Multiplex Methods for Profiling and Programming Protein Degradation

Chase Cameron Suiter

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2025

Reading Committee:

Jay Shendure, Chair

Alice Berger

Lea Starita

Program Authorized to Offer Degree:

Molecular and Cellular Biology

University of Washington

Abstract

Multiplex Methods for Profiling and Programming Protein Degradation

Chase Cameron Suiter

Chair of the Supervisory Committee:

Jay Shendure

Department of Genome Sciences

This thesis focuses on the development of multiplex experimental methods for profiling and programming intracellular protein degradation. In addition, it explores the synergistic application of deep-learning models for understanding and engineering protein function at scale.

Chapter 1 provides an overview of the mechanisms governing proteome homeostasis and outlines the technological gap that has historically limited our ability to study these systems. I review how the convergence of multiplex cellular assays and deep learning models offers a new paradigm for dissecting endogenous regulatory networks and engineering novel protein functions.

Chapter 2 presents COMET (COmbinatorial Mapping of E3 Targets), a pooled assay for mapping E3 ubiquitin ligases to their substrates. By coupling combinatorial libraries of dual-fluorescent reporters with E3-targeting CRISPR guides, we enabled the many-by-many measurement of E3-dependent changes in protein abundance within a single experiment. I apply COMET to identify substrates of SCF complex E3 ubiquitin ligases as well as map the E3s mediating degradation of short-lived transcription factors, revealing that proteolytic regulation is often characterized by complex, many-to-many connectivity rather than simple one-to-one relationships. Finally, I demonstrate the use of deep-learning-based structural prediction models for the *in silico* validation of COMET-nominated E3–substrate pairs, pointing toward a future where computational nomination guides experimental validation.

Chapter 3 shifts from mapping endogenous degradation to programming it. Here, we demonstrate a multiplex framework for the discovery of functional *de novo* designed “proximity handles.” We designed a library of binders targeting various effector proteins and characterized them using LABEL-seq, a multiplex RNA-barcoded protein abundance assay. This approach identified hundreds of designs capable of mediating the stabilization or degradation of a target protein. We further validated a subset of designs in orthogonal assays, demonstrated handle-mediated degradation of the endogenous oncoprotein MCL1, and applied these handles to remodel mitochondrial organization. This study establishes a generalizable pipeline linking computational protein design to high-throughput cell-based readouts.

To conclude, **Chapter 4** discusses future directions at the interface of multiplex technology and deep learning, specifically focusing on the potential for closed-loop design-build-test-learn cycles and the expansion of these methods to enzyme engineering.

ACKNOWLEDGEMENTS

I feel immense gratitude to have the opportunity to write these acknowledgements. There are countless people who have contributed to my development as not only a scientist but as a functioning human, and I owe much to each of them. It is impossible to fully recognize their contributions, but I hope to come as close as possible in these acknowledgements.

Thank you first and foremost to my family. You all know this, but I love you! No family is perfect, but I would not change a single thing about any of you. To my parents, Rick and Gena Suiter, I thank you for your unconditional love for one another. Few are lucky to grow up in a home where true love is exhibited daily, and your example has greatly informed my own relationship. Dad, thank you for your fun-loving nature and level-headed approach to nearly all aspects of life. Mom, thank you for demonstrating consistent compassion for our family but also for others. Thank you for always caring for me and relentlessly supporting me in all my pursuits. I know that you have both sacrificed much to raise us kids, and I can only hope to repay your sacrifices by pursuing my own best life.

To my siblings, Rachel and Caleb Suiter, I thank you both for being my earliest collaborators. Rachel, I thank you for being my role model, and providing me a consistent source of inspiration. The first child has to learn a lot of things on their own, and I appreciate you sharing the lessons you learned with me. Caleb, I hope that I could play the same role for you that Rachel played for me. Although you are my younger brother, I saw and continue to see in you many characteristics I hope to exhibit myself. Thank you for your steadfast commitment to your craft, your ability to lock on to something and see it through to the end is unmatched. Both of you have influenced me in subtle and overt ways that have greatly informed how I carry out my day-to-day life.

The foundation laid by my family encouraged my curiosity, which was subsequently nurtured by many scientific mentors. The first was my Introductory Biology teacher at Minnesota State Community and Technical College, Marc Anderson, who I thank for suggesting I transfer to a four year university if I wanted to pursue research. Following Marc's advice, I transferred to Ole Miss where I had the great fortune of working with two incredible molecular biologists, Brad Jones and Sarah Liljegren. I thank both of them for mentoring me through my earliest experiments with great patience, encouragement and freedom, as well as their support of my scientific pursuits after graduation.

I'm fortunate to have landed in Jun J. Yang's lab at St. Jude as a research technician after graduation. It was my greatest honor to be a member of "Team NUDT15" under the supervision of Rina Nishii and Takaya Moriyama, who showed me two perfect examples of how to be a full-time scientist. They guided me through a two-year crash course on

the day-to-day processes of carrying out a scientific project, providing me exactly what I needed at each step along the way. All of this was supervised by Jun, who saw potential in me that I could not see myself. I recall a one-off conversation with him about applying to graduate school, in which he simply said “You are ready.” Jun is very good at telling you what you need to hear, as opposed to what you want to hear, and that simple endorsement was what I needed to hear at that moment. Jun and his lab set my early scientific career on solid footing, and much of what I have done in the subsequent years is due to their influence.

The Shendure lab was my first rotation at the University of Washington, and I had such a positive experience that I tried to do everything possible to avoid having to rotate in other labs. I’ve been lucky to collaborate and learn from a diverse set of incredible scientists, most notably Junhong Choi, Diego Calderon, Chengxiang Qiu, Xiaoyi Li, and Jean-Benoît Lalanne. Each of these people have had an outsized impact on the way I conduct myself as a person and as a scientist.

Much of the personality of a scientific lab comes from the example set by the PI, which explains the general excellence of the Shendure lab. Many PIs guide their students by providing an example of how to *be a scientist*. Jay is unique in that he goes beyond this by exemplifying how to *be as a scientist*. The distinction is small but important, with the later example informing the characteristics and principles of a truly great scientific investigator. I’m beyond lucky to have been mentored by Jay scientifically and personally, I cannot imagine a more ideal role model.

Finally, I am indebted to my wife, Trisha Lipson. Life is built on much randomness, through which we found each other. What my life would be without you, Trisha, is uncertain, but I am certain that such a life would be immeasurably worse. I thank you for showing me what I could be and pushing me to be the best version of myself. You have been the leading instigator of our journey, putting into action our collective dreams. My greatest hope for you is that you get everything you wish out of life, and I know you will because I know you. You know this, but I love you and I always will.

For my parents

Rick and Gena Suiter

TABLE OF CONTENTS

Chapter 1
INTRODUCTION

1.1 The role of technology in science

The set of scientific questions one could pursue is vast, and the time of an individual scientist is finite. This mismatch necessitates narrowing the space of questions to pursue, for example by focusing on topics of personal interest or perceived utility. Alternatively, some questions are discarded because we lack appropriate technologies to plausibly answer them. By extension, the development of new technology enables the reduction of previously intractable ideas to an experiment. Recent examples of large, technology-induced shifts in scientific reasoning include the Human Genome Project[1], the emergence of CRISPR-based genome editing[2], and, most recently, the development of deep-learning models of biology such as AlphaFold[3]. Beyond reshaping how scientific questions are answered, each of these advances helped redefine which questions could reasonably be asked in the first place.

1.2 Protein degradation as a fundamental layer of regulation

Cells are constantly making and destroying proteins, sometimes within minutes of synthesis[4]. This ongoing turnover is essential for maintaining proteostasis - the balance between protein synthesis, folding, trafficking, and degradation. Without it, damaged or misfolded proteins would accumulate, signaling pathways would lose temporal resolution, and cells would struggle to adapt to changing conditions.

Protein degradation serves several conceptually distinct roles[5]. First, it provides a quality control mechanism for the proteome capable of removing misfolded or otherwise aberrant proteins. Second, degradation provides temporal control of protein effects, allowing stimuli to be rapidly converted into changes in the protein abundance of transcription factors, cell-cycle regulators, and kinases. Controlled clearance of specific proteins allows termination or initiation of a response that is temporally expedient, especially relative to responses requiring transcription and translation. Third, degradation relies on a highly selective and modular recognition system - built from degrons, E3 ligases, adaptors, and autophagy receptors—that determines which substrates are targeted in which contexts, giving the cell fine-grained control over its proteome. The rest of this chapter focuses on how cells implement these tasks through distinct, but interconnected, degradation pathways.

1.3 A brief history of the protein degradation field

Eukaryotic cells use two major pathways for regulation of protein abundance, the ubiquitin-proteasome system (UPS)[6], which degrades many short-lived and regulatory

proteins, and the autophagy pathway[7], which mediates bulk and selective autophagic degradation of proteins, organelles, and other cellular material.

For much of the 20th century, cellular proteins were considered to be largely stable molecules without need for maintenance. Isotope tracing experiments in the 1930's revealed substantial retention of dietary protein in the tissues of a rat fed a diet partially consisting of isotope labeled tyrosine, demonstrating that cellular proteins are subject to turnover[8]. Experiments done in the 1950s by Christian de Duve's group discovered an acidic, intracellular organelle - the lysosome - capable of mediating non-specific protein degradation[9]. Although de Duve earned a share of the 1974 Nobel Prize in Physiology or Medicine for this work, protein turnover was still considered a housekeeping function devoid of the well defined regulatory characteristics of transcription and translation.

The field began to shift when classical pulse-chase experiments with radiolabeled amino acids showed that many intracellular proteins turn over far more rapidly than expected, and that their degradation rates can change in response to hormones, stress, or cell-cycle transitions[10]. These observations hinted that degradation might be selective and regulated, but the underlying machinery was unknown. The first experimental evidence of specific protein degradation came from a series of biochemical studies in the late 1970s and early 1980s from the labs of Avram Hershko, Aaron Ciechanover, and Irwin Rose. Working with reticulocyte extracts, they discovered an ATP-dependent proteolytic activity that could not be explained by lysosomal enzymes, and over several landmark papers they purified and characterized a 76 amino acid protein - ubiquitin - that became covalently attached to substrates destined for degradation and could be recycled after proteolysis[11–14]. These studies revealed a cascade of enzymes, which we now term E1, E2, and E3 that activate, conjugate, and ligate ubiquitin to target proteins, as well as the large multisubunit protease complex we now know as the 26S proteasome.

This work established the UPS as a second proteolytic pathway fundamentally different from autophagy. The UPS was ATP-dependent rather than acid-driven, cytosolic/nuclear rather than lysosome-localized, and, crucially, conferred specificity of protein degradation. Ubiquitin and its conjugation machinery could, in principle, be wired to recognize particular substrates in particular contexts. For discovery of the ubiquitin-proteasome system for intracellular protein degradation, Hershko, Ciechanover, and Rose shared the 2004 Nobel Prize in Chemistry.

Taken together, these historical threads transform our view of protein degradation. Instead of a single, undifferentiating protein degradation system, we now recognize at least two major, interconnected systems - the ubiquitin–proteasome system and autophagy pathway - each with its own dedicated machinery, logic, and regulatory roles. The rest of this chapter focuses on current challenges preventing our comprehensive

understanding of these systems as well as the opportunity to utilize these systems to edit proteomes for basic biology and therapeutic purposes.

1.4. Challenges in Mapping E3-Substrate Interactions at Scale

Despite the central role of E3 ubiquitin ligases in specifying which proteins are degraded, our view of intracellular protein degradation networks remains largely incomplete. One challenge preventing a comprehensive understanding of E3-specificity is the large number of combinatorial possibilities within the all-by-all matrix of substrates (~20,000) and E3s (~600)[5]. Substrate isoforms, expression levels, and post-translational modifications further complicate the scale of possible combinations. Any single experiment necessarily probes only a narrow subset of ligases, substrates, and contexts, resulting in an incomplete and biased understanding of the networks responsible for regulating endogenous protein levels.

A second challenge is that E3-substrate interactions are transient and result in degradation of the substrate protein, which renders direct or indirect interaction measurement technically difficult. Ligases typically bind substrates only long enough to transfer ubiquitin before dissociating, and by the time a substrate is heavily polyubiquitylated it may already be committed to proteasomal degradation. Classic proteomic approaches such as affinity purification - mass spectrometry[15] require stable complexes rather than short-lived catalytic contacts, while proximity-labeling methods[16] report on local neighborhoods rather than on which proteins are actually modified. Perturbation-based strategies, such as genetic disruption or inhibition of an E3[17], can identify proteins whose steady-state levels or half-lives change, but those measurements inevitably mix direct substrates with indirect downstream effects arising from altered transcription, signaling, or stress responses.

The architecture of the degradation network introduces an additional layer of ambiguity. The imbalance between the number of potential substrates (~20,000) and the number of potential E3s (~600) indicates a high level of redundancy in the system. For example, one substrate may be surveilled by more than one E3, and one E3 may surveil diverse substrates or surveil substrates in a context-specific manner. In such a many-to-many system, loss of any individual E3 may produce subtle or context-specific effects on a substrate, whereas strong phenotypes upon perturbing a particular ligase may reflect its position as a regulatory hub rather than a small, well-defined substrate list. Technical constraints compound these issues. Quantitative proteomics is powerful but not yet routine at the scale of hundreds of perturbations, while scalable reporter-based assays often rely on overexpressed, tagged substrates that may not fully recapitulate endogenous behavior.

Together, these features explain why, decades after the discovery of ubiquitin, we lack a comprehensive and context-aware map of E3-substrate relationships. They also motivate approaches that explicitly couple defined E3 perturbations to quantitative,

multiplex measurements of substrate stability, with the goal of moving beyond one-off one-to-one E3-substrate mappings and towards a more global view of the intracellular degradation network.

1.3 Induced proximity and targeted protein degradation

Induced-proximity strategies exploit a simple principle: many enzymes will act on a substrate if they are brought into close and sustained contact, even if that interaction is rare or nonexistent under normal conditions[18]. Proteolysis-targeting chimeras (PROTACs), bifunctional compounds composed of juxtaposed E3 and target binding molecules connected by a linker are one example of this concept. Simultaneous engagement of E3 and target forms a ternary complex, resulting in ubiquitination and degradation of the target protein. PROTACs offer benefits relative to small molecule inhibitors, including **1)** removal rather than simple inhibition of the target protein, **2)** a catalytic mechanism meaning one PROTAC molecule can degrade multiple target molecules, and **3)** the potential to target 'undruggable' proteins such as those without obvious enzymatic active sites.

Similar logic underlies molecular glues[19], which do not have an explicit "linker" but instead stabilize or create interactions between an E3 and a target that are otherwise weak or transient. Either case demonstrates a potential path towards enabling programmable control of the proteome through induced effector-target proximity. This concept has encouraged broader efforts to recruit other effectors - phosphatases, kinases, deubiquitinases, or autophagy receptors - to chosen targets, with the ultimate goal of rewiring proximity to engineer protein fate and ultimately cellular phenotypes.

Despite their appeal, current induced-proximity approaches face substantial constraints[18]. On the chemical side, most clinically advanced PROTACs rely on a narrow set of well-liganded E3 ligases, particularly CRBN and VHL, because suitable ligands with acceptable affinity, selectivity, and pharmacokinetic properties are available. Many other E3s lack such ligands, and many potential targets still have no drug-like binders at all. Even when both ends are ligandable, productive degradation depends sensitively on ternary complex geometry: linker length, flexibility, and orientation of the two binding surfaces can determine whether ubiquitin transfer is efficient or whether the complex is essentially inert. As a result, PROTAC development often involves extensive empirical optimization, and failure modes are not always mechanistically clear.

Biologically, the outcome of induced proximity is coupled to cellular context. The effect of recruiting a particular E3 depends on its expression level, native substrate load, subcellular localization, and regulation in a given cell type. A PROTAC that performs well in one lineage may be ineffective or promiscuous in another. Forced engagement of an E3 with new substrates can also perturb its normal functions, potentially rewiring signaling or transcription in ways that go beyond the intended target. Resistance mechanisms such as loss of the recruited E3, alterations in ubiquitination machinery, or

compensatory changes in degradation pathways have already been observed in experimental systems[18] and are likely to emerge in the clinic.

Protein-based induced-proximity tools, including fusion proteins that tether targets to E3 ligases, deubiquitinases, or autophagy adaptors, offer additional flexibility but introduce new challenges. For example, they are large, require intracellular expression or delivery, and can themselves perturb the systems they are meant to probe. Nonetheless, taken together, small-molecule and protein-based induced-proximity strategies demonstrate that proximity is a powerful handle on protein fate. A goal for the field - and for this thesis - is that fully exploiting this principle will require a deeper, more systematic understanding of effector networks and scalable ways to discover new recruitment handles that operate predictably across different targets, pathways, and cellular contexts.

1.5 The intersection of multiplex experimental methods and deep learning models

Deep learning refers to a subset of machine-learning methods that are trained on large datasets of example inputs (*e.g.* protein sequences) with known outputs (*e.g.* protein structures)[20]. The trained model can then take as input a new example and return as output a prediction. Predictive and generative deep learning models of biology have recently become pervasive. Models for predicting a wide range of biological phenotypes have recently been developed, all of which share several common features that enabled their progress including: **1)** development of enabling computational hardware (GPUs), **2)** availability of massive, well-curated datasets (*e.g.* the Protein Data Bank and large scale sequencing repositories), and **3)** modern deep learning architectures. Together, these developments have produced models that can reason across a wide range of biological inputs and provide non-trivial predictions of biological properties. In contrast to predictive models, generative models are capable of creating novel molecules[21]. Such models provide a way to computationally explore the design space beyond the scale of multiplex assays alone.

What opportunities exist at the intersection of deep learning and high-throughput experimentation? One example is the use of multiplex assays to generate the large-scale datasets required to train deep learning models. Conversely, trained models can guide experimental design by nominating elements of interest that are most likely to be informative, thereby focusing limited experimental capacity. Beyond this one-way flow of information (*i.e.* experiment → model or vice versa), a greater opportunity lies in the closed-loop integration of experimentation and modeling within a design-build-test-learn cycle. Here, many model-nominated elements are characterized in a single multiplex assay, and the resulting data are fed back to the model to improve subsequent iterations. Such cycles may enable a shift from predicting the behavior of biological systems to actively steering them - cellular states under explicit design objectives rather than passively describing existing ones.

1.6 References for Chapter 1

1. Lander, E. S. et al. Initial sequencing and analysis of the human genome. Nature 409, 860–921 (2001).
2. Pacesa, M., Pelea, O. & Jinek, M. Past, present, and future of CRISPR genome editing technologies. Cell 187, 1076–1100 (2024).
3. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589 (2021).
4. Damgaard, R. B. The ubiquitin system: from cell signalling to disease biology and new therapeutic opportunities. Cell Death Differ. 28, 423–426 (2021).
5. Pohl, C. & Dikic, I. Cellular quality control by the ubiquitin-proteasome system and autophagy. Science 366, 818–822 (2019).
6. Ciechanover, A. & Schwartz, A. L. The ubiquitin-proteasome pathway: the complexity and myriad functions of proteins death. Proc. Natl. Acad. Sci. U. S. A. 95, 2727–2730 (1998).
7. Mizushima, N. & Komatsu, M. Autophagy: renovation of cells and tissues. Cell 147, 728–741 (2011).
8. Schoenheimer, R., Ratner, S. & Rittenberg, D. Studies in protein metabolism. J. Biol. Chem. 127, 333–344 (1939).
9. de Duve, C., Pressman, B. C., Gianetto, R., Wattiaux, R. & Appelmans, F. Tissue fractionation studies. 6. Intracellular distribution patterns of enzymes in rat-liver tissue. Biochem. J. 60, 604–617 (1955).
10. Simpson, M. V. The release of labeled amino acids from the proteins of rat liver slices. J. Biol. Chem. 201, 143–154 (1953).
11. Hershko, A., Ciechanover, A. & Rose, I. A. Resolution of the ATP-dependent proteolytic system from reticulocytes: a component that interacts with ATP. Proc. Natl. Acad. Sci. U. S. A. 76, 3107–3110 (1979).
12. Ciehanover, A., Hod, Y. & Hershko, A. A heat-stable polypeptide component of an ATP-dependent proteolytic system from reticulocytes. Biochem. Biophys. Res. Commun. 81, 1100–1105 (1978).
13. Ciechanover, A., Heller, H., Elias, S., Haas, A. L. & Hershko, A. ATP-dependent conjugation of reticulocyte proteins with the polypeptide required for protein degradation. Proc. Natl. Acad. Sci. U. S. A. 77, 1365–1368 (1980).
14. Hershko, A., Ciechanover, A., Heller, H., Haas, A. L. & Rose, I. A. Proposed role of ATP in protein breakdown: conjugation of protein with multiple chains of the polypeptide of ATP-dependent proteolysis. Proc. Natl. Acad. Sci. U. S. A. 77, 1783–1786 (1980).
15. Huttlin, E. L. et al. Dual proteome-scale networks reveal cell-specific remodeling of the human interactome. Cell 184, 3022–3040.e28 (2021).
16. Qin, W., Cho, K. F., Cavanagh, P. E. & Ting, A. Y. Deciphering molecular interactions by proximity labeling. Nat. Methods 18, 133–143 (2021).
17. Rusilowicz-Jones, E. V., Urbé, S. & Clague, M. J. Protein degradation on the global scale. Mol. Cell 82, 1414–1423 (2022).
18. Békés, M., Langley, D. R. & Crews, C. M. PROTAC targeted protein degraders: the past is prologue. Nat. Rev. Drug Discov. 21, 181–200 (2022).
19. Cao, S. et al. Defining molecular glues with a dual-nanobody cannabidiol sensor. Nat. Commun. 13, 1–14 (2022).
20. Sapoval, N. et al. Current progress and open challenges for applying deep learning across the biosciences. Nat. Commun. 13, 1728 (2022).
21. Watson, J. L. et al. De novo design of protein structure and function with RFdiffusion. Nature 620, 1089–1100 (2023).

