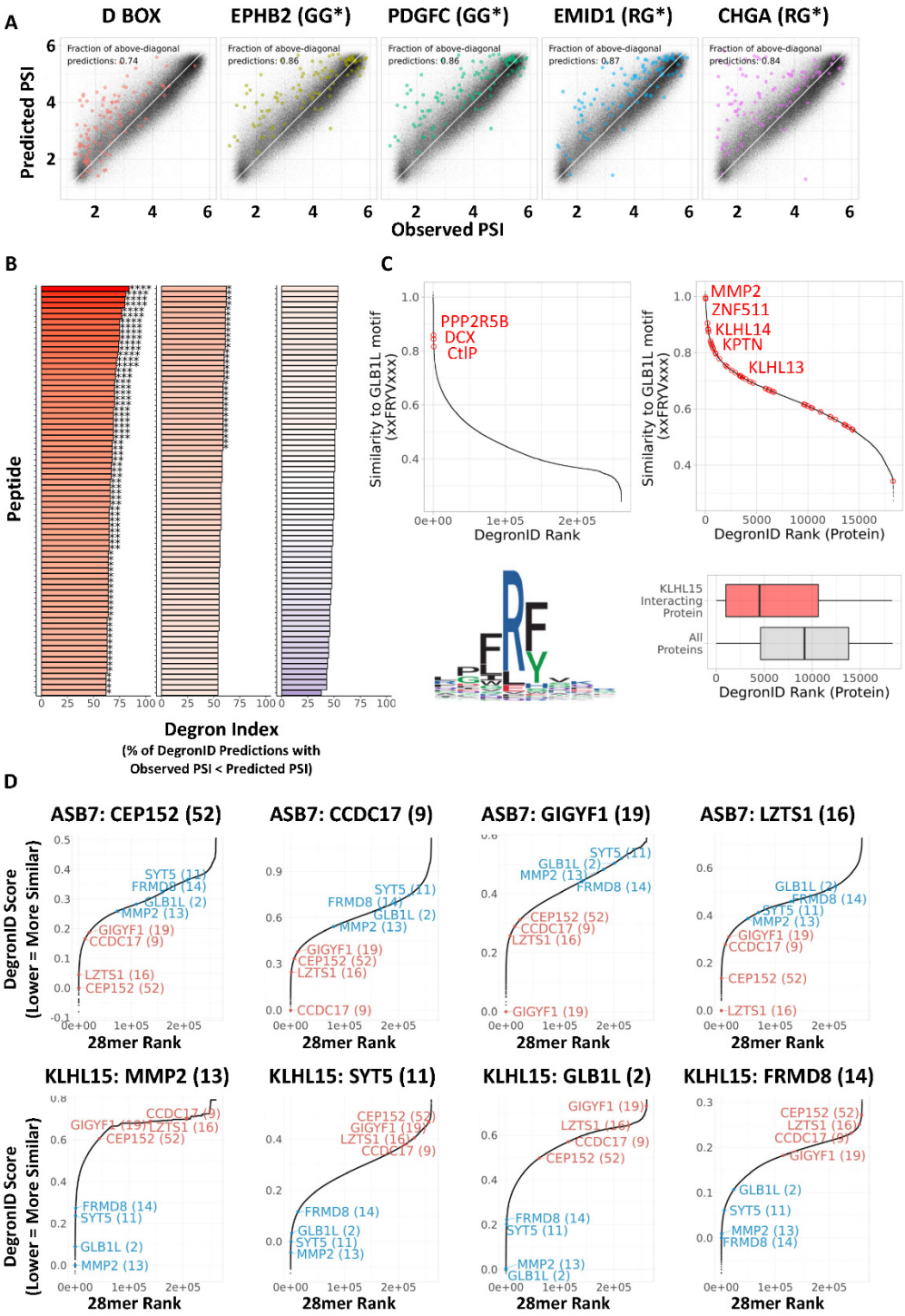

A
D BOX
EPHB2 (GG*)
PDGFC (GG*)
EMID1 (RG*)
CHGA (RG*)
Fraction of above-diagonal predictions: 0.74
Fraction of above-diagonal predictions: 0.86
Fraction of above-diagonal predictions: 0.86
Fraction of above-diagonal predictions: 0.87
Fraction of above-diagonal predictions: 0.84
Predicted PSI
Observed PSI
B
Peptide
Degron Index
(% of DegronID Predictions with Observed PSI < Predicted PSI)
C
PPP2R5B
DCX
CtIP
Similarity to GLB1L motif (xxFRYVxxx)
DegronID Rank
MMP2
ZNF511
KLHL14
KPTN
KLHL13
DegronID Rank (Protein)
KLHL15 Interacting Protein
All Proteins
D
ASB7: CEP152 (52)
ASB7: CCDC17 (9)
ASB7: GIGYF1 (19)
ASB7: LZTS1 (16)
KLHL15: MMP2 (13)
KLHL15: SYT5 (11)
KLHL15: GLB1L (2)
KLHL15: FRMD8 (14)
DegronID Score (Lower = More Similar)
28mer Rank