Chapter 2
COMBINATORIAL MAPPING OF E3 TARGETS

This chapter reproduces previously published work with minimal changes

Suiter, C. C.†#, Calderon, D.†, Lee, D. S., Chiu, M., Jain, S., Chardon, F. M., Lee, C., Daza, R. M., Trapnell, C., Zheng, N., & Shendure, J#. (2025). Combinatorial mapping of E3 ubiquitin ligases to their target substrates. *Molecular Cell*.

† denotes equal contribution; # denotes corresponding authorship

Contributions: C.C.S. and J.S. conceptualized the COMET method with input from D.C. C.C.S. designed all assays and carried out all screens and subsequent validation experiments. C.C.S. and D.S.L. generated AlphaFold-Multimer predictions. M.C. and S.J. assisted with validation experiments. F.M.C. generated the stable K562-Cas9 cell line. C.L. provided assistance with DNA sequencing and guidance around FACS data collection and interpretation. R.M.D. helped develop the tagmentation-based subassembly method. D.C. wrote computational scripts for associating DNA barcodes and ORFs and for processing COMET screen data. C.C.S. analyzed results with inputs from D.S.L., D.C., and J.S. C.C.S., D.C., and J.S. wrote the initial draft of the manuscript with input from all other authors. C.C.S. and J.S. wrote subsequent drafts of the manuscript with input from all other authors. C.T., N.Z., and J.S. supervised the project.

2.1 Abstract

E3 ubiquitin ligases (E3s) confer specificity of protein degradation through ubiquitination of substrate proteins. Yet the vast majority of the >600 human E3s have no known substrates. To identify proteolytic E3-substrate pairs at scale, we developed COMET (COmbinatorial Mapping of E3 Targets), a framework for testing the role of many E3s in degrading many candidate substrates within a single experiment. We applied COMET to SCF ubiquitin ligase subunits that mediate degradation of target substrates (6,716 F-box-ORF combinations) and E3s that degrade short-lived transcription factors (TFs) (26,028 E3-TF combinations). Our data suggest many E3-substrate relationships are complex rather than 1:1 associations. Finally, we leverage deep learning to predict the structural basis of E3-substrate interactions, and probe the strengths and limits of such models. Looking forward, we consider the practicality of transposing this framework, *i.e.* computational structural prediction of all possible E3-substrate interactions, followed by multiplex experimental validation.

2.2 Introduction

Rapid and precise control of cellular protein levels permit cells to regulate their homeostatic state[2] and to respond to changing environments[3]. Such control of intracellular protein degradation is mediated by the ubiquitin-proteasome system (UPS). Ubiquitination is a post-translational covalent modification that can serve roles in transcription, DNA repair and signaling. However, it is most deeply understood as a mark for degradation by the UPS. The UPS mediates protein degradation via E3 ubiquitin ligases (E3s), which provide specificity of degradation by selecting specific substrate proteins[4]. Although there are >600 human E3s[5], the target substrate(s) for the vast majority of E3s remain unknown, partly due to the challenging biochemistry and limited scalability of methods for assessing E3-substrate pairs.

E3-substrate interactions are dynamic and substrates may be rapidly degraded upon ubiquitination[6,7]. Some high-throughput methods to identify E3 targets have been developed, most notably Global Protein Stability (GPS) profiling[8], which leverages a dual-fluorescent reporter to quantify protein abundance on a proteome-wide scale. GPS has been applied to identify substrates of the cullin-RING ligase (CRL) family of E3s, multi-subunit ubiquitin ligases with an interchangeable substrate receptor for modular target specificity[9]. Each GPS experiment relies on a single perturbation that broadly inhibits CRL function, *e.g.* overexpression of dominant-negative cullin fragments or drug-mediated inhibition of the entire proteolysis pathway. A consequence is that GPS lacks specificity to determine the contribution of individual CRL subunits to substrate degradation. With >600 E3s and ~20,000 possible substrates per E3, these limitations curtail the potential of GPS to comprehensively dissect the landscape of proteolytic regulation in human cells.

2.2 Results

2.2.1 A combinatorial screening method for detecting target substrates of E3 ligases

We sought to develop a high-throughput method for screening combinatorial libraries of E3 perturbations and potential substrates. To this end, we adopted a GPS-inspired dual-fluorescent reporter[10] expressing a GFP-fusion protein, which represents the putative ligase substrate, and an mCherry reporter translated from an internal ribosome entry site (IRES) (**Figure 1A**). Expression of the GFP-IRES-mCherry is controlled by the doxycycline-inducible tetracycline response element (TRE). Since the GFP-fusion and mCherry are co-expressed, the GFP:mCherry ratio, hereafter referred to as 'protein abundance', reflects stability of the GFP-fusion protein to degradation. To multiplex this assay, we sequentially clone DNA libraries of E3-targeting CRISPR gRNAs, human ORFs and an ORF-linked DNA barcode (**Figure S1A-F**). A combinatorial COMET plasmid pool contains thousands of gRNA-ORF pairs, with each pair representing a potential E3-substrate interaction (**Figure 1B**).

To facilitate multiplex screening, we generated monoclonal HEK293 and K562 cell lines constitutively expressing reverse tetracycline-controlled transcriptional activator (rtTA) and Cas9, hereafter termed HEK293-rtTA-Cas9 and K562-rtTA-Cas9 (**Figure S1G-H**). The COMET library is integrated into cells at low multiplicity of integration (MOI), such that each cell reports on the abundance of a specific ORF in the presence of a specific gRNA (**Figure 1C**). These cells are sorted using FACS into equal-partition bins based on the GFP:mCherry ratio. Amplicon sequencing is used to quantify gRNA-ORF pairs in each bin. For any given ORF, the read distribution across bins can be compared between cells bearing non-targeting control (NTC) vs. E3-targeting gRNAs, with differences indicating perturbation of that E3 impacts the protein-level abundance of that ORF (**Figure 1D**). In principle, this strategy enables many-by-many testing of E3-substrate pairs.

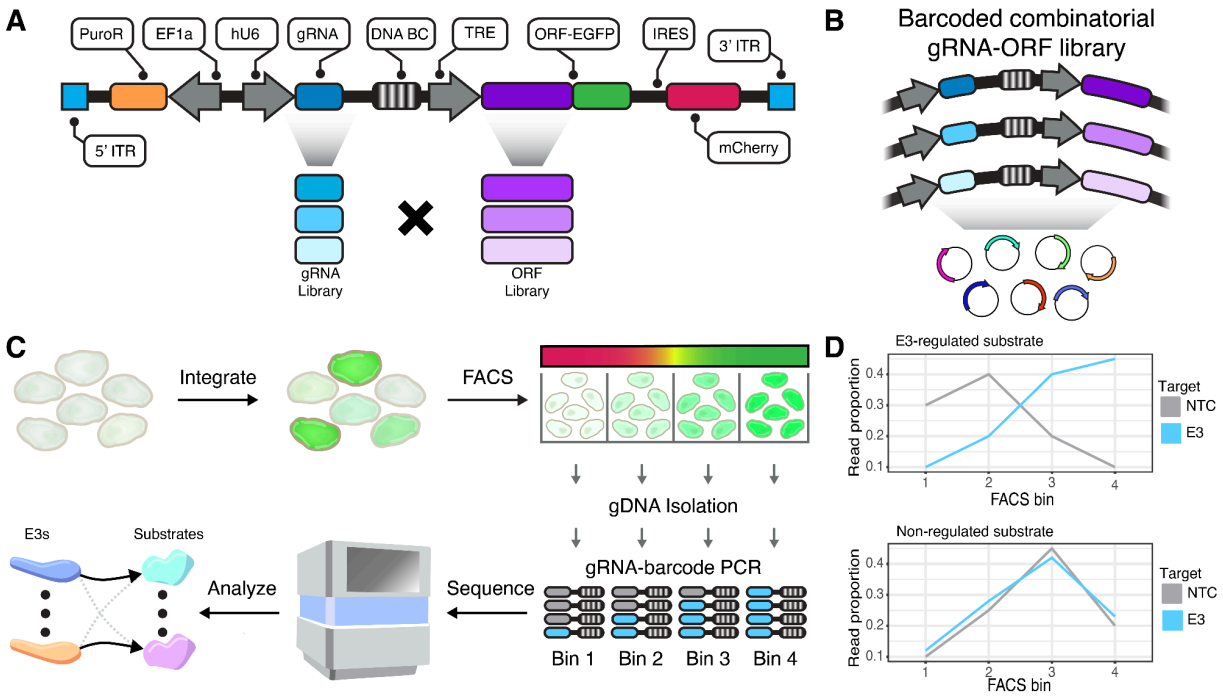

Figure 1. Combinatorial mapping of E3 ubiquitin ligases to their target substrates.

A) Schematic of COMET construct. **B)** Schematic of COMET library where each plasmid represents a unique gRNA-ORF combination. **C)** Cells harboring integrated COMET libraries are sorted on the ratio of GFP:mCherry. Amplicon sequencing of gRNA-barcode pairs measures the relative abundance of perturbation-ORF pairs in each of the four bins. **D)** Illustrative distributions of read counts across 4 FACS bins for an E3-regulated substrate (top panel) and a non-regulated substrate (bottom panel)

2.2.2 Applying COMET to F-box proteins and SCF-linked substrates

For proof-of-concept, we focused on the well-characterized SCF complex[9] (**Figure 2A**). The SCF is the founding member of the CRL superfamily of multisubunit E3s and is composed of a CUL1 scaffold upon which the rest of the complex is assembled. RBX1 and a ubiquitin-charged E2 bind the C-terminus of CUL1, while the N-terminus is occupied by an adaptor protein, SKP1, that mediates interactions with ~70 F-box proteins[11]. Importantly, the SCF is a modular protein degradation system. F-box proteins are interchangeable, with each recruiting a unique set of substrates for ubiquitination, thus diversifying and specifying SCF substrate ubiquitination.

To scalably identify which F-box proteins mediate degradation of which substrates, we cloned a COMET library (**Figure 2B**) in which each plasmid encodes: 1) an F-box-targeting gRNA, 2) a candidate substrate-GFP fusion ORF, and 3) an ORF-linked DNA barcode. We targeted 68 F-box genes[11], core SCF components CUL1, SKP1, and RBX1, and SCF regulators NEDD8 and CAND1, with 3 gRNAs per gene, and also included 23 NTC gRNAs. For candidate substrates, we selected 30 proteins previously annotated as SCF substrates[8], along with 62 randomly chosen proteins that were neither annotated SCF substrates nor previously tested. The 92 candidate substrates had variable endogenous expression levels in K562 cells (**Figure S2A,C**). The resulting library contains 6,716 F-box-ORF combinations (or 242 gRNA x 92 ORF = 22,264 barcoded constructs). Sequencing was used to generate a lookup table of barcode-ORF pairs (**Table S1**). 88% of the ORFs in the library had >300 barcodes, and 92% of barcodes had >90% of their reads associated with a single ORF (**Figure S3A,B**).

The SCF COMET library was integrated via piggyBac[12] transposition into either HEK293-rtTA-Cas9 or K562-rtTA-Cas9 cells (day 0). Cells with integrated constructs were selected with puromycin. Expression of the ORF-GFP-IRES-mCherry reporter was induced with doxycycline on day 10, and cells were sorted into four equally partitioned bins based on the GFP:mCherry ratio on day 12. Genomic DNA was isolated from each bin, from which the gRNA-barcode region was PCR amplified and sequenced to track gRNA-ORF pair proportions across the four bins.

2.2.3 Robust measurement of baseline protein abundance

To evaluate whether our assay reliably measured baseline protein abundances, we focused on data generated from K562 cells, and calculated a Protein Stability Index (PSI) as previously described[8,13] (**Table S2**). The resulting PSI is a weighted average representing the mean bin position of each gRNA-ORF pair, with values ranging from 1 (maximally unstable) to 4 (maximally stable).

We first focused on ORFs paired with NTC gRNAs (*i.e.* unperturbed). The ORFs for BNIP3, CDC25A, FBXL14, and SLC29A3 paired with NTC gRNAs reliably had >75% of their reads falling in the low abundance bin (**Figure 2C**). In contrast, the ORFs for

DNPEP, PTDSS1, TFEB, and UMPS paired with NTC gRNAs had >75% of their reads in the high abundance bin (**Figure 2C**). The PSIs of NTC gRNA-ORF pairs were highly reproducible (Pearson's $R > 0.9$ for all pairwise comparisons; **Figure 2D**).

As COMET is a pooled experiment, we sought to validate protein abundance measurements with individually cloned ORFs. In these experiments, COMET-based PSI values correctly predicted the order of individually assayed protein stabilities (**Figure 2E**), and were highly correlated with mean-fluorescence intensity (MFI) of the GFP:mCherry ratio (Pearson's $R = 0.98$; **Figure S3E**). Altogether, these results show that COMET reproducibly measures protein abundances within the range where effects from altered protein abundance are anticipated to be detected.

2.2.4 Identification of known ligase-substrate interactions

We next sought to test our ability to detect E3 perturbations that stabilize specific substrates. We separately examined data from HEK293 or K562 cells for differences in PSI distribution for each ORF paired with F-box-targeting gRNAs versus NTC gRNAs, which we term ΔPSI ($PSI^{targeting}$ - PSI^{NTC}; **Figure S3C,D**). After filtering for poorly represented gRNA-barcode or gRNA-ORF pairs, we identified 75 and 74 E3-substrate combinations whose PSI was significantly increased in HEK293 and K562 cells, respectively ($p < 0.05$; tests are two-sided t-test unless otherwise stated; *p*-values were corrected using the Benjamini-Hochberg method; **Table S3**).

Focusing first on well-established E3-substrate pairs that we had included in the experiment, we successfully identified the founding F-box family member CCNF as controlling the abundance of SLBP[14], in both HEK293 (ΔPSI = 0.23; $p < 6e\text{-}6$) and K562 (ΔPSI = 0.47; $p < 6e\text{-}13$) cells. Knockout of FBXW7 also increased the abundance of TP53 in K562 cells (ΔPSI = 0.50; $p < 2e\text{-}3$), a pairing that has been described under physiological conditions[15]. To visualize these effects, we plotted distributions of PSIs calculated at the barcode-guide level (**Figure 3A**). This confirms that the shifts in PSI distributions for SLBP and TP53, when paired with NTC versus CCNF- or FBXW7-targeting gRNAs, are not driven by outlier barcodes. Further, PSI increases for SLBP and TP53 were consistent across four transfection replicates, and across different targeting gRNAs (**Figure 3B,C**). Finally, we individually cloned SLBP and TP53 ORFs into the COMET reporter and integrated the construct in K562-Cas9-rtTA cells. Transfection of SLBP or TP53 reporter cells with gRNA plasmids targeting CCNF or FBXW7 resulted in a 1.8-fold and 1.7-fold increase in the mean GFP:mCherry ratio, respectively (**Figure 3D**).

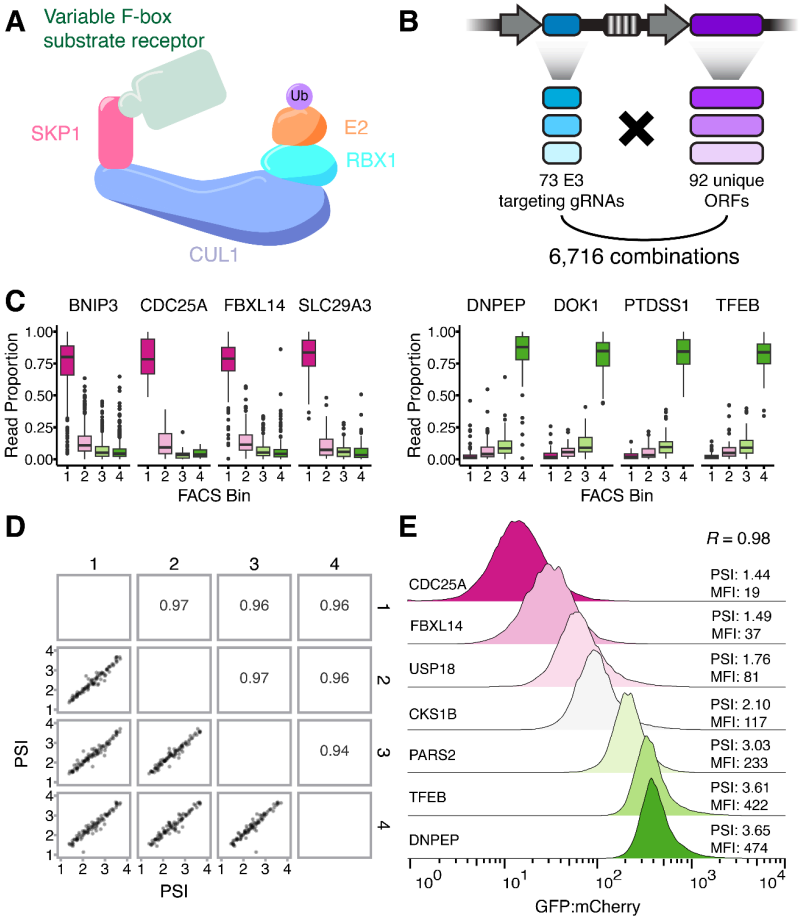

Figure 2. Robust measurement of baseline protein abundance with COMET.

A) Schematic of SCF E3 ubiquitin ligase complex. **B)** Schematic of combinatorial library of F-box-targeting gRNAs and ORFs. **C)** Examples of ORFs with consistently low (left) or high (right) abundance as estimated by the distribution of their barcodes across FACS bins. **D)** Pairwise comparison of PSI between experimental transfection replicates. Pearson's *R* shown. **E)** Validation of COMET-based PSI values with individually measured GFP:mCherry ratios for indicated proteins. Histograms display the FACS-measured GFP:mCherry ratio of cells. PSI values shown on the right are derived from COMET while MFI values reflect individually measured GFP:mCherry ratios. All data in this figure are generated from K562 cells.

2.2.5 Reproducibility within and between cell lines

For each cell line, we arbitrarily collapsed reads from four independent transfection replicates to two replicates, and recalculated *p*-values and ΔPSIs. There were 57 (HEK293) and 126 (K562) E3-substrate pairs that were significant in one or both replicates. The ΔPSI values of pairs that were significant in both replicates were highly consistent (green points; **Figure S4A,B**; Pearson's R = 0.97 and 0.98 for HEK293 and K562 respectively). Finally, considering all pairs that were significant in either replicate, we observe concordant direction of effects (HEK293: 55/ 57; $p < 2e\text{-}14$; K562: 124/126; $p < 3e\text{-}16$; binomial test).

We next examined the consistency of effects between cell lines. Across both HEK293 and K562 experiments, 209 unique E3-substrate pairs significantly shifted ΔPSI values either positively or negatively, with 46/209 pairs achieving significance in both cell lines (**Figure 3E,F**). Direction of effects were overwhelmingly concordant for shared significant associations (green points in **Figure 3E;** 45/46; $p < 7e\text{-}13$, binomial test; 19 destabilizing, 26 stabilizing). This directional concordance extended to pairs that were significant in only one cell line (orange [HEK293] and blue [K562] points in **Figure 3E**; 124/163; $p < 1e\text{-}11$, binomial test). Furthermore, the overlap in hits between cell lines was highly significant ($p < 1e\text{-}78$, hypergeometric test; **Figure 3G**). These patterns suggest that a larger proportion of the 209 pairs are shared, and might have been called as significant in both cell lines with greater experimental power.

In both HEK293 and K562, we observed a clear excess of significant *p*-values (**Figure 3H**), including numerous perturbation-ORF pairs discussed above that result in a significant PSI increase (HEK293: 75; K562: 74). Surprisingly however, we also observed many perturbation-ORF pairs that resulted in a significant PSI decrease (HEK293: 36; K562: 109). Assuming all effects are direct and E3 ligases are solely destabilizing, knocking out an E3 is only expected to increase the abundance of its target substrate, such that this is difficult to explain.

The majority of destabilizing pairs involved two ORFs, CTNNBIP1 (HEK293: 13; K562: 38) and SERINC3 (HEK293: 8; K562: 36). What might explain their recurrent destabilization? Multi-subunit E3s such as the SCF are dynamic complexes that exist in an equilibrium in which substrate receptors and substrates are constantly exchanged. It is possible that knockout of one substrate receptor disrupts this equilibrium leading to increased degradation of particular substrates. As our statistical tests detect differences between the means of the NTC and targeting gRNA PSI distributions, another possibility is that DNA damage caused by Cas9 double-strand breaks (DSBs) leads to a cellular environment in which CTNNBIP1 and SERINC3 are degraded.

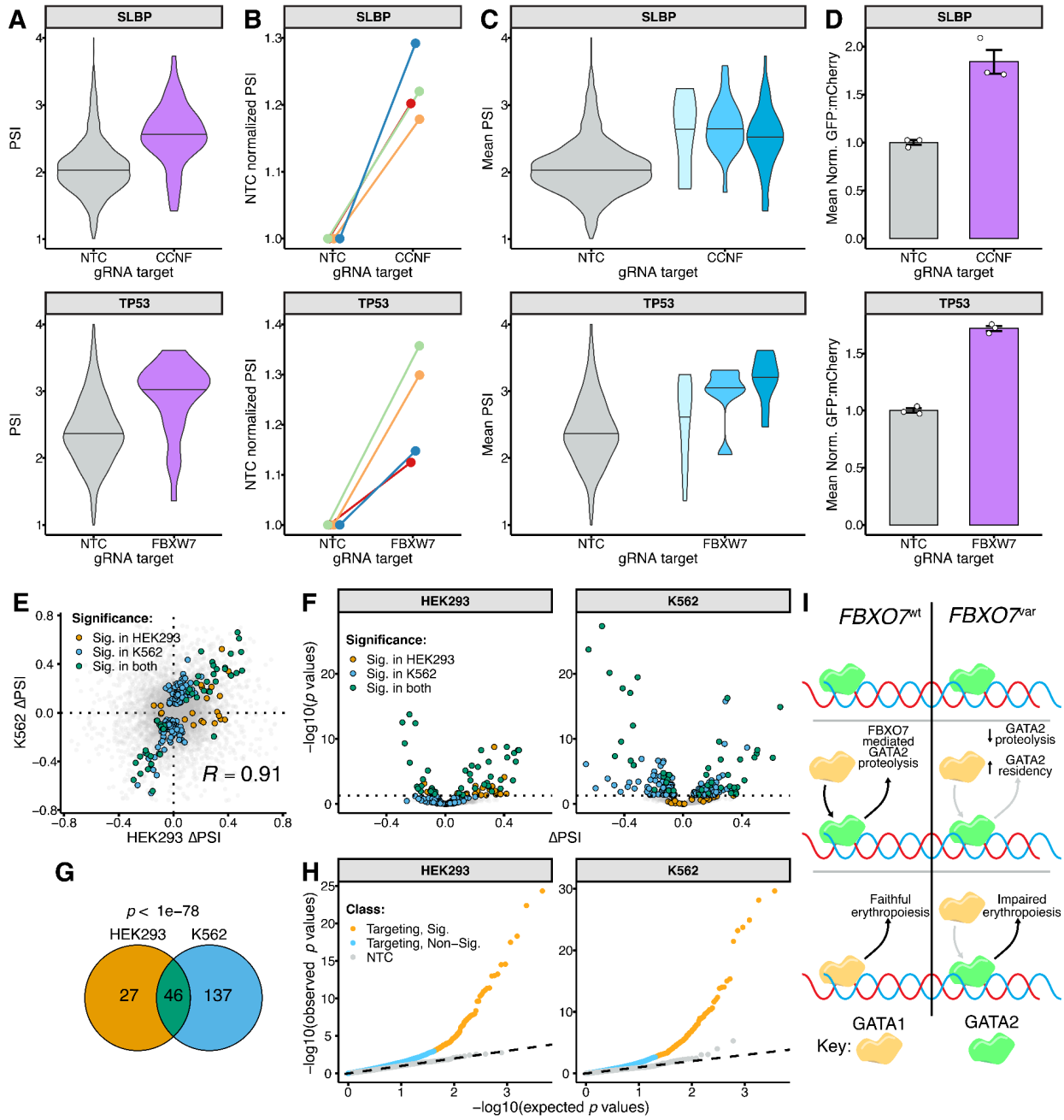

Figure 3. COMET identifies F-box substrate receptors for previously annotated SCF substrates.

A) Violin plots of PSI for SLBP (top) and TP53 (bottom) in presence of NTC or targeting gRNAs for F-box genes CCNF (top) or FBXW7 (bottom). **B)** Line plots of

NTC-normalized PSIs in each of four independent transfection replicates, as in panel **A** for CCNF-SLBP (top) and FBXW7-TP53 (bottom). **C)** Violin plots of mean PSIs in presence of NTC or three distinct gRNAs, as in panel **A** for CCNF-SLBP (top) and FBXW7-TP53 (bottom). **D)** Flow cytometry validation of CCNF-SLBP (top) and FBXW7-TP53 (bottom) effects. NTC-normalized mean GFP-mCherry ratios are shown. **E)** Comparison of ΔPSI estimates between HEK293 vs. K562**.** Pearson's *R* calculated based on subset of interactions that are significant in either or both cell lines. **F)** Volcano plot of ΔPSI versus *p*-values in HEK293 (left) or K562 (right). Points colored based on whether the corresponding E3-substrate pair was significant in HEK293 (orange), K562 (blue), both (green) or neither (gray) cell line. **G)** Overlap of significant associations between HEK293 and K562 experiments. Significance of overlap calculated with hypergeometric test. **H)** Quantile-quantile plots from the HEK293 (left) or K562 (right) showing enrichment of measured *p*-values for both significant (yellow) and nonsignificant (blue) targeting gRNAs or NTC gRNAs (gray) over the null distribution of *p*-values (dashed line). **I)** Schematic of potential mechanism of GATA2→GATA1 switch regulation by FBXO7.

We reasoned that if gain-in-activity of an untargeted substrate receptor underlies the recurrent destabilization of CTNNBIP1 and SERINC3, then these might share a common E3 knockout that stabilizes their abundance. Intriguingly, CTNNBIP1 and SERINC3 are stabilized by knockout of FBXW7 in K562 cells (as is SERINC3 in HEK293T cells). FBXW7 is a transcriptional target of the p53 TF[16]. p53 has a well-studied role as a regulator of the DNA-damage-repair pathway[17], with p53 protein levels increasing in response to DNA damage. This suggests a mechanism in which the CRISPR-Cas9 DSBs leveraged by COMET to disrupt E3s leads to p53-mediated increases in FBXW7 levels and consequent degradation of the FBXW7 substrates CTNNBIP1 and SERINC3. On the other hand, we note that common SCF component knockouts (NEDD8, SKP1, RBX1) are three of the four most significant hits for these two ORFs, in the destabilizing direction. This supports an alternative hypothesis in which the SCF plays a direct or indirect role in stabilizing these two ORFs (and possibly other ORFs).

2.2.6 Features of potentially novel ligase-substrate interactions

Among the 75 significantly destabilizing F-box-ORF pairs in HEK293 cells, there were 26 unique ORFs, of which 16 (62%) are established SCF substrates. Considering only 30/92 ORFs in this experiment were previously SCF-linked, this represents a 1.9-fold enrichment (p < 6e-6, hypergeometric test). We observed a similar result in K562 cells (26 unique ORFs among 74 destabilizing pairs, 17 (65%) of which are established SCF substrates; 2.0-fold enrichment; p < 5e-6, hypergeometric test).

Although previously implicated as SCF substrates by GPS[8], the F-box proteins responsible for their degradation were largely unknown since this information is not

recoverable from a GPS experiment. Each of these ORFs was, on average, associated with multiple perturbations (mean 2.9 [HEK293] and 2.8 [K562]). ORFs associated with 3+ perturbations were dominated by previously annotated SCF substrates (HEK293: 11/14; K562: 9/9). Because previous GPS experiments utilized a dominant negative CUL1 fragment, our results suggest screens utilizing such fragments are biased towards recovering substrates surveilled by many individual F-boxes.

Degradation of these ORFs was mediated by 22 (HEK293) and 24 (K562) unique F-box or core SCF component proteins. Core SCF components RBX1 and SKP1, and the SCF regulator NEDD8, were recurrently implicated, often paired with substrates with additional F-box hits. For example, in K562 cells, we found CKS1B PSIs were increased when any one of the core SCF components (CUL1, SKP1, RBX1) or the CRL regulator, NEDD8, were knocked out, or an additional 7 F-box substrate receptors.

One might have predicted knocking out core complex components would have resulted in more substrate associations or perhaps a superset of associations. Of the 26 ORFs with at least one significant, stabilizing association in K562 cells, only 8 had at least one significant CUL1, SKP1, RBX1, or NEDD8 hit. However, this paucity could be due in part to insufficient power consequent to the highly essential functions of these core SCF components. This interpretation is supported by our observation that targeting of SKP1, RBX1, NEDD8, or CUL1 was underrepresented among gRNAs recovered from gDNA in both cell lines (**Figure S4C,D,F,G**). However, it is challenging to conclusively interpret a lack of an association because these potential false negatives may also be attributable to unknown factors, such as buffering effects of residual protein or subunit redundancy.

FBXW7 was the most recurrently observed F-box protein among K562 hits, paired with many cell cycle related proteins (*e.g.* CDK2AP1, CDK4, CDKN1A, CKS1B), consistent with the known role of FBXW7 as a tumor suppressor. In our data, FBXW7 knockout resulted in stabilization of GATA2, a key regulator of hematopoiesis[18]. While an association between FBXW7 and GATA2 has been previously reported[19], our results suggest GATA2 may be degraded by two additional F-box proteins, FBXL16 and FBXO7. GATA2 auto-regulates its own expression during early hematopoiesis, eventually being displaced by GATA1 (“GATA switching”), leading to repression of *GATA2* expression and the installation of gene expression programs that drive erythropoiesis. GATA factor switching is driven at least partially by the drastic difference in half-lives for GATA1 (>4 hours) and GATA2 (~1 hour)[20]. Interestingly, a genome-wide association study associated *FBXO7* variants with red blood cell phenotypes, specifically an increase in mean cell hemoglobin[21]. Further, an *Fbxo7*-knockout mouse model exhibited impaired erythropoiesis and anemia with increased mean cell hemoglobin concentration[22]. We hypothesize that FBXO7-mediated GATA2 degradation enhances GATA switching by reducing GATA2 abundance and facilitating GATA1-driven

erythropoiesis (**Figure 3I, left**). Reduction or loss of FBXO7 may impair the ability of GATA1 to displace GATA2, leading to impaired erythropoiesis (**Figure 3I, right**).

2.2.7 Massively parallel testing for E3 ligases that surveil TFs with short half-lives

Our results with GATA2 led us to wonder whether proteolytic degradation of short-lived TFs is a general phenomenon. Regulation of gene expression by TF proteolysis is well-established yet understudied[23–25], with therapeutic implications as E3-TF interactions can be pharmacologically targeted for stabilization[26,27] or abrogation[28]. Leveraging a recent study that identified short-lived proteins (half-life ≤ 8 hours) across a range of cell types[29], we focused on 108 TFs and sought to ask whether their rapid turnover was attributable to proteasomal degradation, and if so, via which E3s (**Figure 4A**). Endogenous expression of these TFs in K562 cells was variable (**Figure S2B,D**).

While our initial experiment focused on perturbing F-boxes, we expanded the scope of our gRNA library to 723 gRNAs targeting 241 genes encoding components of the seven CRL families (SCF, CRL2, CRL3, CRL4A, CRL4B, CRL5, CRL7) and the anaphase-promoting complex (APC/C) ubiquitin ligase, supplemented with 50 NTC gRNAs. Altogether, we cloned a COMET library with 83,484 (773 gRNA x 108 ORFs) barcoded constructs and integrated these to K562-Cas9-rtTA cells (**Figure 4B**).

At 48 hours after reporter expression induction, cells were sorted based on the GFP:mCherry ratio, genomic DNA isolated from each of four FACS bins, and the gRNA-barcode region amplified and sequenced. Testing for significantly altered PSI distributions was performed as previously, revealing 9 TFs whose abundances were significantly increased by 13 ligase perturbations. Similar to the SCF-focused experiments, read proportions for gRNAs targeting CRL subunits were significantly different from all other gRNAs in the pool (**Figure S4E,H**). This phenomenon is less pronounced in the TF screen, which may be attributable to its expanded scale leading to less power to detect fitness effects and/or to the inclusion of more non-essential CRL subunits.

We identified PTTG1 and SOX9 as stabilized upon knockout of *FZR1*, a subunit of APC/C, a multisubunit protein assembly[30] that controls cell cycle progression by degrading cell-cycle proteins such as cyclin B[31] and PTTG1[32] (Securin) (**Figure 4C**), supporting the former and suggesting the latter as APC/C substrates, results we validated in singleton experiments We validated these results in singleton experiments (**Figure 4D**). The APC/C is known to recognize its protein substrates through linear degradation motifs called degrons, including the D-box[33] (RxxL) and KEN-box[34] (KEN) motifs. There are two canonical APC/C coactivators responsible for APC/C substrate recognition, CDH1 (encoded by *FZR1*) and CDC20. Although PTTG1 was stabilized by knockout of *FZR1*, we did not detect a significant stabilization of PTTG1 following knockout of *CDC20* (ΔPSI = 0.15, p = 0.49).

The PTTG1-encoding ORF used in our screen has an N-terminal KEN-box and a more centrally located D-box; however the SOX9 ORF contains only a D-box. It is thought that degrons occur in disordered regions of substrate proteins[35]. However, the putative SOX9 D-box degron occurs in a highly structured alpha-helix within the DNA-binding domain. SOX9 is known to contain two nuclear-localization signals[36] (NLS) and one nuclear-export signal[37] (NES). Interestingly, the SOX9 D-box directly overlaps one of these NLS signals, suggesting APC/C may serve to not only degrade SOX9 but also to alter SOX9 localization.

A quantile-quantile plot again showed an excess of signal for targeting gRNAs (**Figure 4E**). As previously, upon recalculating ΔPSIs on independent pairs of replicates, we observe strong reproducibility for both the directionality and magnitude of significant ΔPSI values (Pearson's R = 0.93; 200/207 directionally concordant; **Figure 4F**). Compared with our SCF-focused proof-of-concept, this scaled screen identified relatively few E3 perturbations that stabilize substrate abundance, with the PSIs of only 9 substrates increasing in response to 13 E3 perturbations (**Figure 4G,H**). This could result from having fewer test proteins with *a priori* evidence of being a E3 substrate, as well as reduced power due to this screen being more complex (83,484 rather than 22,264 barcoded constructs).

Once again, we also observed *destabilizing* effects, including 5 ORFs (ZNF593, CDKN2A, SLBP, TFAP4, TCEAL8) associated with >=25 destabilizing perturbations, and 7 ligases that, when knocked out, destabilized >=10 ORFs (DCAF7, ELOC, SKP2, WDR5, ASB16, SKP1, VHL). Recurrently destabilized ORFs may be sensitive to DSBs from Cas9 or some other general indirect interaction. Recurrently destabilizing perturbations may indicate a general role for the corresponding components in protein stabilization. For example, although WDR5 was included in our list of multi-subunit CRLs, it is better recognized as a core component of COMPASS complexes[38], suggesting that COMPASS complexes may serve to stabilize certain TFs.

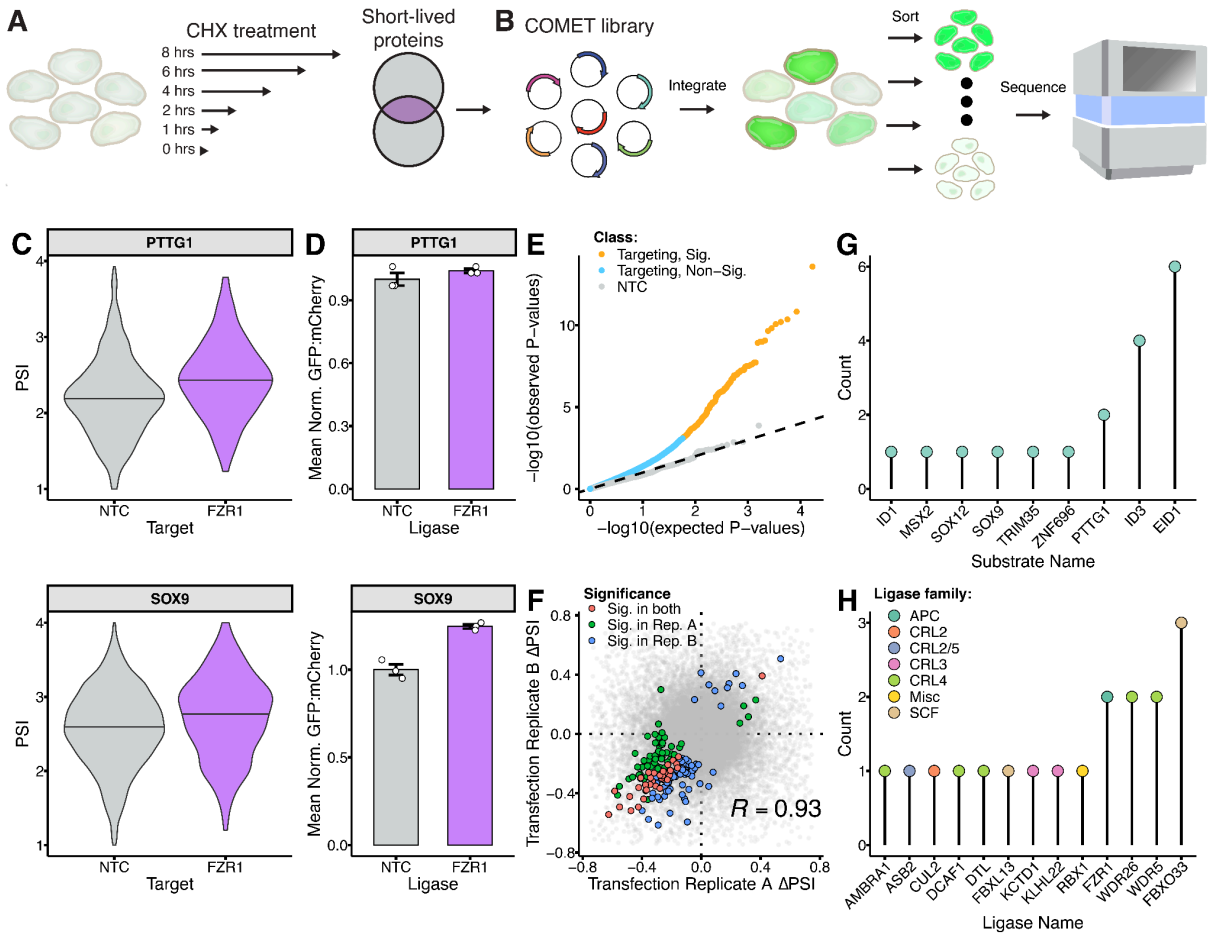

Figure 4. Screening for CRLs that degrade short-lived TFs.

A) Schematic of experiment identifying short-lived TFs[29]. **B)** Schematic of application of COMET to these TFs. **C)** PSI distributions for PTTG1 (top) and SOX9 (bottom) when paired with NTC or *FZR1*-targeting gRNAs. **D)** Validation of *FZR1* knockout effects in PTTG1 and SOX9 reporter cell lines by flow cytometry. NTC-normalized mean GFP-mCherry ratios are shown. **E)** Quantile-quantile plot showing enrichment of measured *p*-values for targeting (yellow & blue points) or NTC gRNAs (gray points) over the null distribution of *p*-values (dashed line). **F)** Reproducibility of ΔPSI estimates. For this analysis, reads from two pairs of independent transfection replicates were combined to Replicates A & B, and ΔPSIs and significance was recalculated separately for each. Points colored based on whether corresponding E3-ORF pair was significant in Replicate A (blue), Replicate B (red) or both (green), or involves a NTC gRNA (gray). Pearson correlation calculated based on subset of interactions significant in Replicate A and/or Replicate B. **G)** Lollipop plot of 9 ORFs observed among 18 significant, positive interactions. **H)** Lollipop plot of 13 E3 perturbations observed among 18 significant, positive interactions.

2.2.8 Classifying the landscape of ligase-substrate interactions

We sought to ask whether any general patterns could be discerned in the distribution of E3-substrate interactions. For this analysis, associations involving core ligase components such as RBX1 or cullin scaffolds were excluded. Filtering further for significant, stabilizing associations ($p < 0.05$, $\Delta PSI > 0$), COMET associated 48 E3s (SCF: 37, TF: 11) with 51 substrates (SCF: 42, TF: 9) for a total of 103 interactions across both experiments (SCF: 87, TF:16). We enumerated four classes of E3-substrate relationships: one-to-one, many-to-one, one-to-many, and many-to-many. For example, in our SCF HEK293 dataset, FBXW7 regulates 18 substrates including TP53, which itself is regulated by a single ligase, and so the interaction between FBXW7 and TP53 is categorized as many-to-one. With the caveat that we focused on only CRLs and did not test all possible substrates in our screens, we find examples of each class in each experiment (**Table S4**; **Figure S5**). Although we have not yet tested all E3s against all potential substrates, these results give us a preliminary sense for the connectivity of the proteolytic regulatory network.