Figure S2

Figure S2. Degron ID classified saturation mutagenesis motifs into clusters based on their sequence similarities. Related to Figure 4.

(A) Benchmarking of the DegronID method. For D-box degrons from APC/C, along with saturation mutagenesis heatmaps from previously characterized C-terminal degrons the DegronID algorithm was performed and the 200 peptides that most closely resemble the starting motif are colored.

(B) 219 individual degron prediction models sorted by percentage of the top 200 predicted hits from each degron prediction model that are less stable than expected based on amino acid composition. Asterisks indicate FDR significance level (**** FDR < 0.0001, *** FDR < 0.001, ** FDR < 0.01, * FDR < 0.1). Significance calculated relative to a null distribution defined by repeated random sampling of 100 peptides from our library and enumeration from each of the resultant random samples what percentage of randomly selected peptides are less stable than expected based on amino acid composition.

(C) Benchmarking of the DegronID algorithm. (Top left). Substrates known to be degraded by KLHL15, marked in red, score in the top 0.2% of our library for similarity to the GLB1L(2) FRYV motif characterized in our study. (Top right). Distribution of BioPlex 3.0 KLHL15 interacting proteins compared to all human proteins in our library for similarity to the GLB1L(2) FRYV motif. (Bottom right) Boxplot showing that KLHL15 interacting proteins skew towards scoring as similar to the GLB1L(2) FRYV motif. (Bottom left) FR[FY] motif obtained from sequence alignment of KLHL15 interacting proteins, using sequence intervals from each protein that score as most similar to GLB1L(2) by DegronID.

(D) Illustration of ranked degron similarity search based on DegronID score for eight selected degrons with validated cognate E3 either ASB7 (top, red) or KLHL15 (bottom, blue). X axis indicates rank of the 260k peptides in our library, y axis indicates DegronID score for each peptide. Peptides for the eight selected degrons are colored and labelled, with color indicating cognate E3.