2.2.9 Computational assessment of COMET-nominated E3-substrate interactions

COMET nominates E3-substrate pairs for which one-by-one experimental validation (*e.g.* by co-immunoprecipitation) would be challenging to scale, and potentially to interpret due to technical confounders (*e.g.* transient E3-substrate interactions or substrate degradation). As a more scalable means of orthogonal assessment, we turned to *AlphaFold-Multimer*[39,40], computing models for: 1) all 103 significantly stabilizing COMET-nominated E3-substrate pairs (class: 'COMET-linked'); 2) As negative controls, 1,000 nonsignificant E3-substrate pairs randomly selected from the sets tested in our COMET experiments (class: 'screen-negative'); 3) As additional negative controls, 1,000 pairs obtained by randomly sampling the canonical isoforms of all human proteins (class: 'true-random'); 4) As positive controls, we extracted 1,000 E3-protein pairs from the BioPlex[41] dataset (class: 'BioPlex'), which is a proteome-scale map of experimentally derived protein-protein interactions generated via affinity-purification mass spectrometry.

Altogether, we modeled 3,103 protein-protein pairs with *AlphaFold-Multimer*, with 5 individual models per pair for a total of 15,515 unique models. To systematically evaluate support for a given interaction, we extracted interchain contacts from each model. Interchain contacts were defined based on the following criteria: 1) a maximum distance of 5 Å between any two atoms in the interacting residues; 2) a predicted aligned error (PAE) of ≥10 for each residue pair involved in the contact; and 3) a predicted local distance difference test (pLDDT) score of ≥50 for both residues. This strategy was adapted from a recent *AlphaFold-Multimer*-based *in silico* screen[42].

Two metrics were calculated to summarize model quality: 1) *Inverse normalized PAE:* The mean PAE for all contacts of each individual model were calculated. These values

were min-max normalized across all four prediction classes and inverted such that 0 corresponds to the lowest, and 1 to the highest, PAE values. 2) *Normalized contact consistency:* We summarized contact consistency as the average number of models each contact was observed in (range 1-5), similarly normalizing these values such that 0 corresponds to the lowest, and 1 to the highest, contact consistency. Plotting these metrics revealed a subset of models exhibiting high confidence and consistency in the COMET-linked and BioPlex classes but not the screen-negative or true-random classes (**Figure 5A-D**; note there are 10-fold fewer models in COMET-linked class than each control class). We also calculated a summary score which gave equal weights to confidence and consistency metrics, ranging from 0 (worst) to 1 (best). Encouragingly, summary score distributions for the COMET-linked and BioPlex classes were significantly higher than both negative control classes (**Figure 5E**).

We visualized a cumulative density function for each prediction class, and heuristically set a requirement that at least 2 of the 5 models for a given protein pair pass a threshold summary score of 0.4 to be considered supported (**Figure 5F**). 16.5% (17/103) of COMET-nominated pairs met these requirements, which is 1.3-fold higher than the BioPlex pairs (12.3%, 123/1000), 2.8-fold higher than the screen-negative pairs (5.8%, 58/1000) and 5.2-fold higher than the true-random pairs (3.2%, 32/1000). In summary, the odds of COMET-nominated pairs accruing *AlphaFold-Multimer* support was similar to E3-protein pairs from BioPlex (p < 3e-1, OR = 0.7, 95% CI = 0.4-1.3), and significantly greater than either screen-negative controls (p < 3e-4, odds ratio (OR) = 3.2, 95% CI = 1.7-5.9, Fisher's exact test) or true-random controls (p < 5e-7, OR = 6.0, 95% CI = 3.0-11.6, Fisher's exact test) (**Figure 5G**).

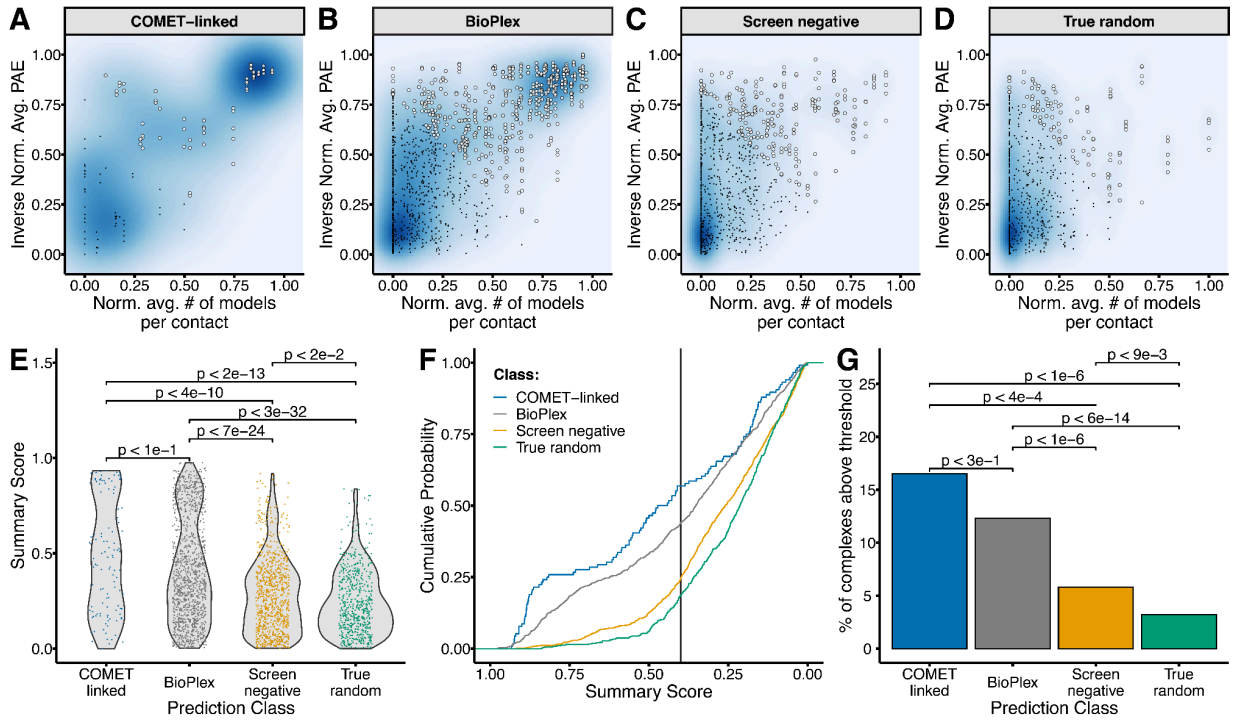

Figure 5: *AlphaFold-Multimer* modeling of COMET-nominated E3-substrate interactions

A) Scatterplot of inverse normalized average PAE values plotted against the normalized average number of contacts per residue for the 103 COMET-nominated E3-substrate *AlphaFold-Multimer* predictions. Each point represents a single model. White points indicate models with a summary score >= 0.4. Black points represent models with a summary score <= 0.4. **B)** Same as in **A**, but for 1000 E3-protein pairs from BioPlex. **C)** Same as in **A**, but for 1000 nonsignificant E3-substrate pairs randomly sampled from those tested in COMET experiments. **D)** Same as in **A**, but for 1000 randomly sampled protein-protein pairs randomly sampled from the human proteome. **E)** Violin plot showing distribution of summary scores for COMET-nominated (blue), BioPlex (grey), screen-negative (orange) or true-random (green) prediction classes. Significance calculated using Wilcoxon rank sum test. **F)** Cumulative density function showing proportion of models exceeding range of summary score thresholds for each of the four prediction classes. Vertical black line indicates the summary score threshold (0.4) that we used. **G)** Barplot of proportion of complexes with ≥2 models passing the summary score threshold for each prediction class. Significance calculated using Fisher's exact test, *p* values are adjusted with the Benjamini–Hochberg procedure.

2.2.10 Visualization of in silico models of E3-substrate interactions

We next sought to examine selected *AlphaFold-Multimer*-derived models of COMET-nominated E3-substrate interactions in more detail, beginning with the well-established interaction of PTTG1 (Securin) and the APC/C adapter protein FZR1. As discussed above, FZR1 interacts with its substrates via degrons such as the

KEN-box and D-box (RXXL) motifs. PTTG1, a known FZR1 substrate, contains KEN-box (residues 9-11) and more central D-box (RKAL, residues 61-64) motifs. To investigate this interaction further, we used *AlphaFold* to fold PTTG1 without FZR1 (**Figure 6A**; red curve). The resulting pLDDT values varied greatly across the protein, but were unremarkable near the KEN-box and D-box motifs. However, when we fold PTTG1 with FZR1 via *AlphaFold-Multimer*, we observe two striking increases in local pLDDT values that precisely coincide with the degrons (**Figure 6A**; blue curve). Furthermore, computational substrate-E3 co-folding induced a reduction of pLDDT values in all regions outside of these degrons. This result suggests the potential for co-folding of substrates with their cognate E3s to result in (relatively) increased model confidence localized to degron motifs, *i.e.* presumably corresponding to E3 engagement with the degron(s).

We also examined three COMET-nominated E3-substrate interactions that are novel, supported by structural modeling, and exhibit coincident patterns with respect to co-folding confidence and putative degron locations. In the first of these, we folded SOX12 with vs. without its COMET-associated E3, AMBRA1. Once again, we observe a relative jump in the pLDDT complex/monomer ratio, in this case a narrow C-terminal region, suggesting this region may correspond to a degron (**Figure 6B**). SOX12 and AMBRA1 have not been shown previously to interact. However, AMBRA1 was recently shown to recognize its various Cyclin D substrates through a conserved TP motif[43] and the *AlphaFold-Multimer* model contains a TP motif at the center of the putative degron (**Figure S6E,F**). Interestingly, this putative degron (residues 282-304) also overlaps the majority of the SOX12 transcriptional activation domain (residues 283-315), consistent with the observation that degron motifs within TFs often overlap the transcriptional activation domain[25].

The third example involves the cyclin-dependent kinase inhibitor CDKN1A and F-box protein FBXW5 (**Figure 6C**). Here, the CDKN1A peptide predicted to be bound by FBXW5 was confidently predicted even when CDKN1A is modeled on its own (**Figure 6C**)**.** Although the pLDDT values around the predicted interacting region are not increased as in the previous two examples, the range of high-confidence pLDDT values is narrowed upon co-folding. As previously mentioned, CDKN1A functions as a cyclin-dependent kinase inhibitor. Cyclins interact with their binding partners through Cy motifs (*e.g.* RxL or RxI) on the partner protein. CDKN1A contains one N-terminal (RRL, residues 19-21) and one C-terminal (RRL, residues 155-157) Cy motif. The C-terminal Cy motif is just outside of the predicted interacting region, raising the possibility that FBXW5 may sequester CDKN1A or otherwise compete for binding with other CDKN1A interactors. Beyond our experimental and computational association of FBXW5 to the cell cycle related CDKN1A, FBXW5 has previously been shown to degrade other cell cycle proteins such as the centriole assembly factor SAS-6[44] and the actin regulator EPS8[45].

The fourth example involves APBB1IP and FBXO21. Similar to the CDKN1A-FBXW5 interaction, the putative APBB1IP degron (residues 139-154) exhibited consistently high pLDDT values in the monomer prediction (**Figure 6D**). However, the pLDDT values of the predicted interacting residues decreased following upon co-folding with FBXO21. Of note, the putative APBB1IP degron is the only example presented here that is structured; this suggests co-folding partially disrupts the predicted secondary structure of this region leading to the observed decrease in pLDDT. A previous study demonstrated that FBXO21 is responsible for degrading the short-lived protein EID1[46] and mapped the minimal EID1 degron to residues 160-172. Alignment of APBB1IP residues 1-37 notably revealed two APBB1IP residues (F12 and L16) were conserved with EID1 and its paralog, EID2, as well as two semi-conserved, hydrophobic residues (L21 and L22) (**Figure S6G,H**). Plots visualizing PAE values for each of the above examples are provided in **Figure S6A-D**.

To test whether *AlphaFold-Multimer*-nominated degrons mediate degradation in an E3-specific manner, we cloned a series of reporters expressing full-length substrate proteins or their putative degrons fused to GFP, integrated these into HEK293-rtTA-Cas9 cells together with a construct expressing either an E3-specific or NTC gRNA, and measured GFP:mCherry ratios 14 days post-transfection. Cells co-transfected with constructs bearing an N-terminal GFP fusion of PTTG1 and an FZR1-targeting gRNA exhibited a 1.4-fold increase in their GFP:mCherry ratio relative to controls (**Figure S7A**; p = 0.009, one-sided t-test). This ratio was similar when only the degron region of PTTG1 (residues 1-84) was fused to GFP (1.3-fold increase; p = 0.018). Splitting the degron into its KEN-box (residues 1-42) or D-box (residues 42-84) components resulted in no or modest (1.1-fold) stabilization, respectively, suggesting both motifs are required. Cells expressing a C-terminal GFP fusion of SOX12 and an AMBRA1-targeting gRNA exhibited 1.2-fold stabilization relative to controls (p = 0.018) while the putative degron region of SOX12 (residues 282-304) did not (0.9-fold of original) (**Figure S7B**). Cells transfected with C-terminal GFP fusions of full length CDKN1A or the CDKN1A degron (residues 130-164) showed no or modest (1.1-fold increase; p = 0.027) stabilization, respectively (**Figure S7C**), in the presence of a FBXW5-targeting gRNA. Finally, we tested N-terminal GFP fusions of full-length APBB1IP or the putative APBB1IP degron (residues 1-40), which strikingly exhibited 1.5-fold (p < 5e-6) and 1.8-fold (p = 0.003) stabilization, respectively, in the presence of a FBXO21-targeting gRNA (**Figure S7D**). Although further investigation into the mechanisms by which these putative degrons mediate substrate proteolysis is warranted, the observation of E3-dependent, degron-mediated proteolysis in several of the cases tested (FZR1-PTTG1, FBXW5-CDKN1A, FBXO21-APBB1IP) supports the utility of leveraging *AlphaFold-Multimer* as a means of degron discovery.

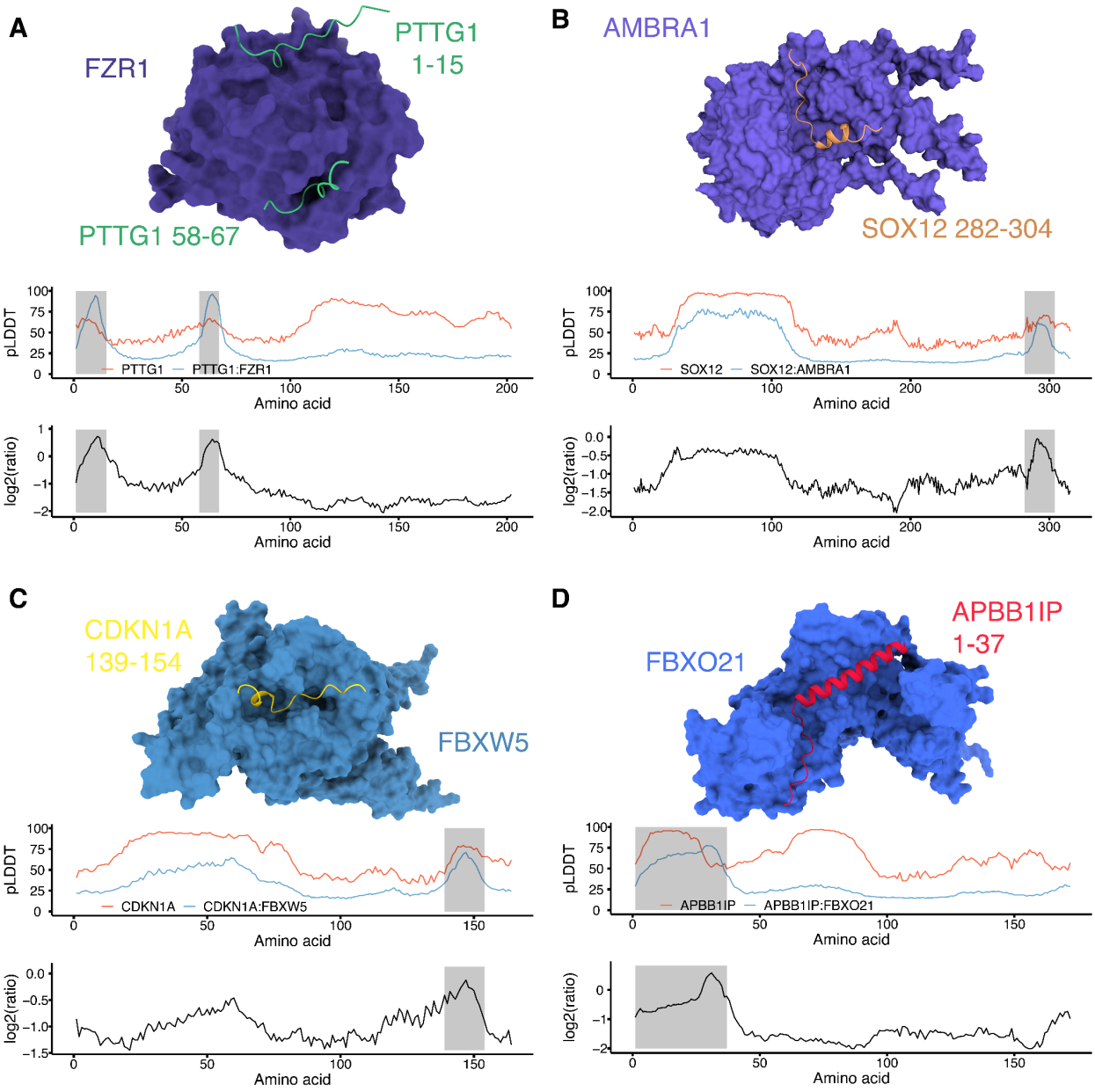

Figure 6. Computational models of E3-substrate interactions.

A-D) (Top) Visualizations of FZR1-PTTG1 (**A**), AMBRA1-SOX12 (**B**), FBXW5-CDKN1A (**C**) and FBXO21-APBB1IP (**D**) modeling are presented. E3-substrate pairs were folded together with *AlphaFold-Multimer* and interacting substrate peptides highlighted. (Middle) Line plots of pLDDT values across the substrate protein are visualized in the context of a folded substrate monomer (red) or the substrate co-folded with the COMET-nominated E3 (blue). Finally, the log2-scaled ratio (monomer/complex) of pLDDT values at each residue are plotted, with peaks indicating regions of elevated pLDDT values in the complex prediction relative to the monomer. Note that scales for log2 ratios are different in each panel, *i.e.* adjusted for each plot to maximize contrast.

2.3 Discussion

Protein degradation is an essential component of cellular regulation, but relatively few relationships between specific E3s and specific target substrates are known. Towards addressing this, we developed COMET, a combinatorial experimental framework wherein the consequences of perturbing many E3 ubiquitin ligases on the stability of each of many overexpressed candidate substrates can be tested in a single experiment. COMET is highly scalable, with tens-of-thousands of potential E3-substrate pairs assayable in a single experiment. In contrast with GPS screens which rely on dominant-negative fragments or pharmacologic inhibition, the CRISPR perturbations leveraged by COMET can be directed at non-CRL, monomeric ubiquitin ligases, *e.g.* RING, HECT, and RBR E3s[47]. Moreover, although unexplored in this study, the potential exists to use CRISPR activation of E3s in COMET screens.

By expanding the number of experimentally supported E3-substrate interactions, we anticipate that COMET may advance models of proteolytic regulatory networks (**Figure 7A**). Such links may also be valuable for targeted protein degradation, *e.g.* via proteolysis-targeting chimeras (PROTACs) or molecular glues (MGs), which hold great therapeutic promise but are hindered by the dearth of endogenous E3-substrate interactions, *e.g.* to target for stabilization with MGs (**Figure 7B**). COMET may also be useful for contextualizing genetic variants in E3s and/or substrates (**Figure 7C**). As one example, dissecting the GWAS association of FBXO7 with red blood cell phenotypes is daunting given that the phenotype could arise from FBXO7 degrading nearly any substrate protein. However, the COMET-derived linkage of FBXO7 and GATA2 suggests a plausible mechanism for the GWAS association (**Figure 3I**). Of note, a recent analysis of 33 cancer types showed 19% of mutated cancer driver genes affect protein degradation[48].

Although in principle, COMET makes it possible to test all possible E3s vs. all possible substrates, library construction and FACS would undoubtedly become bottlenecks, as testing 241 CRLs vs. ~20,000 human genes would require an experiment 185-fold larger than what we have demonstrated here. However, this challenge might be eased by upfront ORFeome-wide screens to identify bonafide proteasome or CRL substrates to be tested by COMET. For example, 1,554 proteins were recently shown to be stabilized by the pan-CRL inhibitor MLN4924, suggesting they are CRL substrates[49]. The comprehensive application of COMET to these substrates (241 CRLs x 1554 CRL substrates) would require an experiment only 14-fold larger than the short-lived TF COMET screen reported here, a challenging but approachable scale.

An alternative would be to invert the current framework (experimental nomination→computational validation) by instead first identifying candidate E3-substrate pairs through scaling of deep learning-based modeling to all possible

interactions (computational nomination→experimental validation; **Figure 7D**). In this scenario, the computational resource requirements would be dramatically larger, but given the sparseness of the set of true interactions, experimental validation could likely be achieved in a single COMET experiment. A limitation is that the false negative rate of computational modeling of E3-substrate interactions is not known, but the same could be said of experimental screening methods.

From a resources perspective, what would this take? Here we generated *AlphaFold-Multimer* predictions for 3,103 E3-substrate pairs (*i.e.* 103 COMET-linked E3-substrate pairs and 3,000 control pairs). Each prediction required ~1 GPU-hour (mean: 1.8 GPU-hr; median: 1 GPU-hr), with total compute time of ~157 GPU-days. Although the tradeoff between prediction quality and speed could be adjusted, assuming identical parameters as used here (5 models, 30 recycles, model similarity tolerance of 0.5), we estimate modeling interactions between the ~375,000 possible pairs of 241 CRLs and 1554 CRL substrates would require ~43 GPU-years, which is substantial but certainly not impossible. Expanding this to all ~600 human E3s vs. all ~20,000 human proteins (~12 million pairs) would be more challenging (~1369 GPU-years), but for perspective we note that there are already over 200 million predictions in the AlphaFold Protein Structure Database (June 2024). Furthermore, predictions reported here were generated using a combination of Nvidia A100 and L40 GPUs; simply switching to faster GPU architectures would likely result in shorter prediction times.

Here we described COMET, a combinatorial framework for testing the role of many specific E3s in degrading many specific substrates within a single experiment, and furthermore demonstrated the potential for targeted computational modeling to provide orthogonal support for experimentally nominated E3-substrate pairs. However, given the pace at which structural modeling is improving and computational resources are scaling, we envision a future in which this framework is inverted, and methods like COMET are used for the multiplex experimental validation of predictions generated by computational modeling of a vast number of potential interactions (**Figure 7D**).

2.4 Limitations of the Study

As COMET is a pooled assay, measured abundances are relative to other proteins in the pool, which makes it challenging to compare effect sizes between different experiments. This could potentially be mitigated by including protein spike-ins with known abundances to facilitate batch correction and/or by expanding the number of FACS bins.

Although our detection of some established E3-substrate pairs shows COMET can detect effects from directly interacting partners, it is possible that novel, COMET-nominated E3-substrate pairs have an indirect relationship, *e.g.* if a targeted E3 regulates an untargeted E3 which then regulates a targeted substrate. This could

potentially be mitigated by computational support for a given interaction, or by scaling the screen to target all E3s and deubiquitinases.

Finally, as with any large-scale screen, there exists the potential for both false negatives and false positives. False negatives could arise from insufficient power, ineffective gRNAs, essentiality of perturbed genes, substrates that lack the correct post-translational modifications necessary for ubiquitination, lack of E3 expression, or masking of degrons near the C-terminus by the EGFP fusion. False positives could arise from non-specific effects, *e.g.* off-target events or cellular responses induced by Cas9-induced DSBs[50].

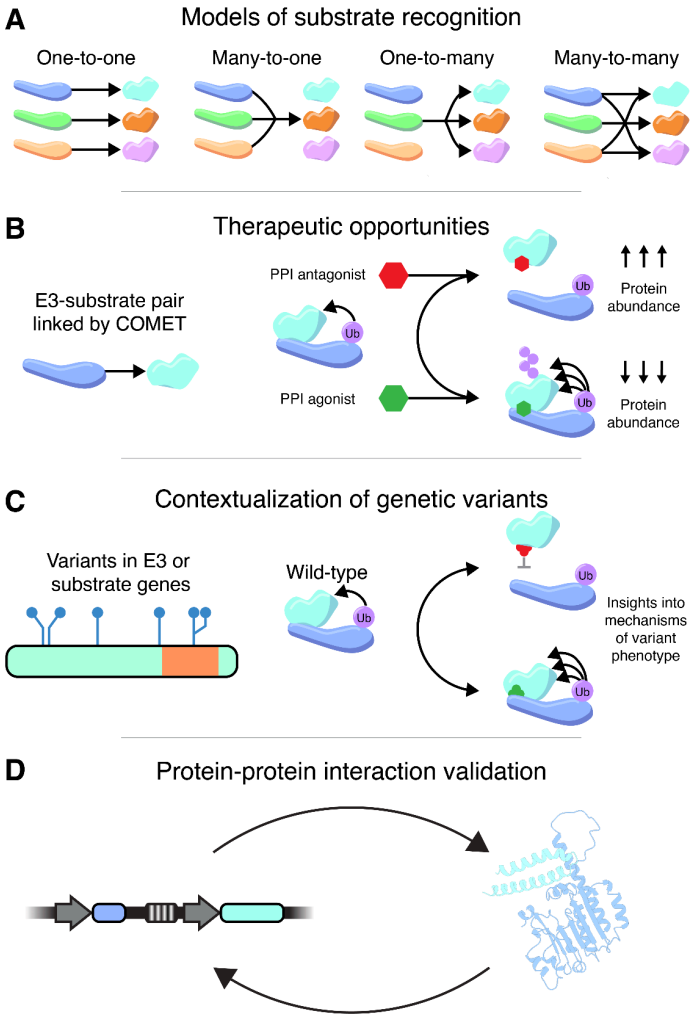

Figure 7. Potential applications of COMET.

A) By linking E3s to their substrates, COMET advances our understanding of the connectivity of regulatory networks controlling protein abundance. Visualized are four possible classifications of E3-substrate relationships based on the number of interactions. **B)** Individual E3-substrate interactions represent potential therapeutic targets for either stabilization or inhibition. **C)** E3-substrate pairs linked via COMET may provide insights into the mechanisms by which E3 or substrate variants lead to disease phenotypes. **D)** Reciprocal validation of COMET or *AlphaFold-Multimer*-nominated E3-substrate interactions.

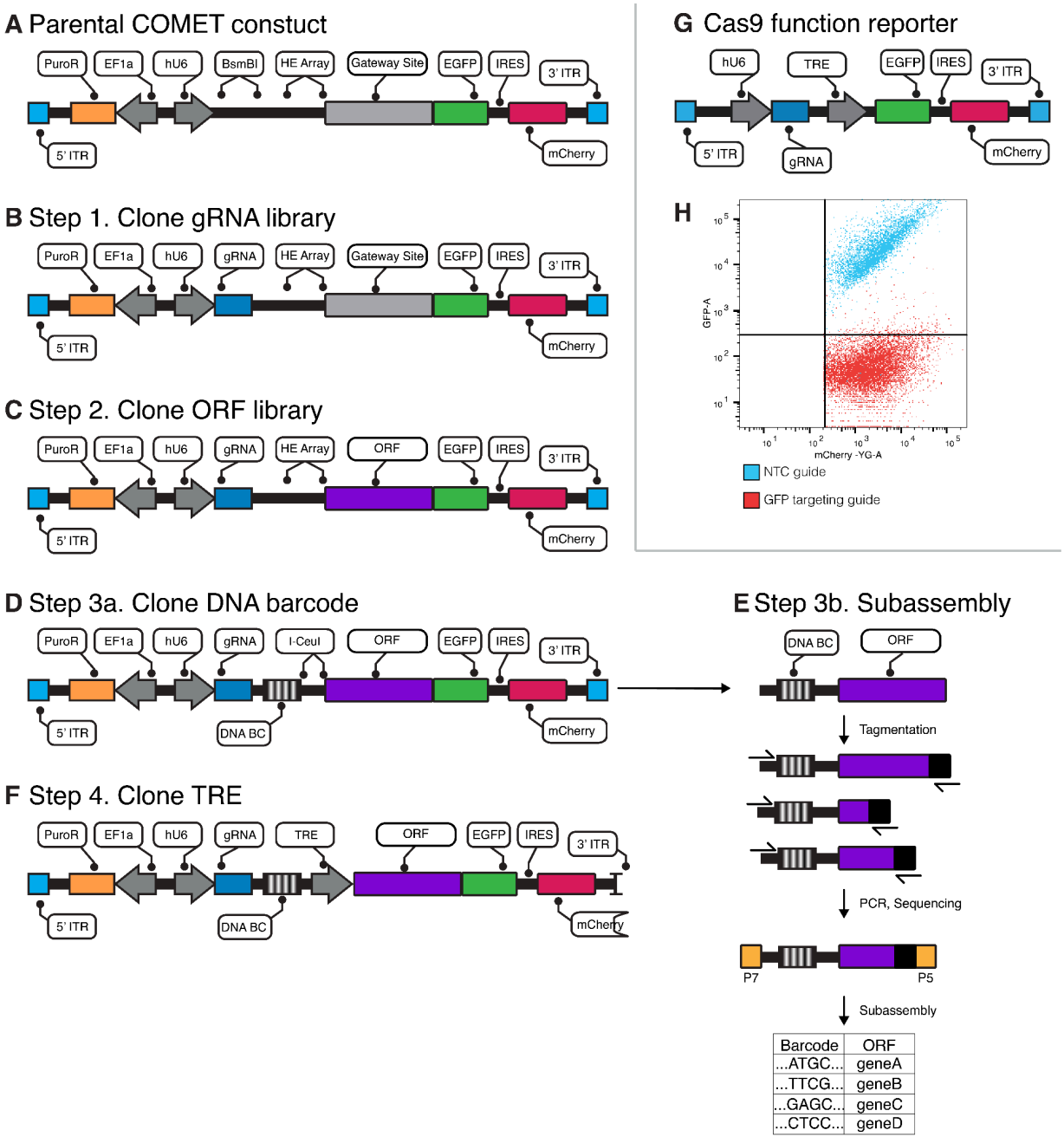
A Parental COMET constuct
PuroR
EF1a
hU6
BsmBI
HE Array
Gateway Site
EGFP
IRES
3' ITR
5' ITR
mCherry
B Step 1. Clone gRNA library
gRNA
C Step 2. Clone ORF library
ORF
D Step 3a. Clone DNA barcode
I-CeuI
DNA BC
E Step 3b. Subassembly
Tagmentation
PCR, Sequencing
P7
P5
Subassembly
Barcode
ORF
...ATGC...
geneA
...TTCG...
geneB
...GAGC...
geneC
...CTCC...
geneD
F Step 4. Clone TRE
TRE
G Cas9 function reporter
gRNA
H
GFP-A
mCherry -YG-A
NTC guide
GFP targeting guide

Figure S1. Schematic of COMET library cloning workflow and Cas9 function reporter.

A-F) The parental COMET construct (**A**) is digested with BsmBI and a gRNA library is cloned into the backbone (**B**). ORF libraries are then recombined into the gRNA containing plasmid library (**C**). The gRNA and ORF containing library is then digested with two homing endonucleases (targeting the HE Array shown in Step 2 schematic), and a 25mer DNA oligo is cloned into the backbone in order to barcode the library (**D**). These barcodes are associated to ORFs via subassembly (**E**). Finally, the barcoded gRNA and ORF library is digested at dual I-CeuI sites and a tetracycline response element (TRE) promoter is cloned into the backbone (**F**). **G)** Schematic of reporter used to test for cell line reporter induction as well as Cas9 activity. Two versions of this reporter were cloned, one with a GFP-specific gRNA and one with an NTC gRNA. **H)** Flow cytometry plot of K562-rtTA-Cas9 clone 3 cells transfected with the GFP-specific gRNA containing reporter (red) as well as the NTC gRNA containing reporter (blue).

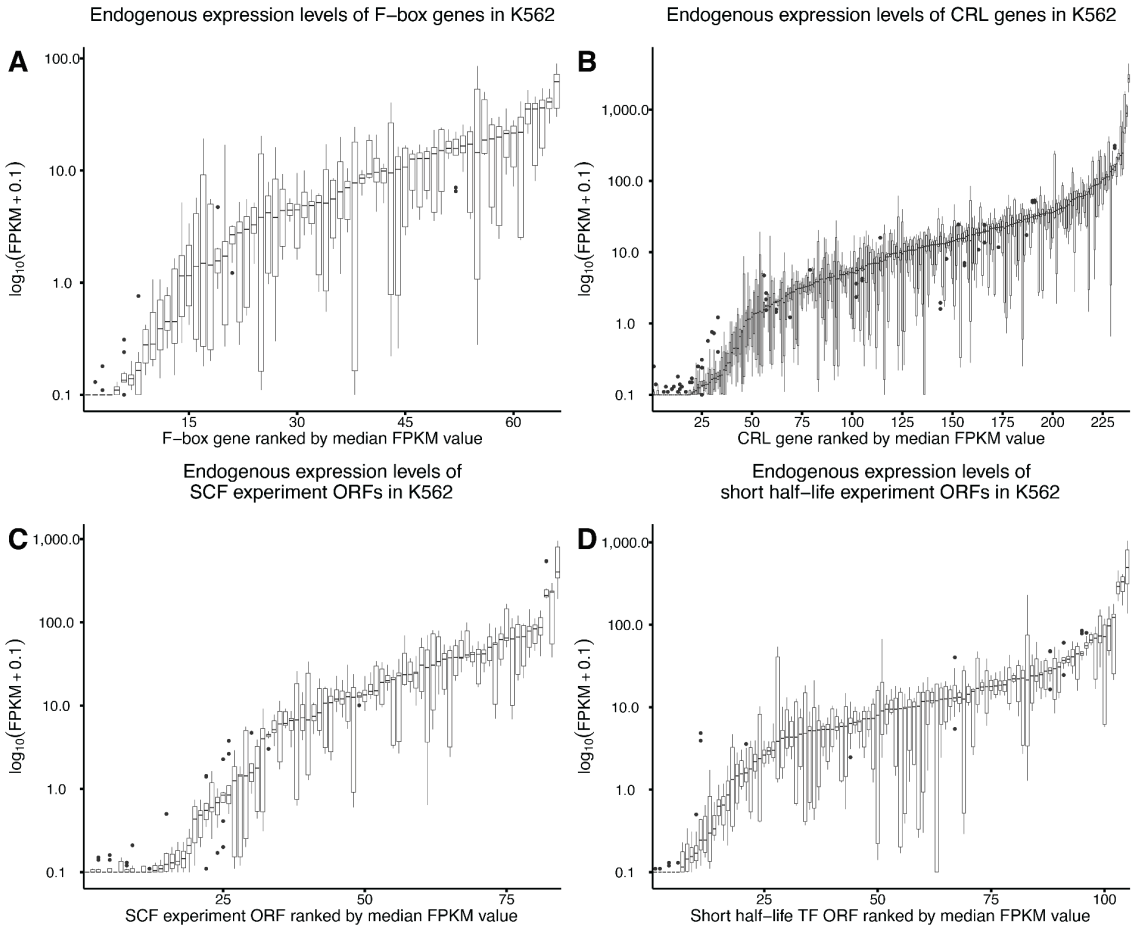

Figure S2. E3 ubiquitin ligase and substrate gene expression levels in K562 cells.

A) Boxplots showing endogenous gene expression levels in K562 cells of 68 F-box genes. Genes are ranked by median fragments per kilobase of transcript per million fragments mapped (FPKM). **B)** Boxplots showing endogenous gene expression levels in K562 cells of 241 CRL genes, ranked by median FPKM. **C)** Boxplots showing endogenous gene expression levels of the 92 putative substrate-endocoding ORFs from the SCF experiment in K562 cells. **D)** Boxplots showing endogenous gene expression levels in K562 cells of the 108 putative substrate-encoding ORFs tested in the short half-life TF experiment.

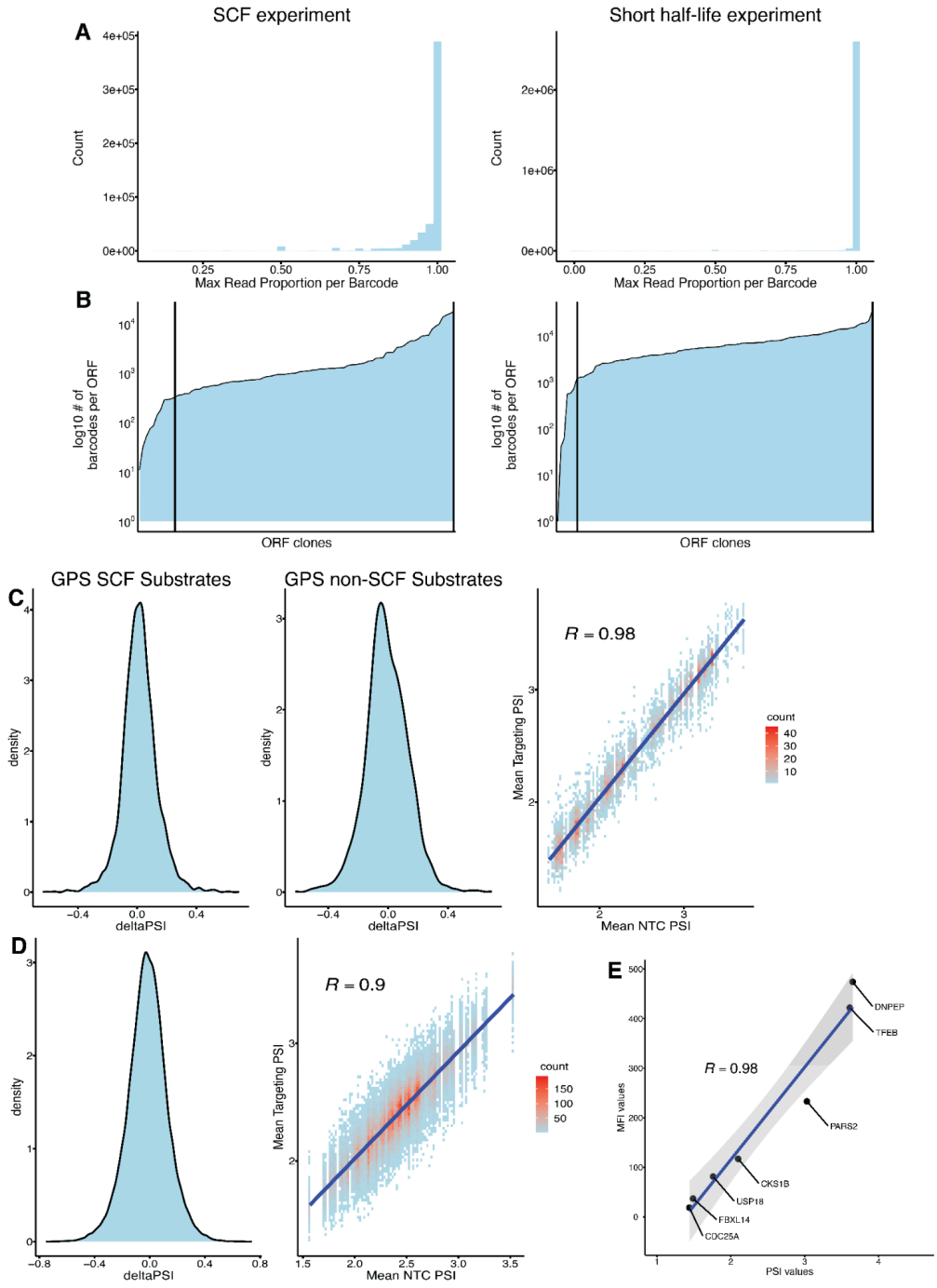
A
SCF experiment
Short half-life experiment
Count
Max Read Proportion per Barcode
B
log10 # of barcodes per ORF
ORF clones
C
GPS SCF Substrates
GPS non-SCF Substrates
density
deltaPSI
R = 0.98
Mean Targeting PSI
Mean NTC PSI
count
D
R = 0.9
E
MFI values
PSI values
DNPEP
TFEB
PARS2
CKS1B
USP18
FBXL14
CDC25A

Figure S3. Visualization of barcode specificity, PSI distributions, and PSI/MFI correlation.

A) Histograms of the max read proportion of individual barcodes for the SCF (left) and short half-life experiments (right). For an individual barcode, the proportion of reads associated to any ORF was computed, and the vast majority (92% SCF, 98%) of barcodes associate to a single ORF >90% of the time. **B)** Line plots of the number of barcodes per ORF for the SCF (left) and short half-life experiments (right). Vertical black lines indicate ORFs with > 300 barcodes (left, SCF experiment) or > 1300 barcodes (right, short half-life experiment). **C)** Density plot of SCF experiment PSIs for GPS-annotated SCF substrate proteins (left) as well as proteins not identified as SCF substrates (middle). Scatter plot colored by point density (right) showing correlation between PSIs calculated for NTC gRNAs (x-axis) or targeting gRNAs (y-axis). **D)** Same as panel **C,** but for the short half-life TF experiment. **E)** Scatterplot comparison of flow cytometry derived Mean Fluorescence Intensities (MFI; y-axis) and computationally calculated Protein Stability Indexes (PSI; x-axis). R denotes Pearson's correlation coefficient.