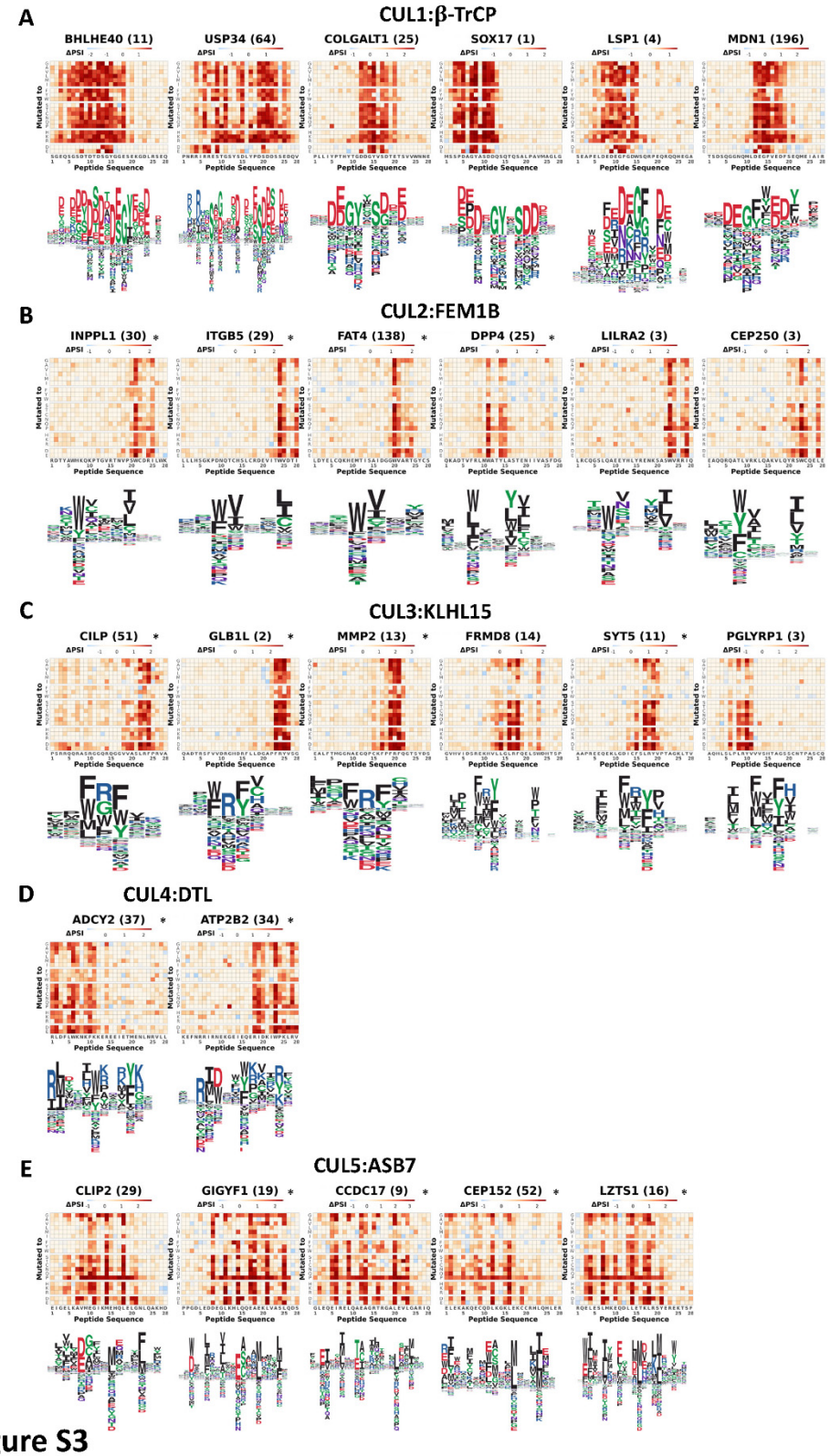

A
CUL1:β-TrCP
BHLHE40 (11)
USP34 (64)
COLGALT1 (25)
SOX17 (1)
LSP1 (4)
MDN1 (196)
B
CUL2:FEM1B
INPPL1 (30) *
ITGB5 (29) *
FAT4 (138) *
DPP4 (25) *
LILRA2 (3)
CEP250 (3)
C
CUL3:KLHL15
CILP (51) *
GLB1L (2) *
MMP2 (13) *
FRMD8 (14)
SYT5 (11) *
PGLYRP1 (3)
D
CUL4:DTL
ADCY2 (37) *
ATP2B2 (34) *
E
CUL5:ASB7
CLIP2 (29)
GIGYF1 (19) *
CCDC17 (9) *
CEP152 (52) *
LZTS1 (16) *
ΔPSI
Mutated to
Peptide Sequence

Figure S3

Figure S3. DegronID predicted peptides with similar degron motifs in the same cluster. Related to Figure 5.

For each meta-cluster indicated in Figure 4, including (A) CUL1:β-TrCP, (B) CUL2:FEM1B, (C) CUL3:KLHL15, (D) CUL4: DTL, (E) CUL5:ASB7, the mutagenesis heatmaps and logo plots of selected peptides with similar motifs were shown. Individual peptides mapped by CRISPR screening, or validated to be the substrates of the corresponding cognate E3 ligase were marked with asterisks.

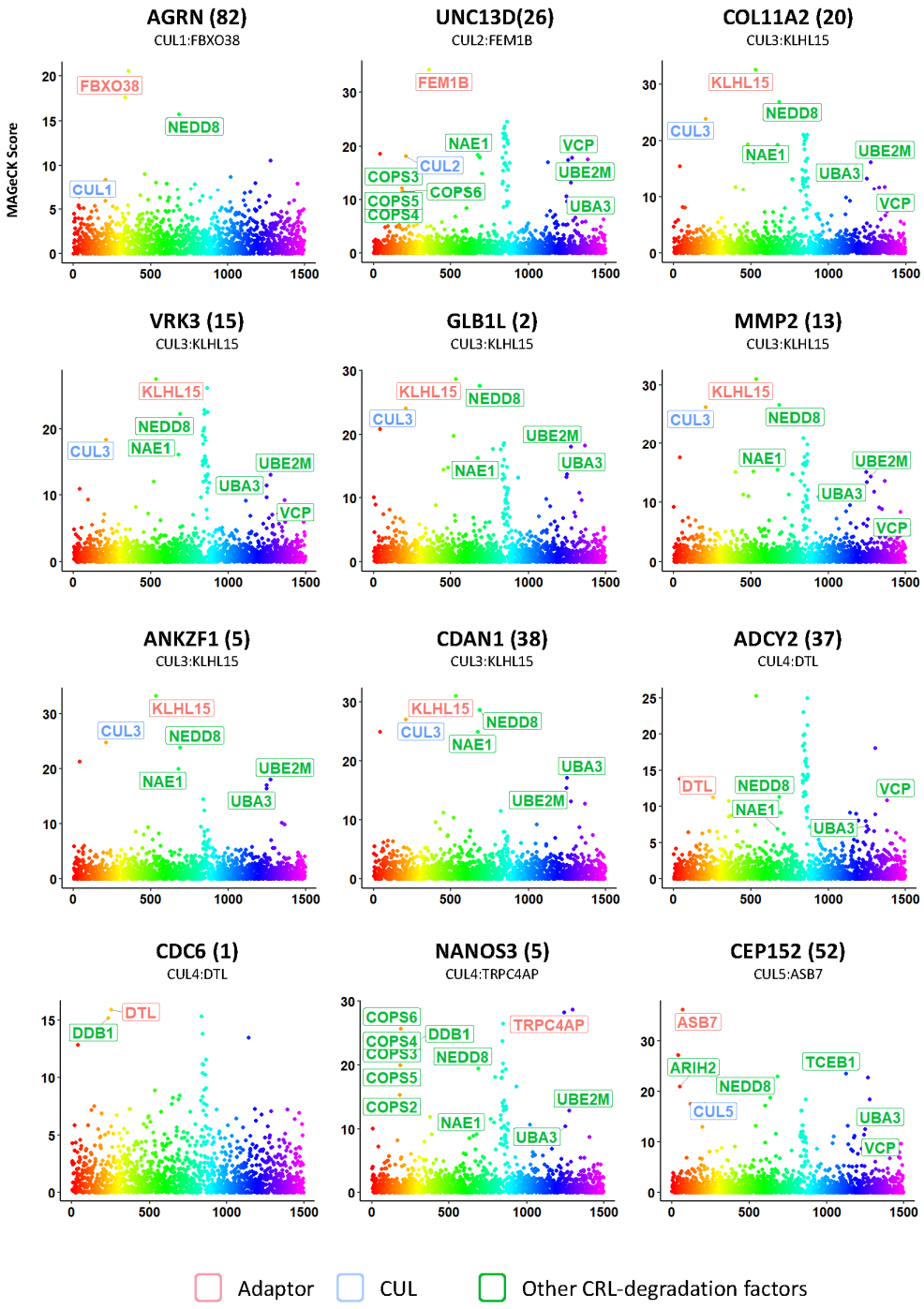

AGRN (82)
CUL1:FBXO38
UNC13D(26)
CUL2:FEM1B
COL11A2 (20)
CUL3:KLHL15
VRK3 (15)
CUL3:KLHL15
GLB1L (2)
CUL3:KLHL15
MMP2 (13)
CUL3:KLHL15
ANKZF1 (5)
CUL3:KLHL15
CDAN1 (38)
CUL3:KLHL15
ADCY2 (37)
CUL4:DTL
CDC6 (1)
CUL4:DTL
NANOS3 (5)
CUL4:TRPC4AP
CEP152 (52)
CUL5:ASB7
MAGeCK Score
Adaptor
CUL
Other CRL-degradation factors

Figure S4

Figure S4. Additional CRISPR screens that identified the cognate E3 ligases for CRL-degron peptides. Related to Figure 6.

For each indicated peptide, CRISPR screening results are shown. For the CRISPR screens, the results were first analyzed by MAGeCK and the significantly enriched Cullin-adaptor pairs were identified and shown (Table S7; see STAR Methods for details).