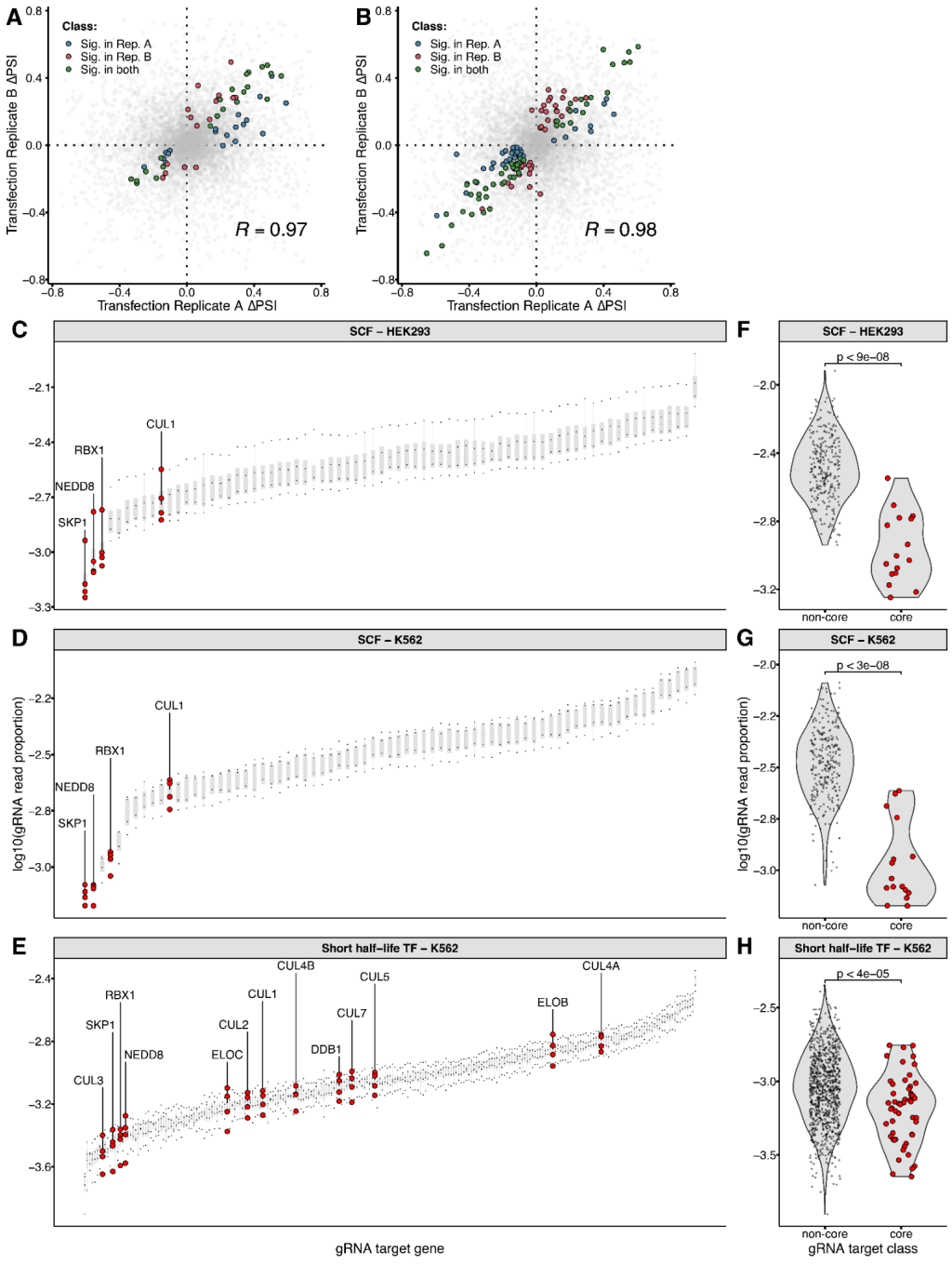
A
Class:
Sig. in Rep. A
Sig. in Rep. B
Sig. in both
Transfection Replicate B ΔPSI
Transfection Replicate A ΔPSI
R = 0.97
B
Class:
Sig. in Rep. A
Sig. in Rep. B
Sig. in both
Transfection Replicate B ΔPSI
Transfection Replicate A ΔPSI
R = 0.98
C
SCF – HEK293
CUL1
RBX1
NEDD8
SKP1
D
SCF – K562
log10(gRNA read proportion)
CUL1
RBX1
NEDD8
SKP1
E
Short half–life TF – K562
CUL4B
CUL5
CUL1
RBX1
CUL7
SKP1
CUL2
NEDD8
ELOC
DDB1
ELOB
CUL4A
CUL3
gRNA target gene
F
SCF – HEK293
p < 9e–08
non-core
core
G
SCF – K562
p < 3e–08
log10(gRNA read proportion)
non-core
core
H
Short half–life TF – K562
p < 4e–05
non-core
core
gRNA target class

Figure S4. Assessment of ∆PSI correlation in two cell lines and core CRL subunit essentiality.

A) Reproducibility of ∆PSI estimates between collapsed replicates in HEK293 cells from the pilot SCF COMET experiment. Reads from four independent transfection replicates were collapsed into two replicates, and ∆PSIs and significance were recalculated. Points are colored based on significance in Replicate A (blue), replicate B (red), or both (green). Pearson correlation was calculated based on the subset of interactions that were significant in either or both cell lines. **B)** Same as panel **A** but for K562 cell line. **C-E)** Plots of gRNA counts isolated from genomic DNA 12 days after library transfection for the HEK293 SCF (**C**), K562 SCF (**D**), and the K562 short half-life TF (**E**) screens. Core SCF components such as cullin scaffolds and substrate adaptors are labeled and read proportions from four independent transfections replicates are indicated with red points. **F-H)** Read proportions for gRNAs targeting core CRL components were significantly different than read proportions for gRNAs targeting non-core CRL subunits. Significance was calculated using a two-sided t-test.

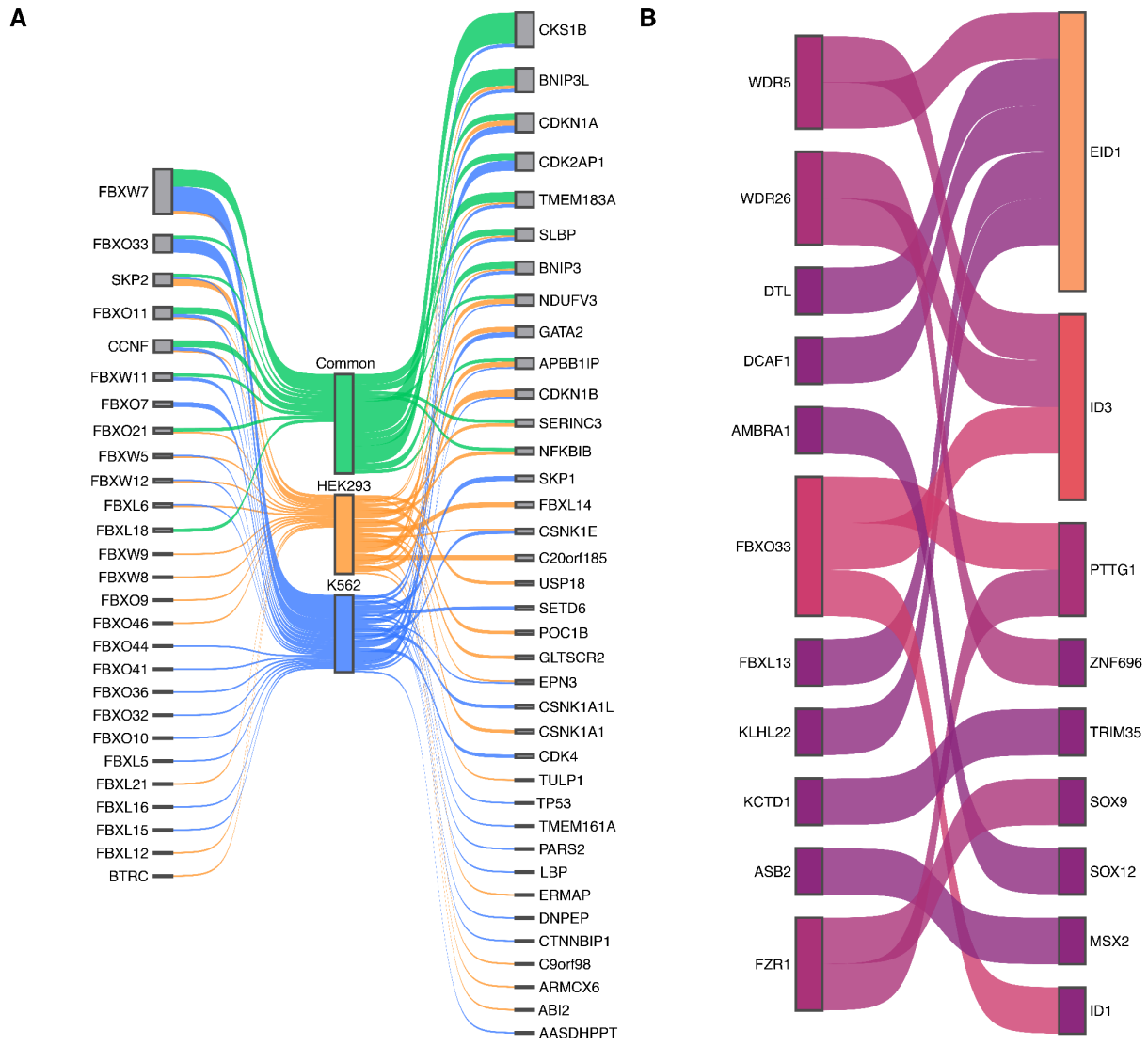

Figure S5. COMET maps E3 ligases to their target substrates.

A) Visual summary of results from the SCF experiment. E3s listed on the left are connected to all COMET-identified significant target substrates ($p < 0.05$, $\Delta PSI > 0$). Flow colors correspond to the cell line in which the association was identified (orange = HEK293, blue = K562, and common = green). **B)** Same as panel **A**, except visually summarizing results from the TF experiment. Flow colors originating from the left correspond to the number of substrates linked to each E3.

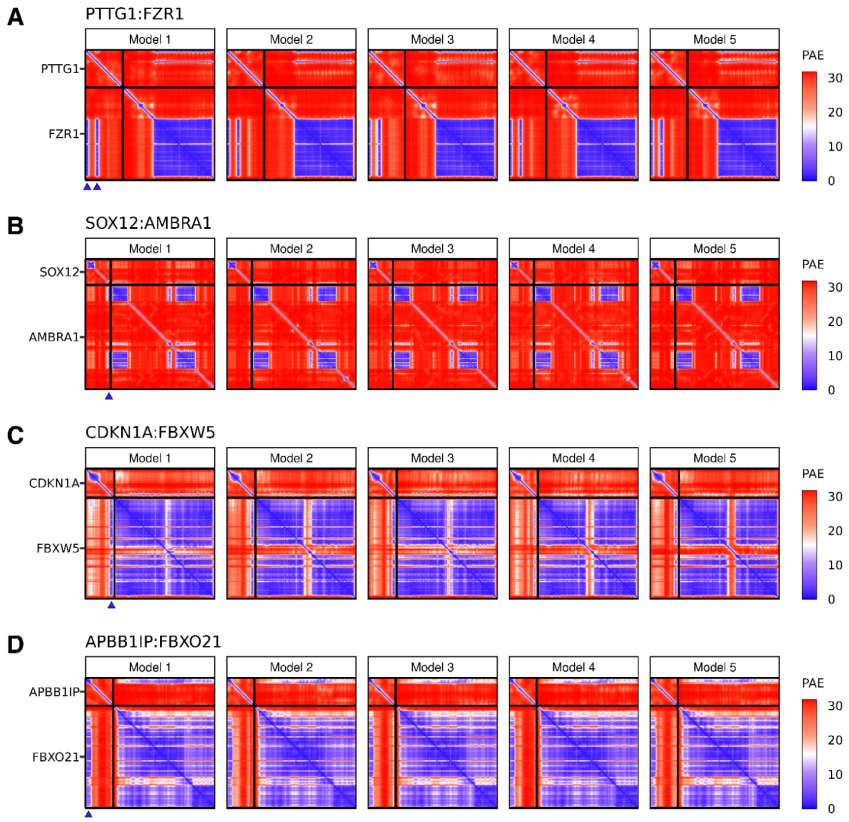
A
PTTG1:FZR1
Model 1
Model 2
Model 3
Model 4
Model 5
PTTG1
FZR1
PAE
30
20
10
0
B
SOX12:AMBRA1
SOX12
AMBRA1
C
CDKN1A:FBXW5
CDKN1A
FBXW5
D
APBB1IP:FBXO21
APBB1IP
FBXO21
E
SOX12 282 TSHFEFPDYCTPEVTEMIAGDWR 304
CCND1 276 EEEEEVDLACTPTDVRDVDI--- 295
CCND2 270 KSEDELDQASTPTDVRDIDL--- 289
CCDN3 273 SSSQGPSQTSTPTDVTAIHL--- 292
G
APBB1IP 1 MGESSEDIDQMFSTLLGEMDLLTQSLGVDTLPPPDPN 37
EID1 160 ----------AFIEELFSLMVVN---------- 172
EID2 209 ----------TFTIALTASEVIN---------- 221
F
S283
I299
C291
V296
T292
P293
SOX12
H
L21
L22
T23
L16
F12

Figure S6. Predicted Aligned Error (PAE) plots for select substrate-E3 models and alignment of *AlphaFold-Multimer* annotated degron motifs to known degrons motifs.

A-D) PAE plots of E3-substrate pairs modeled using AlphaFold-Multimer. Plots are shown for each of 5 individual models generated per E3-substrate pair. Plots are provided for (**A**) PTTG1:FZR1, (**B**) GATA2:FBXL16, (**C**) CDKN1A:FBXW5, and (**D**) APBB1IP:FBXO21, which correspond to **Figure 6A-D** respectively. In each set of plots, quadrants corresponding to either the substrate or the E3 are indicated on the y-axis. Regions in the substrate that are predicted to interact with the E3 are colored blue in the lower left quadrant, and vice versa in the upper right quadrant. For all pairs, blue triangles in the Model 1 panels indicate putative degrons displayed in **Figure 6**. **E)** Alignment of the putative degron region of SOX12 and previously annotated AMBRA1-specific degron motifs in human D-type cyclins. The conserved TP motif through which AMBRA1 recognizes D-type cyclins is highlighted in red. Conserved residues (*), conservative substitutions (:), and semi-conservative substitutions (.) are indicated below the alignment. **F)** Visualization of the predicted AMBRA1 (purple) and SOX12 (tan) interaction interface with side chains of residues with any level of conservation from the alignment in panel **E** shown. **G)** Alignment of the putative degron region of APBB1IP against previously annotated FBXO21-specific degrons in EID1 and its paralog EID2. **H)** Visualization of the predicted FBXO21 (blue) and APBB1IP (red) interaction interface with side chains displayed of residues with any level of conservation from the alignment in panel **G**.

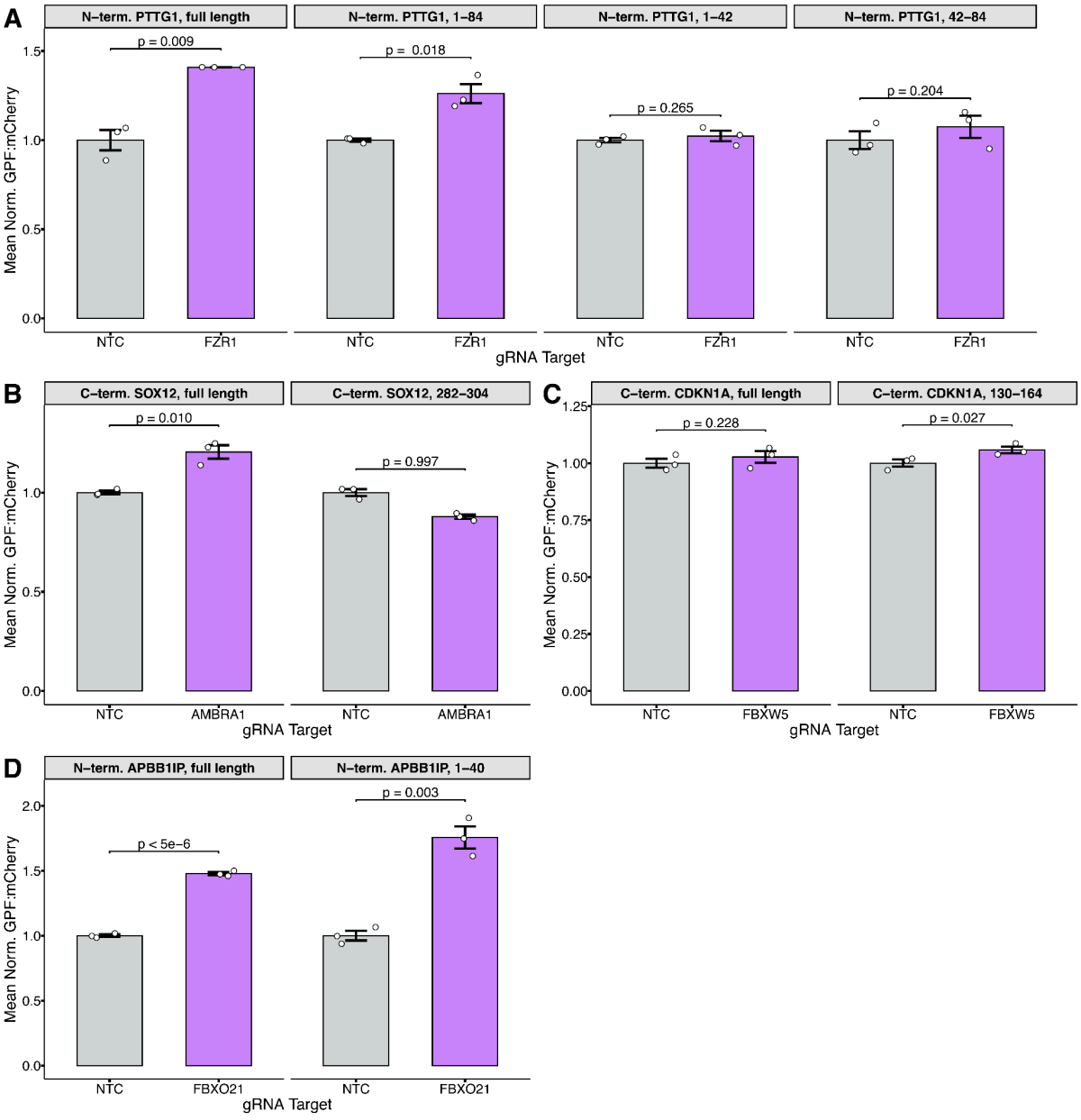
A
N–term. PTTG1, full length
N–term. PTTG1, 1–84
N–term. PTTG1, 1–42
N–term. PTTG1, 42–84
p = 0.009
p = 0.018
p = 0.265
p = 0.204
Mean Norm. GPF:mCherry
NTC
FZR1
gRNA Target
B
C–term. SOX12, full length
C–term. SOX12, 282–304
p = 0.010
p = 0.997
AMBRA1
C
C–term. CDKN1A, full length
C–term. CDKN1A, 130–164
p = 0.228
p = 0.027
FBXW5
D
N–term. APBB1IP, full length
N–term. APBB1IP, 1–40
p < 5e–6
p = 0.003
FBXO21

Figure S7. Assessment of E3-specific proteolysis of AlphaFold-predicted degrons.

NTC-normalized GFP:mCherry ratios as measured by flow cytometry are presented for *AlphaFold-Multimer* nominated degrons presented in **Figure 6**. Panels **A-D** correspond to **Figure 6A-D** respectively. Cells were transfected with the indicated combinations of GFP:mCherry reporter and gRNA expression plasmid. **A)** GFP:mCherry ratios measured in cells expressing various PTTG1 reporters and either an NTC or FZR1 gRNA plasmid. **B)** GFP:mCherry ratios measured in cells expressing various SOX12 reporters and either an NTC or AMBRA gRNA plasmid. **C)** GFP:mCherry ratios measured in cells expressing various CDKN1A reporters and either an NTC or FBXW5 gRNA plasmid. **D**) GFP:mCherry ratios measured in cells expressing various APBB1IP reporters and either an NTC or FBXO21 gRNA plasmid. Significance was calculated using a one-sided t-test.