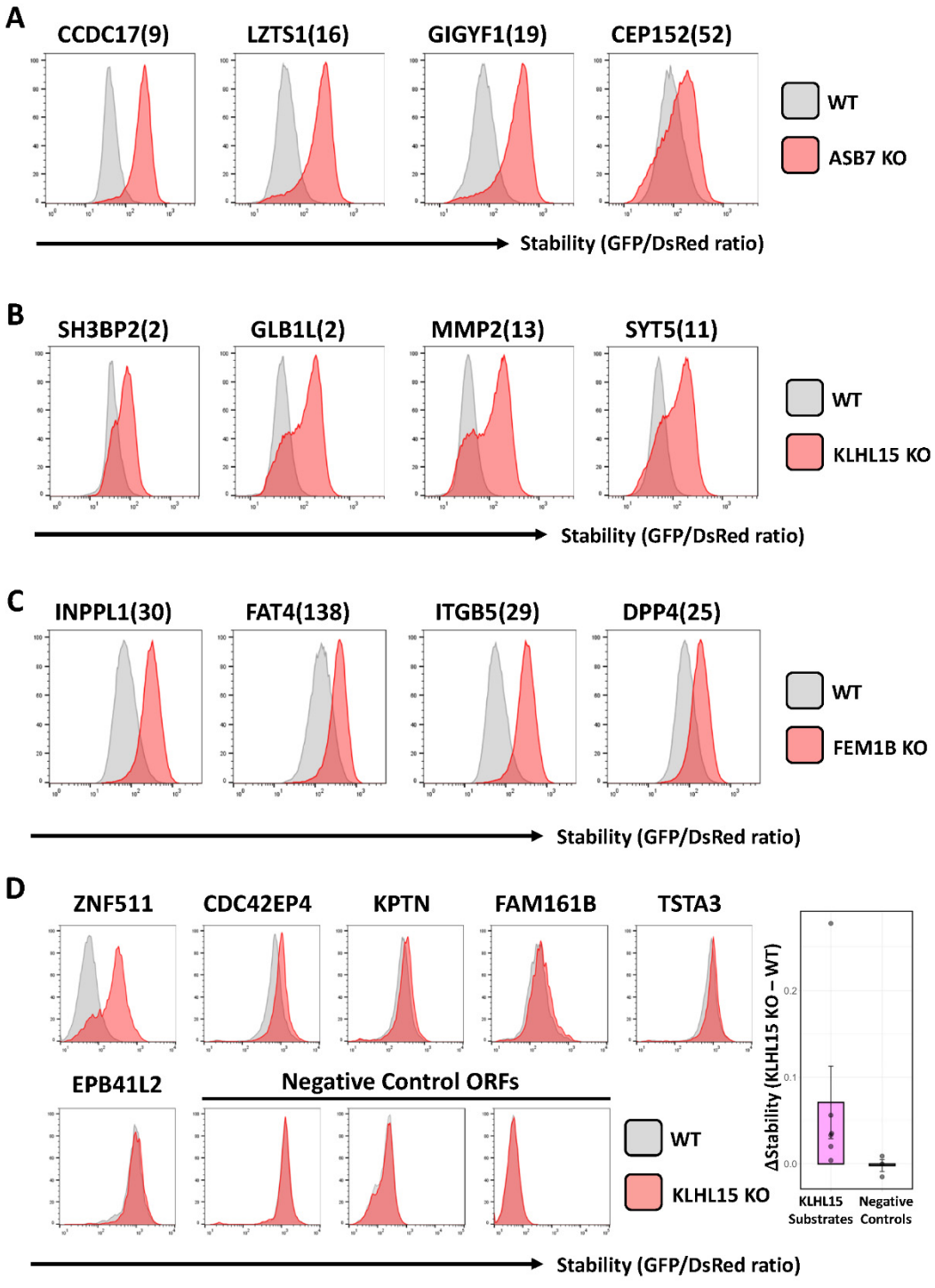
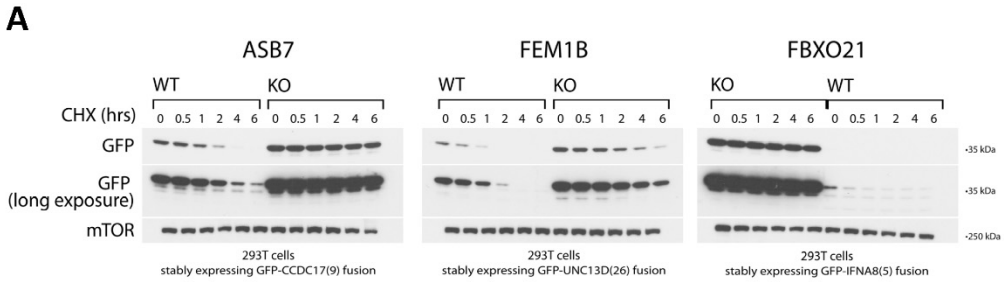

A
CCDC17(9)
LZTS1(16)
GIGYF1(19)
CEP152(52)
WT
ASB7 KO
Stability (GFP/DsRed ratio)
B
SH3BP2(2)
GLB1L(2)
MMP2(13)
SYT5(11)
WT
KLHL15 KO
Stability (GFP/DsRed ratio)
C
INPPL1(30)
FAT4(138)
ITGB5(29)
DPP4(25)
WT
FEM1B KO
Stability (GFP/DsRed ratio)
D
ZNF511
CDC42EP4
KPTN
FAM161B
TSTA3
EPB41L2
Negative Control ORFs
WT
KLHL15 KO
Stability (GFP/DsRed ratio)
ΔStability (KLHL15 KO – WT)
0.2
0.1
0.0
KLHL15 Substrates
Negative Controls

Figure S5

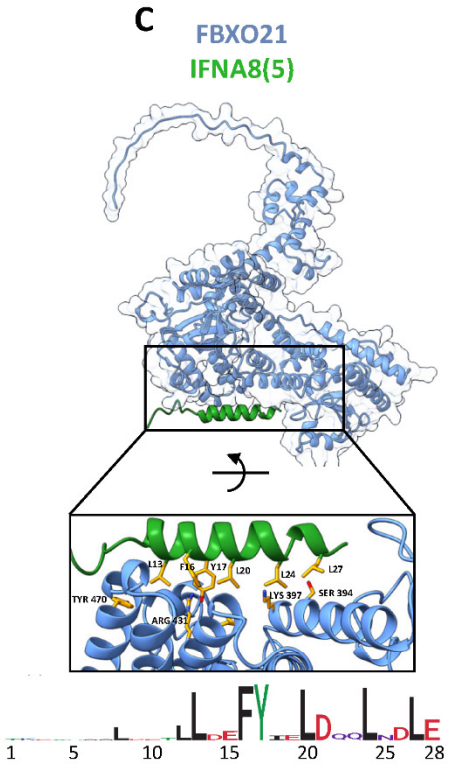
Figure S5. Selected CRL-degron peptides encoding similar motifs were validated as substrates of the same cognate E3 ligase individually. Related to Figure 6.

Peptides encoding motifs resembling that of (A) ASB7; (B) KLHL15; (C) FEM1B were individually validated as the substrate of the E3 ligase by comparing their stabilities in WT and E3 KO condition. The KO was generated by lentiviral infection of an sgRNA targeting the CRL-adaptor, followed by puromycin selection to select the cells expressing the sgRNA. (D) ORFs containing motifs for KLHL15 and identified in BioPlex 3.0 as KLHL15 interacting proteins were individually validated as KLHL15 substrates by comparing their stabilities in WT and KLHL15 KO cells. Three other randomly selected proteins lacking a FRY-like domain or any known association to KLHL15 were also measured as negative controls. In the bar plot to the right, each point represents the change in modal stability between the KLHL15 KO vs. WT cells for one ORF. Putative KLHL15 substrates are gathered on the left column, overlaying the pink bar, and negative control ORFs overlay the grey bay. The bar heights represent the mean of all points in that group. Error bars represent standard error of the mean.



B

```
ASB7:CEP152_52
   1  ELEKAKQECQDLKGKLEKCCRHLQHLER 28
      CCHHHHHHHHHHHHHHHHHHHHHHHCC
      957999999999988999999999969
ASB7:GIGYF1_19
   1  PPGDLEDDEGLKHLQQEAEKLVASLQDS 28
      CCCCCCHHHHHHHHHHHHHHHHHHCCC
      998875799999999999999998589
ASB7:LZTS1_16
   1  RQELESLMKEQDLLETKLRSYEREKTSF 28
      CHHHHHHHHHHHHHHHHHHHHHHHCCC
      848999999888999999888876789
FBXO21:IFNA8_5
   1  TKDSSAALDETLLDEFYIELDQQLNDLE 28
      CCCCHHHHHHHHHHHHHHHHHHHHCCC
      998689999999999999999997799
FBXO38:ZFHX4_40
   1  VSSERSLPEEEWRAVIGDIYQCKLCNYN 28
      CCCCCCCCHHHHHHHHCCCEEEEEEECC
      988888868999998689665568867 9
```