2.5 References for Chapter 2

1. King, R. W., Deshaies, R. J., Peters, J. M. & Kirschner, M. W. How proteolysis drives the cell cycle. Science 274, 1652–1659 (1996).
2. Tan, X. et al. Mechanism of auxin perception by the TIR1 ubiquitin ligase. Nature 446, 640–645 (2007).
3. Berndsen, C. E. & Wolberger, C. New insights into ubiquitin E3 ligase mechanism. Nat. Struct. Mol. Biol. 21, 301–307 (2014).
4. Deshaies, R. J. & Joazeiro, C. A. P. RING domain E3 ubiquitin ligases. Annu. Rev. Biochem. 78, 399–434 (2009).
5. Iconomou, M. & Saunders, D. N. Systematic approaches to identify E3 ligase substrates. Biochem. J 473, 4083–4101 (2016).
6. Pierce, N. W., Kleiger, G., Shan, S.-O. & Deshaies, R. J. Detection of sequential polyubiquitylation on a millisecond timescale. Nature 462, 615–619 (2009).
7. Yen, H.-C. S. & Elledge, S. J. Identification of SCF ubiquitin ligase substrates by global protein stability profiling. Science 322, 923–929 (2008).
8. Petroski, M. D. & Deshaies, R. J. Function and regulation of cullin–RING ubiquitin ligases. Nat. Rev. Mol. Cell Biol. 6, 9–20 (2005).
9. Yen, H.-C. S., Xu, Q., Chou, D. M., Zhao, Z. & Elledge, S. J. Global protein stability profiling in mammalian cells. Science 322, 918–923 (2008).
10. Jin, J. et al. Systematic analysis and nomenclature of mammalian F-box proteins. Genes Dev. 18, 2573–2580 (2004).
11. Yusa, K., Zhou, L., Li, M. A., Bradley, A. & Craig, N. L. A hyperactive piggyBac transposase for mammalian applications. Proc. Natl. Acad. Sci. U. S. A. 108, 1531–1536 (2011).
12. Emanuele, M. J. et al. Global identification of modular cullin-RING ligase substrates. Cell 147, 459–474 (2011).
13. Dankert, J. F. et al. Cyclin F-Mediated Degradation of SLBP Limits H2A.X Accumulation and Apoptosis upon Genotoxic Stress in G2. Mol. Cell 64, 507–519 (2016).
14. Cui, D. et al. FBXW7 Confers Radiation Survival by Targeting p53 for Degradation. Cell Rep. 30, 497–509.e4 (2020).
15. Kimura, T., Gotoh, M., Nakamura, Y. & Arakawa, H. hCDC4b, a regulator of cyclin E, as a direct transcriptional target of p53. Cancer Sci. 94, 431–436 (2003).
16. Williams, A. B. & Schumacher, B. p53 in the DNA-Damage-Repair Process. Cold Spring Harb. Perspect. Med. 6, (2016).
17. Vicente, C., Conchillo, A., García-Sánchez, M. A. & Odero, M. D. The role of the GATA2 transcription factor in normal and malignant hematopoiesis. Crit. Rev. Oncol. Hematol. 82, 1–17 (2012).
18. Nakajima, T. et al. Regulation of GATA-binding protein 2 levels via ubiquitin-dependent degradation by Fbw7: involvement of cyclin B-cyclin-dependent kinase 1-mediated phosphorylation of THR176 in GATA-binding protein 2. J. Biol. Chem. 290, 10368–10381 (2015).
19. Lurie, L. J., Boyer, M. E., Grass, J. A. & Bresnick, E. H. Differential GATA Factor Stabilities: Implications for Chromatin Occupancy by Structurally Similar Transcription

Factors. Biochemistry 47, 859–869 (2008).
20. van der Harst, P. et al. Seventy-five genetic loci influencing the human red blood cell. Nature 492, 369–375 (2012).
21. Randle, S. J., Nelson, D. E., Patel, S. P. & Laman, H. Defective erythropoiesis in a mouse model of reduced Fbxo7 expression due to decreased p27 expression. J. Pathol. 237, 263–272 (2015).
22. Mark, K. G. & Rape, M. Ubiquitin-dependent regulation of transcription in development and disease. EMBO Rep. 22, e51078 (2021).
23. Lipford, J. R. & Deshaies, R. J. Diverse roles for ubiquitin-dependent proteolysis in transcriptional activation. Nat. Cell Biol. 5, 845–850 (2003).
24. Geng, F., Wenzel, S. & Tansey, W. P. Ubiquitin and proteasomes in transcription. Annu. Rev. Biochem. 81, 177–201 (2012).
25. Kronke, J. et al. Lenalidomide Causes Selective Degradation of IKZF1 and IKZF3 in Multiple Myeloma Cells. Science vol. 343 301–305 Preprint at https://doi.org/10.1126/science.1244851 (2014).
26. Lu, G. et al. The myeloma drug lenalidomide promotes the cereblon-dependent destruction of Ikaros proteins. Science 343, 305–309 (2014).
27. Vassilev, L. T. et al. In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. Science 303, 844–848 (2004).
28. Li, J. et al. Proteome-wide mapping of short-lived proteins in human cells. Mol. Cell 81, 4722–4735.e5 (2021).
29. Peters, J.-M. The anaphase promoting complex/cyclosome: a machine designed to destroy. Nat. Rev. Mol. Cell Biol. 7, 644–656 (2006).
30. Clute, P. & Pines, J. Temporal and spatial control of cyclin B1 destruction in metaphase. Nat. Cell Biol. 1, 82–87 (1999).
31. Hagting, A. et al. Human securin proteolysis is controlled by the spindle checkpoint and reveals when the APC/C switches from activation by Cdc20 to Cdh1. J. Cell Biol. 157, 1125–1137 (2002).
32. Glotzer, M., Murray, A. W. & Kirschner, M. W. Cyclin is degraded by the ubiquitin pathway. Nature 349, 132–138 (1991).
33. Pfleger, C. M. & Kirschner, M. W. The KEN box: an APC recognition signal distinct from the D box targeted by Cdh1. Genes Dev. 14, 655–665 (2000).
34. Guharoy, M., Bhowmick, P., Sallam, M. & Tompa, P. Tripartite degrons confer diversity and specificity on regulated protein degradation in the ubiquitin-proteasome system. Nat. Commun. 7, 1–13 (2016).
35. Südbeck, P. & Scherer, G. Two Independent Nuclear Localization Signals Are Present in the DNA-binding High-mobility Group Domains of SRY and SOX9*. J. Biol. Chem. 272, 27848–27852 (1997).
36. Gasca, S. et al. A nuclear export signal within the high mobility group domain regulates the nucleocytoplasmic translocation of SOX9 during sexual determination. Proc. Natl. Acad. Sci. U. S. A. 99, 11199–11204 (2002).
37. Cenik, B. K. & Shilatifard, A. COMPASS and SWI/SNF complexes in development and disease. Nat. Rev. Genet. 22, 38–58 (2020).
38. Evans, R. et al. Protein complex prediction with AlphaFold-Multimer. bioRxiv

2021.10.04.463034 (2022) doi:10.1101/2021.10.04.463034.
39. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589 (2021).
40. Huttlin, E. L. et al. Dual proteome-scale networks reveal cell-specific remodeling of the human interactome. Cell 184, 3022–3040.e28 (2021).
41. Lim, Y. et al. In silico protein interaction screening uncovers DONSON's role in replication initiation. Science 381, eadi3448 (2023).
42. Simoneschi, D. et al. CRL4AMBRA1 is a master regulator of D-type cyclins. Nature 592, 789–793 (2021).
43. Puklowski, A. et al. The SCF–FBXW5 E3-ubiquitin ligase is regulated by PLK4 and targets HsSAS-6 to control centrosome duplication. Nat. Cell Biol. 13, 1004–1009 (2011).
44. Werner, A. et al. SCFFbxw5 mediates transient degradation of actin remodeller Eps8 to allow proper mitotic progression. Nat. Cell Biol. 15, 179–188 (2013).
45. Zhang, C. et al. Peptidic degron in EID1 is recognized by an SCF E3 ligase complex containing the orphan F-box protein FBXO21. Proceedings of the National Academy of Sciences 112, 15372–15377 (2015).
46. Morreale, F. E. & Walden, H. Types of Ubiquitin Ligases. Cell 165, 248–248.e1 (2016).
47. Tokheim, C. et al. Systematic characterization of mutations altering protein degradation in human cancers. Mol. Cell 81, 1292–1308.e11 (2021).
48. Timms, R. T. et al. Defining E3 ligase–substrate relationships through multiplex CRISPR screening. Nat. Cell Biol. 25, 1535–1545 (2023).
49. Lazar, N. H. et al. High-resolution genome-wide mapping of chromosome-arm-scale truncations induced by CRISPR-Cas9 editing. bioRxiv 2023.04.15.537038 (2023) doi:10.1101/2023.04.15.537038.
50. Doench, J. G. et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat. Biotechnol. 34, 184–191 (2016).
51. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74 (2012).
52. Mirdita, M. et al. ColabFold: making protein folding accessible to all. Nat. Methods 19, 679–682 (2022).
53. Słabicki, M. et al. The CDK inhibitor CR8 acts as a molecular glue degrader that depletes cyclin K. Nature 585, 293–297 (2020).

Chapter 3
MULTIPLEX DESIGN AND DISCOVERY OF PROXIMITY HANDLES FOR PROGRAMMABLE PROTEOME EDITING

This Chapter is adapted with minimal changes from a manuscript under review

Suiter, C. C.†#, Ahn, G.†, Chiu, M., Fu, Y., Sadre, S., Simon, J. J., Lee, D. S., Fowler, D. M., Maly, D. J., Baker, D.#, & Shendure, J.# (2025). Multiplex design and discovery of proximity handles for programmable proteome editing. bioRxiv(p.2025.10.13.681693). https://doi.org/10.1101/2025.10.13.681693

† denotes equal contribution; # denotes corresponding authorship

Contributions: C.C.S. and G.A. conceptualized the application of de novo designed proteins for targeted protein degradation and stabilization with input from D.B. and J.S.. G.A. designed binders against FBXL12, GABARAP, GABARAPL2, MAP1LC3A, and UCHL1. S.S. designed binders against USP38 and TRAF2. C.C.S. designed and executed screening experiments with assistance from M.C. and input from J.J.S. and J.S.. C.C.S. analyzed screen data with input from J.S.. C.C.S. performed fluorescent reporter validation experiments with assistance from M.C. and D.S.L.. G.A. performed western blotting and apoptosis experiments with assistance from C.C.S.. Y.F. performed confocal microscopy of mitochondria with assistance from C.C.S.. C.C.S., G.A., D.B, and J.S. wrote the initial draft of the manuscript with input from all other authors. D.M.F., D.J.M., D.B., and J.S. supervised the project.

3.1. ABSTRACT

Although we now have a rich toolset for genome editing, an equivalent framework for manipulating the proteome with a comparable flexibility and specificity remains elusive. A promising strategy for “proteome editing” is to use bifunctional molecules (*e.g.* PROteolysis-Targeting Chimeras or PROTACs[51]) that bring a target protein into proximity with a degradation or stabilization effector, but their broader application is constrained by a limited repertoire of well-characterized target or effector “handles”. We asked whether coupling *de novo* protein design to a multiplex screening framework could address this gap by accelerating the discovery of effector handles for intracellular protein degradation, stabilization, or relocalization. Using LABEL-seq[52], a sequencing-based assay that enables multiplex, quantitative measurement of protein abundance, we screened 9,715 *de novo* designed candidate effector handles for their ability to recruit a target protein to components of the ubiquitin–proteasome system[53] (UPS) (FBXL12, TRAF2, UCHL1, USP38) or the autophagy pathway[54] (GABARAP, GABARAPL2, MAP1LC3A). In a single experiment, we discovered hundreds of *de novo* designed effector handles that reproducibly drove either intracellular degradation (n = 277) or stabilization (n = 204) of a reporter protein. Validation of a subset of these hits in an orthogonal assay confirmed that sequencing-based measurements from the primary screen reliably reflected changes in intracellular abundance of the target protein. Successful effector handles were discovered for both the UPS (n = 194) and autophagy (n = 287) pathways, which provide complementary routes for programmable proteome editing. Autophagy-recruiting effector handles generalized to endogenous targets, as substituting the reporter-specific target handle with a high-affinity MCL1 binder[55] reduced endogenous levels of this intracellular oncoprotein[56]. Moreover, directing autophagy-recruiting effector handles to the outer mitochondrial membrane dramatically perturbed mitochondrial networks in a manner consistent with synthetic tethering and sequestration[57,58]. Beyond generating a diverse repertoire of protein abundance or localization effector handles, our results establish a scalable, low-cost platform that links deep learning–guided protein design to functional cellular readouts, and chart a course toward a general framework for programmable proteome editing.

3.2 INTRODUCTION

The past decade has seen transformative advances in genome editing, with CRISPR-based technologies[59] now enabling the precise, programmable manipulation of virtually any genomic locus (DNA) and, increasingly, transcribed nucleic acids (RNA). However, our ability to manipulate the remaining layer of the central dogma—proteins—with comparable flexibility remains primitive. Unlike nucleic acids, amino acid polymers lack predictable base-pairing rules, and thus resist the kinds of programmable targeting that have made genome and transcriptome editing routine. Advancing from genome editing to proteome editing will require not only a “phone book”

for targeting proteins but also generalizable means of modulating their abundance, localization, and activity with precision within living cells and tissues.

One promising strategy for proteome editing is proximity induction[60], in which a target protein is brought into physical proximity with a cellular effector to trigger a desired outcome—analogous to how guide RNAs and Cas9-effector fusions recruit chromatin-modifying activities to specific genomic loci. Targeted protein degradation with PROTACs[51] exemplifies this concept[61]. By recruiting a protein-of-interest to endogenous degradation machinery, these bifunctional small molecules can selectively eliminate intracellular proteins. However, despite their therapeutic promise, most PROTACs rely on a narrow set of effector “handles” (e.g., CRBN, VHL), leaving the vast majority of the human E3 ligase repertoire—as well as other potential effector pathways—untapped. PROTACs’ broader application is further limited by the requirement for well-defined binding pockets on both the target and the effector, which are not always available, and by the challenge of scaling high-affinity ligand discovery across the proteome. As a result, the broader potential of proximity induction as a strategy for proteome editing is constrained not by concept, but by execution[62–64].

Deep-learning–based *de novo* protein design[65,66] offers a potentially powerful route to overcoming these constraints. Computational design tools can now rapidly generate candidate binders to virtually any surface epitope, bypassing the structural constraints that restrict small molecules and, in principle, enabling programmable recruitment of almost any target protein to a broad spectrum of effector functions defined by the nature and location of that recruitment. This potential is underscored by numerous examples of bifunctional proteins that use natural binding domains, antibody fragments, or combinations thereof for proximity induction[67–70], yet these too remain limited by the range of available reagents. In principle, *de novo* protein design opens the door to the generalization of proximity induction-based proteome editing. However, while modern computational pipelines can now readily generate thousands of candidate binders against virtually any protein, most efforts to date have focused on screening for *in vitro* affinity rather than intracellular function.

A central challenge is how to scale functional screening to match the scale of *de novo* design. For example, if the goal is to build bifunctional proteome editors that modulate protein abundance, this requires quantitatively assessing which designs successfully recruit and degrade a target protein inside cells. Historically, methods for the multiplexed quantification of protein abundance have been limited. However, LABEL-seq[52] offers a powerful solution. By linking the abundance of a target protein to an RNA barcode, LABEL-seq enables protein levels to be measured across thousands of cellular contexts within a single experiment.

Here we describe how integrating three key elements—*de novo* protein design, intracellular RNA barcoding of proteins, and the cell as a compartment in which designs

and barcoded targets are brought together—creates a framework for scaling the discovery of reagents for intracellular proteome editing. We show that this combination enables sequencing-based quantification of how a target protein's abundance is modulated by each of ~10,000 candidate effector handles in one experiment, allowing not only detection of successful designs but also their ranking by intracellular potency, characterization of their effect-size distributions, and comparisons across target–effector pairings or cellular contexts. Together, these capabilities bridge protein design and cell-based functional screens, opening a path to the generation of a "phone book" for programmable proteome editing in human cells.

3.3 RESULTS

3.3.1 A strategy for the multiplex, intracellular assessment of de novo designed proteome editors

We set out to develop a multiplex workflow that quantitatively measures how libraries of bifunctional proteins—each comprising a target-binding domain ("target handle") and an effector-binding domain ("effector handle")—alter the intracellular abundance of a target protein. Classical proteomic methods (*e.g.* Western blots, mass spectrometry) are not readily scaled or multiplexed. Coupling fluorescence-activated cell sorting (FACS) with nucleic acid barcodes ("sort-seq")[71,72] provides an attractive alternative, but typically relies on a limited number of bins, reducing quantitative resolution, and is also constrained to a single fluorescently labeled target or effector per experiment.

To overcome these limitations, we turned to LABEL-seq[52], which converts intracellular protein abundance into a sequencing-based readout. In this approach, a protein-of-interest (POI) is fused to an RNA-binding domain (RBD) and co-expressed with a uniquely barcoded RNA species that binds the RBD via an MS2 hairpin-MS2 coat protein (MCP) interaction[73]. Following immunoprecipitation (IP) of the RBD–POI fusion, the co-enriched RNA barcodes are sequenced, yielding a digital, quantitative measurement of protein abundance. In a *cis* configuration, sequence variants of the POI are tagged with unique barcodes, *e.g.* enabling multiplex measurement of their effects on protein stability[52] (**Fig. S1A**). However, in a *trans* configuration, an invariant, RBD-tagged target can be labeled with an RNA barcode that instead encodes the identity of a concurrently delivered perturbation, *e.g.* a bifunctional proteome editor designed to modulate the RBD-tagged target's abundance (**Fig. S1B**).

A critical feature of the *trans* configuration setup is that the cell itself serves as a compartment that links a specific perturbation (here, a bifunctional proteome editor) to a specific RNA barcode. As the RNA barcode remains bound to the target protein even after cells are lysed via the strong MS2-MCP interaction[52], the intracellular effects of many candidate editors can potentially be measured in a single multiplex experiment in which the primary readout is massively parallel sequencing. Although this strategy is compatible with the screening of either target handles, effector handles, or combinations

thereof, our proof-of-concept focused on evaluating computationally designed effector handles. In particular, we sought to evaluate thousands of designed binders to E3 ligases or autophagy-related proteins for their ability to mediate the intracellular degradation or stabilization of a target protein.

A first key component of our experimental setup for this proof-of-concept is a plasmid library encoding thousands of barcoded bifunctional proteome editors (**Fig. 1A**). In brief, a constant target handle (NbALFA, a high-affinity antibody to the ALFA tag[74]) is fused to a library of designed effector handles (**Fig. S1C**). Each effector handle design is associated with a DNA barcode on the same construct that will later be expressed as a circular RNA barcode[52]. By using degenerate DNA barcodes and subassembly procedures developed for massively parallel reporter assays[75,76], we associate each effector handle design with multiple barcodes (**Fig. S2A**). Importantly, the resulting plasmid library is promoterless, such that its components cannot be expressed until after successful integration to a genomic landing pad via Bxb1-mediated recombination (**Fig. S1D**), as described further below.

A second key component is a dual reporter construct encoding the target protein and a “standard” protein (**Fig. 1B**). The target protein is fused to a FLAG epitope (for IP), an MCP RBD (to recruit the LABEL-seq barcode) and an ALFA epitope (to recruit the bifunctional protein editor). The standard protein is tagged with a Myc epitope and the MCP RBD, but lacks any target recruitment epitope. In the LABEL-seq framework[52], the Myc-tagged standard serves as an internal control. In any given cell, both the target and standard protein are bound by the same RNA barcode via the MS2-MCP interaction. Thus, target-associated barcode counts (FLAG pull-down) can be normalized to standard-associated barcode counts (Myc pull-down), isolating design-dependent changes in target abundance from technical or cell-to-cell variation.

We used Bxb1-mediated recombination to introduce the library of barcoded proteome editors into engineered HEK293 cells bearing a single copy of a landing pad[52]. Prior to recombination, the landing pad consists of an attP site sandwiched between convergent U6 and tetracycline response element (TRE) promoters (**Fig. S1D**). Successful recombination positions the proteome editor component downstream of the doxycycline-inducible TRE promoter, and the RNA barcode component downstream of the constitutive U6 promoter (**Fig. S1E**). Following selection for successful landing pad integrants, expression of the proteome editor is induced with doxycycline, and the dual reporter concurrently introduced via transient transfection.

At this stage, each successfully transfected cell should contain: (**i**) a single bifunctional proteome editor with a fixed target handle (NbALFA) and a variable (designed) effector handle; (**ii**) a single circular RNA species, bearing an MS2 hairpin and a barcode associated with the *cis*-encoded design; (**iii**) Target protein, bearing FLAG epitope, MCP domain, and ALFA epitope; and (**iv**) Standard protein, bearing Myc

epitope and MCP domain. Within each cell, the MS2-bearing circRNA binds the MCP domain of both target and standard proteins, effectively barcoding individual protein molecules[52]. The levels of circRNA-barcoded target proteins may be modulated by successful effector handles, *e.g.* by recruiting them to proteasomes or lysosomes (**Fig. 1C**), while those of standard proteins should not. Forty-eight hours after editor induction and reporter transfection, the cells are lysed and the lysate split to either an anti-FLAG or anti-Myc IP (**Fig. 1D**). As per the LABEL-seq protocol, circRNAs recovered from each IP are then subjected to RT-PCR and massively parallel sequencing.

From the resulting sequencing data, editor-associated barcodes are counted, and their proportional contributions to the anti-FLAG IP-derived library and the anti-Myc IP-derived library are calculated and ranked. In this *trans* configuration setup of LABEL-seq, the bifunctional proteome editors whose associated barcodes are relatively enriched in the anti-FLAG IP-derived library can be inferred to have promoted stabilization of the target protein, while those whose associated barcodes are relatively depleted can be inferred to have promoted its degradation (**Fig. 1D**).