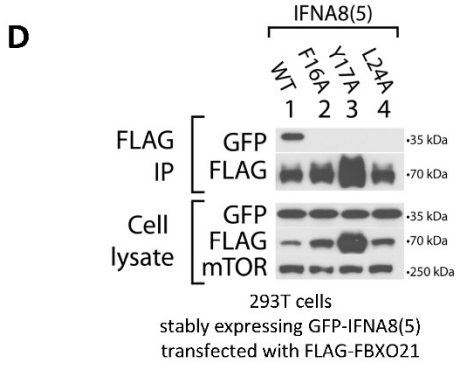


Figure S6

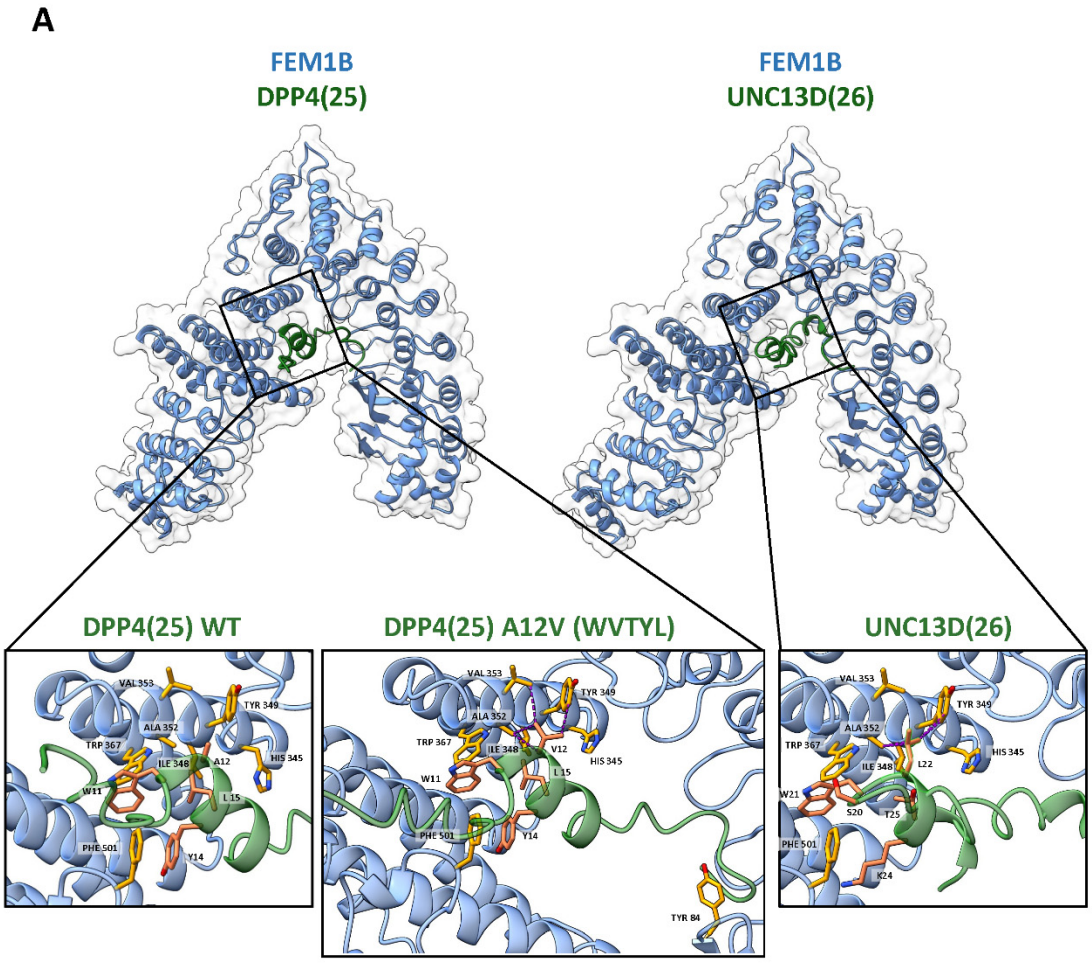
Figure S6. Half-life assay, peptides regulated by FBXO38, FBXO21 and ASB7 are predicted to encode alpha-helical structures by secondary structure prediction program PROTEUS2. Related to Figure 7.

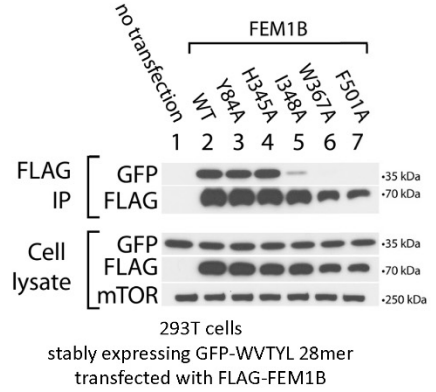
(A) Cycloheximide chase immunoblot for FBXO21:IFNA8(5), ASB7:CCDC17(9), and FEM1B:UNC13D(26).

(B) Alpha-helical structure predictions are shown for the indicated peptides.

(C) Docking structure for FBXO21:IFNA8(5) as predicted by AlphaFold2. Sidechains of critical degron residues mapped by saturation mutagenesis are shown and labelled with single-letter amino acid codes. Sidechains of potential degron-interacting FBXO21 residues are shown and labelled with three-letter amino acid codes.

(D) Co-immunoprecipitation immunoblot data for FBXO21 KO 293T cells stably expressing IFNA8(5) (WT or mutant) fused to GFP and transfected with Flag-FBXO21.


A
FEM1B
DPP4(25)
FEM1B
UNC13D(26)
DPP4(25) WT
DPP4(25) A12V (WVTYL)
UNC13D(26)
VAL 353
TYR 349
ALA 352
TRP 367
ILE 348
A12
HIS 345
W11
L 15
PHE 501
Y14
V12
TYR 84
L22
W21
S20
T25
K24


B
no transfection
FEM1B
WT
Y84A
H345A
I348A
W367A
F501A
1 2 3 4 5 6 7
FLAG IP
GFP
FLAG
Cell lysate
GFP
FLAG
mTOR
35 kDa
70 kDa
250 kDa
293T cells
stably expressing GFP-WVTYL 28mer
transfected with FLAG-FEM1B

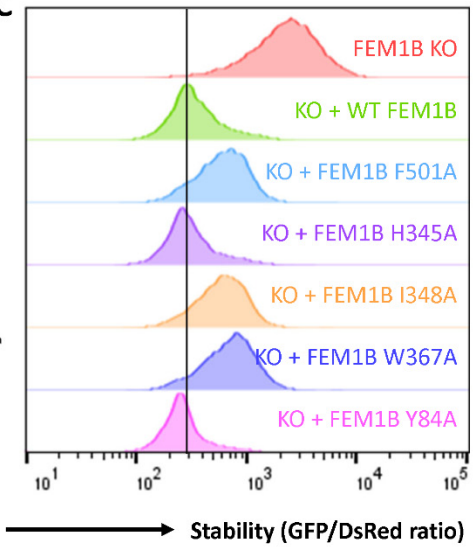

C
FEM1B KO
KO + WT FEM1B
KO + FEM1B F501A
KO + FEM1B H345A
KO + FEM1B I348A
KO + FEM1B W367A
KO + FEM1B Y84A
10^1
10^2
10^3
10^4
10^5
Stability (GFP/DsRed ratio)

Figure S7

Figure S7. E3-degron docking by AlphaFold2 identified critical degron residues consistent with that revealed by saturation mutagenesis, co-immunoprecipitation, and GPS. Related to Figure 7.

(A) Docking structures for FEM1B to identified degron peptides from DPP4(25) and UNC13D(26). In the zoomed-in windows, candidate critical residues on both the ligase and degron are highlighted. For DPP4(25), docking for WT peptide (…WATYL…) and optimized mutant peptide (…WVTYL…) are shown. Shown degron residues are labelled with single-letter amino acid codes. Sidechains of FEM1B residues that were mutagenized in validation experiments are shown and labelled with three-letter amino acid codes. Purple dashed lines indicate atomic contacts made by FEM1B with DPP4(25) A12, WVTYL 28mer V12, or UNC13D(26) L22. Contacts were identified through ChimeraX as intermolecular pairs of atoms with VDW overlap > -0.4Å, which produced similar results to using atomic center-center distance < 4.0Å.

(B) Co-immunoprecipitation immunoblot data for FEM1B KO 293T cells stably expressing an optimized FEM1B degron peptide (…WVTYL…) fused to GFP and transfected with Flag-FEM1B (WT or mutant).

(C) Flow stability data for GFP-optimized FEM1B degron peptide (…WVTYL…) with KO, stably expressed WT or stably expressed mutant FEM1B. The residues Y84 and H345 do not touch the degron and their mutations show WT stability and continue to bind the degron peptide as expected.