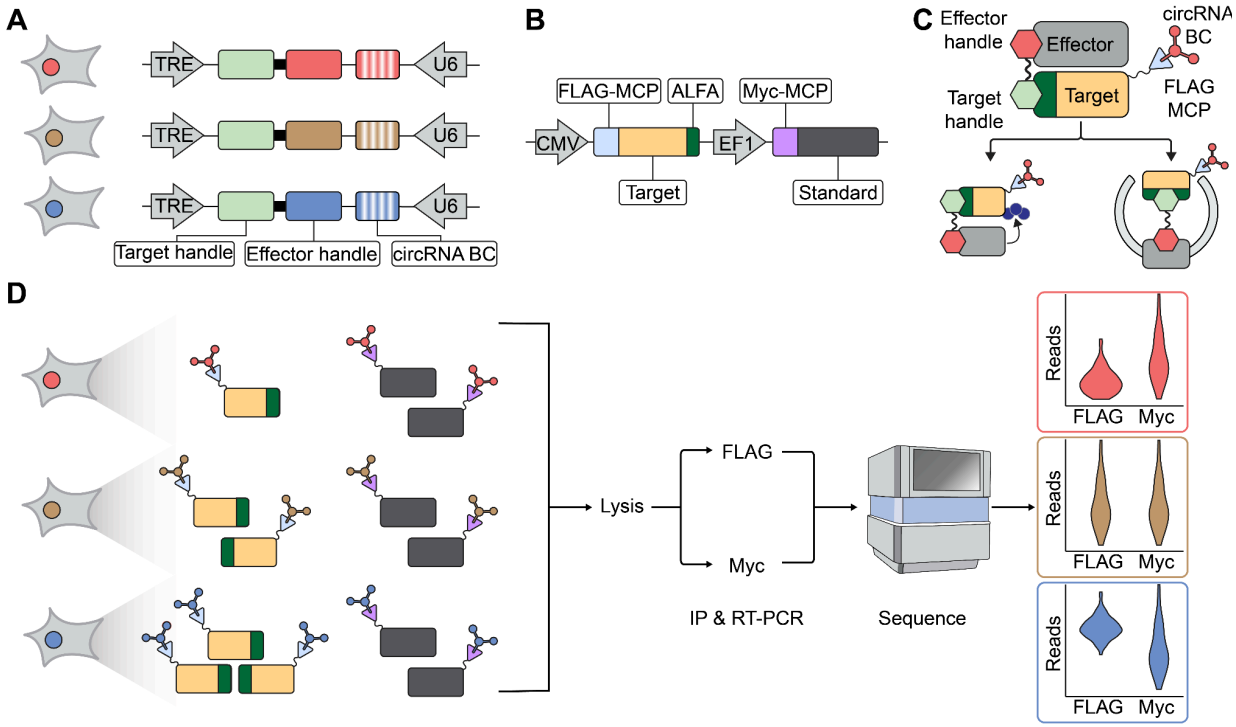

Figure 1. Multiplexed screening platform for de novo designed proteome editors using LABEL-seq.

A) Schematic of genomic landing pad in engineered HEK293 cells following successful recombination. The locus encodes a doxycycline-inducible bifunctional proteome editor consisting of a constant target handle (green, NbALFA) fused to a variable effector handle generated through computational design (red, brown or blue; solid). For LABEL-seq[52], the locus also encodes a constitutively expressed RNA species containing a barcode subsequence, MS2 hairpin, and terminal ribozymes (red, brown or blue; vertical stripes). Prior to recombination into the landing pad, each barcode subsequence is linked to a specific effector handle design via subassembly. Terminal ribozymes mediate circularization, generating a circular RNA barcode capable of binding the MCP domain via the MS2 hairpin. **B)** Schematic of the transiently transfected dual reporter construct. The construct consists of a CMV-driven target reporter (tan), which bears a FLAG-tagged MCP domain (pale blue) and ALFA epitope (dark green), and a EF1-driven standard reporter (dark grey), which bears a Myc-tagged MCP domain (purple). **C)** Schematic of a bifunctional proteome editor that successfully induces proximity of target and effector proteins. Also shown on the target protein is a FLAG-MCP domain (pale blue triangle) bound to a circRNA barcode that identifies the effector handle. Proximity induction can result in target degradation via ubiquitination (left, navy blue circles) or autophagy (right). **D)** Schematic of experimental workflow. Recombined HEK293 cells, each expressing one bifunctional proteome editor and one circular RNA barcode, are transiently transfected with the dual reporter construct. Design-identifying barcodes label both the target and standard reporters. Interaction of the editor with the target protein may result in its degradation (red), no effect (brown) or stabilization (blue). However, editors are not expected to interact with the standard protein. After 48 hours, cells are lysed, and the resulting lysate split to anti-FLAG and anti-Myc IPs. Co-immunoprecipitated circular RNA barcodes are subjected to RT-PCR and massively parallel sequencing. Effector handle designs that successfully promote degradation (red) or stabilization (blue) of the target protein are expected to shift the distribution of barcode counts in the anti-FLAG IP (target-specific) relative to the baseline defined by the anti-Myc IP (standard-specific), while those that have no effect are not (brown).

3.3.2 De novo protein design of candidate effector handles

While developing this experimental framework, we also sought to select effector proteins for a proof-of-concept, and to computationally design binders to them as our candidate effector handles for bifunctional proteome editors. A recent proteome-wide study systematically profiled endogenous human proteins for their capacity to mediate proximity-dependent protein degradation or stabilization[69]. Guided by these results, we selected seven representative effectors associated with the major protein degradation pathways (**Fig. 2A**). Specifically, we chose FBXL12 and TRAF2 as components of distinct E3 ubiquitin ligase complexes, UCHL1 and USP38 as deubiquitinases (DUBs), and GABARAP, GABARAPL2, and MAP1LC3A as autophagy-related proteins (ATGs). In the cited study[69], FBXL12 and the three ATGs were classified as degraders, while TRAF2 and the two DUBs were classified as stabilizers.

To design candidate binders for these seven effectors, we employed deep-learning-based *de novo* protein design. First, RF Diffusion[65] was used to build protein backbones predicted to engage each effector. We then used Protein Message Passing Neural Network (Protein-MPNN)[66] to propose amino acid sequences for each backbone. Finally, the resulting candidates were structurally evaluated with AlphaFold2 (AF2)[40]. For each effector, we performed two iterative rounds of partial diffusion and sequence resampling to optimize binder–effector interfaces, and moved forward only with designs that passed AF2 confidence thresholds for interface predicted aligned error (PAE) and predicted local distance difference test (pLDDT) scores.

Altogether, we designed 10,225 candidate effector handles across seven effector targets, ranging from 125 for FBXL12 to 4632 for GABARAP (**Fig. S2B**; **Tables S1**, **S6**). Each design was 100 amino acids in length. These sequences were synthesized as a pool of oligonucleotides (Twist Biosciences), PCR-amplified, and cloned into a plasmid backbone. A 16-nucleotide (nt) degenerate barcode was then introduced downstream of each construct, generating the barcoded library that was ultimately recombined into the genomic landing pad (**Fig. 1A**). A high-confidence dictionary of mappings of barcodes to designs was generated by subassembly[75] (**Figs. S2A**,**C-D**). Because the library was generated and maintained at high complexity, each effector handle was represented by multiple unique barcodes (mean 43 ± 41; **Fig. S2E**; **Table S2**). This is a key feature, as each barcode mitigates against technical variation by providing a quasi-independent internal replicate for a given design within the context of a single multiple experiment.

3.3.3 ~10,000-plex functional assessment of bifunctional proteome editors

This barcoded library—encoding 10,225 bifunctional proteome editors, each composed of a constant target handle and a variable, computationally designed effector handle—was introduced into HEK293 landing pad cells via Bxb1-mediated

recombination (**Fig. 1A**; **Fig. S1C-E**). Following selection for successful recombinants, cells were split and transfected in duplicate with the dual reporter construct (**Fig. 1B**), and simultaneously treated with doxycycline (dox) to express the proteome editors. We chose glucokinase (GCK) as the target protein for this proof-of-principle experiment, primarily because GCK is monomeric, moderately sized (52 kDa) and cytoplasmically localized.

Replicate cultures were harvested, lysed, and divided for two immunoprecipitations (IPs): anti-FLAG to enrich the target protein and anti-Myc to enrich the standard protein (**Fig. 1D**). Circular RNA barcodes co-immunoprecipitated in each IP were reverse-transcribed, PCR-amplified, and subjected to massively parallel sequencing (**Fig. 1D**). To quantify the effects of individual editors on target protein abundance, barcode counts were first tabulated for each of the four samples (two biological replicates × two IPs; **Table S3**). For each barcode–design pair in the barcode dictionary, read counts were normalized to total reads per sample to obtain a per-sample read proportion for each barcode.

A key technical concern is whether recombination to the landing pad bottlenecked the complexity of the original library. To test this, we tallied the number of barcodes associated with each design. Reassuringly, “summary read proportions” were highly correlated between the plasmid library and the experimental libraries (Pearson’s r = 0.98, Spearman’s ρ = 0.98). However, ~2% of designs exhibited substantial “dropout” when introduced to cells (**Fig. S3A**). This phenomenon was consistent between replicates, but differed across effector targets (**Fig. S3B**). For example, effector handles designed to bind DUBs were 8.5-fold more likely to cause >50% dropout than those designed to bind other effector targets (10.2% of DUB binders [79/776] vs. 1.2% [110/9,200] of other binders). We speculate that these designs are toxic to HEK293 cells when highly expressed, either intrinsically or through their interaction with effectors.

After excluding barcodes with low representation ($<10^{-7}$ of reads in any of the four samples) and then designs represented by fewer than three barcodes, 9,715 designs (95%) were retained for further analysis. The effects of effector handle designs on the ratio of FLAG-to-Myc read proportions spanned a three-fold range and were reproducible between technical replicates (**Fig. 2B**; comparing the median ratio across all barcodes associated with a given design in Rep-1 vs. Rep-2: Pearson’s r = 0.72, Spearman’s ρ = 0.67; **Table S4**). Across the two replicates, there were 3,699 effector designs that significantly stabilized or destabilized the target protein ($p < 0.05$; Wilcoxon signed-rank test in which the FLAG vs. Myc read proportion rank distributions are compared for barcodes associated with a given design; p values were corrected for multiple-hypothesis testing using the Benjamini-Hochberg procedure). Among these, 2,016 (55%) were significant in both replicates, with perfectly concordant directional effects (2,016/2,016 [100%]). This directional concordance overwhelmingly extended to

designs significant in only one replicate (1,575/1685 [94%]). Encouraged by these results, we merged the sets of design-associated Flag and Myc read proportions observed in the two replicates (**Table S5**). In the combined data, 55% (5,299/9,715) of designed effector handles significantly stabilized or destabilized the target protein (p < 0.05; Wilcoxon signed-rank test; p values corrected using the Benjamini-Hochberg procedure).

Although we are using conservative procedures in testing for statistical significance (non-parametric test; hypothesis control), we performed an additional check in which we simply permuted the barcode-design assignments and recalculated significance by the same procedures. In the resulting quantile-quantile plot, we observe a massive excess of small p-values with the true barcode-binder assignments, while the permuted barcode-binder assignments result in a distribution of p-values that closely tracks expectation (**Fig. 2C**).

Overall, these analyses show that most designed effector handles (**Table S6**) exhibit significant, reproducible, directionally consistent effects on the levels of the target protein. However, most of these effects are small in magnitude. For our further investigations, we focused our attention on the 481 bifunctional proteome editors with the largest effect sizes, including 277 “degraders” (p < 0.05 & median log2(fold-change) < -0.5) and 204 “stabilizers” (p < 0.05 & median log2(fold-change) > 0.3) (**Fig. 2D**). With these thresholds, functional designs were obtained for both the UPS (130 degraders, 64 stabilizers) and autophagy (147 degraders, 140 stabilizers) pathways. Of note, degraders were more likely to have stronger effect sizes than stabilizers, *e.g.* 35 degraders, but no stabilizers, altered target protein abundance by >2-fold (**Fig. 2E**).

Our “hit rate” in terms of whether designs would succeed as either degraders or stabilizers varied substantially by effector, from 2% for GABARAPL2 (5/231) to 12% for UCHL1 (20/166) (**Fig. 2F**). As noted above, FBXL12 and the three ATGs were previously classified as degraders, while TRAF2 and the two DUBs were classified as stabilizers[69]. However, only FBXL12 and GABARAPL2 were fully concordant with this prior. For example, 9/9 successful FBXL12 binders were degraders rather than stabilizers. In contrast, successful designs for the other effectors included both degraders and stabilizers (**Fig. 2F**). For example, we expected binders that recruited the target protein to either of the two DUBs (UCHL1 and USP38) to increase its levels. However, not only did these two effectors have the highest “hit rates” (12% and 8%, respectively), their associated designs were mostly degraders (14/20 [70%] and 40/44 [91%], respectively).

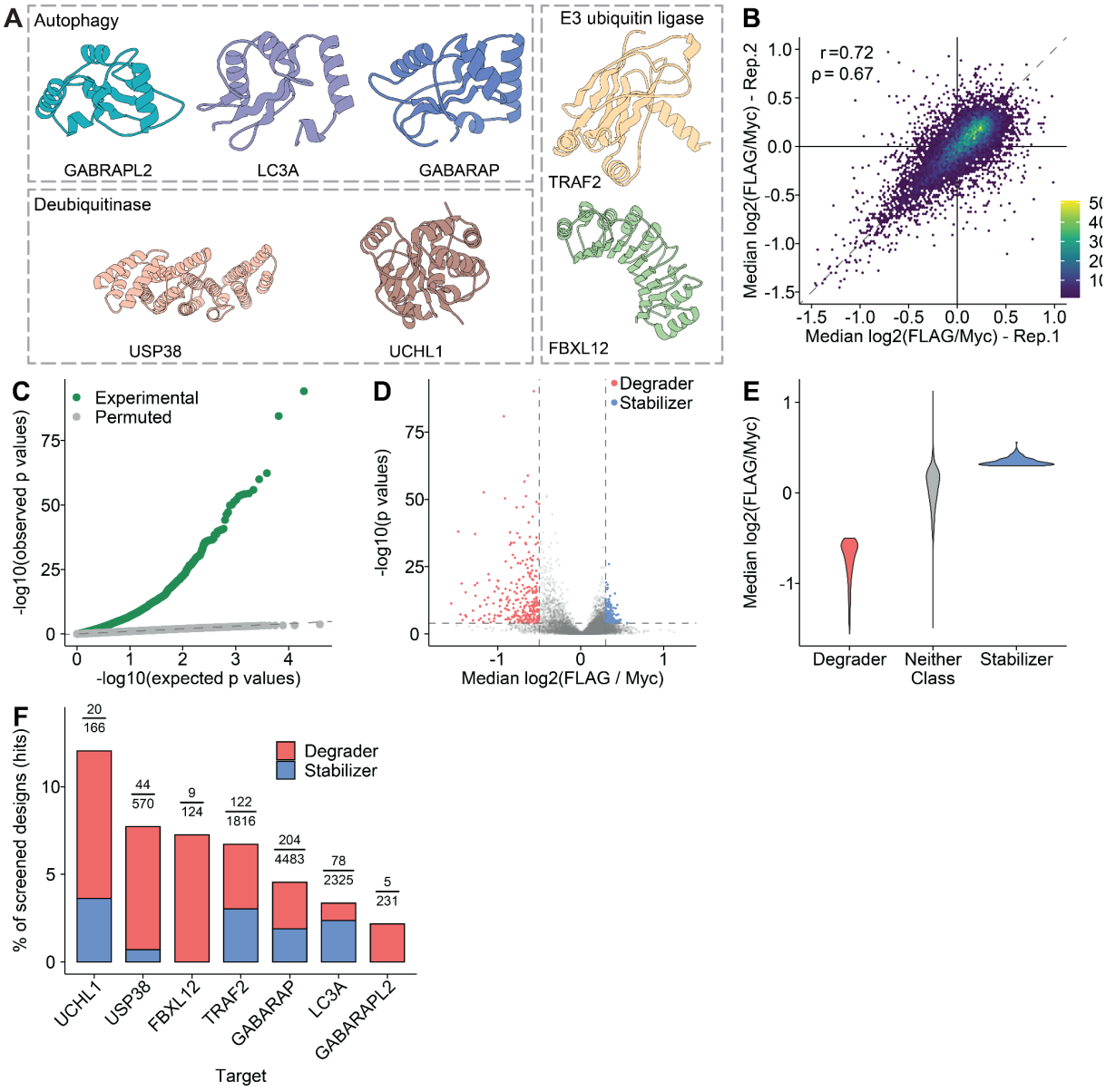
A
Autophagy
GABRAPL2
LC3A
GABARAP
Deubiquitinase
USP38
UCHL1
E3 ubiquitin ligase
TRAF2
FBXL12
B
r=0.72
ρ= 0.67
Median log2(FLAG/Myc) - Rep.2
Median log2(FLAG/Myc) - Rep.1
C
Experimental
Permuted
-log10(observed p values)
-log10(expected p values)
D
Degrader
Stabilizer
-log10(p values)
Median log2(FLAG / Myc)
E
Median log2(FLAG/Myc)
Degrader
Neither
Stabilizer
Class
F
% of screened designs (hits)
20/166
44/570
9/124
122/1816
204/4483
78/2325
5/231
Degrader
Stabilizer
UCHL1
USP38
FBXL12
TRAF2
GABARAP
LC3A
GABARAPL2
Target

Figure 2. Identification of functional effector handle designs with LABEL-seq.

A) Structures of effector proteins against which *de novo* protein binders were computationally designed. **B)** Hexagonal bin plot comparing the median log2(FLAG/Myc) fold-change for each design between two replicates. **C)** Quantile-quantile plot showing excess of low p-values (green points) in observed data, but no such excess when binder-barcode associations are permuted (grey points). Dashed line corresponds to null (x=y). **D)** Volcano plot showing effect sizes (*x*-axis; median log2(FLAG/Myc) fold-change) and significance (*y*-axis; -log10(p-value)) for each effector handle design. **E)** Violin plot showing effect distributions for effector handle designs classified as degraders (red; n = 277; $p < 0.05$ & median log2(fold-change) < -0.5), stabilizers (blue; n = 204; $p < 0.05$ & median log2(fold-change) > 0.3), or neither (gray). **F)** Barplot showing the percentage of binders classified as hits (y-axis) for each of 7 effector targets. Each bar indicates the percentage of hits classified as degraders (red) or stabilizers (blue). Above each bar, we also list the fraction on which the percentage was calculated (number of hits over number of designs for that effector).

3.3.4 Singleton validation of bifunctional protein degraders and stabilizers

We next sought to validate selected bifunctional protein degraders and stabilizers identified in the ~10,000-plex cellular screen. We chose not to pursue further validation of designed effector handles targeting GABARAPL2 due to the paucity of designs exhibiting large effect sizes. For the remaining six effectors, we selected the two degraders and one stabilizer for singleton validation experiments (**Fig. S4**; **Table S7**). As a negative control, we also selected a binder that consistently exhibited no effect on the levels of the target protein. We then generated constructs encoding a constitutively expressed NbALFA (target handle; C-terminus) fused to one of these 19 binders (effector handle; N-terminus). These constructs also express mTagBFP2 from an independent EF1 promoter as a transfection marker.

To measure protein abundance changes in an assay orthogonal to the original LABEL-seq screen, we employed a bicistronic GFP-IRES-mCherry protein abundance reporter that enables ratiometric quantification of GFP-fusion protein abundance normalized to mCherry. We generated two dox-inducible variants of this reporter. The first reporter encoded a FLAG-GFP-GCK-ALFA-IRES-mCherry-2A-Puro cassette. As a negative control, we sought to abolish the interaction between the binder and the ALFA-tagged GFP protein. To accomplish this, a second version of the reporter was cloned in which the ALFA tag was deleted in-frame, leaving FLAG-GFP-GCK only. Stable polyclonal HEK293T lines expressing either reporter were generated via piggyBac transposition of the reporter plasmids. Effector handle-encoding plasmids were transiently transfected into each reporter line in triplicate. Reporter expression was induced with dox for 24 hrs post-transfection, and then the GFP:mCherry ratios of mTagBFP2-positive cells were assessed by flow cytometry at 48 hrs post-induction.

Within each reporter line, GFP:mCherry ratios measured in the presence of each protein editor (*i.e.* NbALFA fused to an effector handle) were normalized to the mean GFP:mCherry ratio of the negative control binder (**Fig. 3A-B**, grey points). For 12/12 degraders, we observed ratios consistently below that of the no-effect control (**Fig. 3A**, red points), while for 5/6 of the stabilizers, we observed ratios consistently above that of the no-effect control (**Fig. 3A**, blue points). Altogether, 17/18 (94%) effector handle designs subjected to singleton validation exhibited directional concordance with expectation. The absence of the ALFA tag in the GFP:mCherry reporter abolished the effects of these effector handles, providing further evidence of their specificity (**Fig. 3B**). These effects can also be visualized as a slope plot to more directly compare the GFP:mCherry ratios induced by each binder in each cell line (**Fig. 3C**). The GABABRAP-D1 effector handle induced the strongest degradation effect, with a mean normalized GFP:mCherry ratio of 0.26 indicating a 3.9-fold decrease in abundance of the reporter protein. TRAF2-S1 induced the strongest stabilization effect, a 1.4-increase in levels of the reporter protein.

We also compared effects estimated from the ~10,000-plex experiment with LABEL-seq against effects measured via the orthogonal GFP:mCherry reporter assay, and found these to be highly concordant (Pearson's r = 0.93, Spearman's ρ = 0.98; **Fig. 3D**). Finally, we extended these validations from HEK293T cells to HepG2 cells and once again observed strong concordance (Pearson's r = 0.91, Spearman's ρ = 0.93; **Fig. S5**), suggesting that these bifunctional proteome editors may function in all cell types where the effector proteins to which they bind are active.

Finally, we then examined the design models of binder-effector complexes for effector handles that induced >50% degradation (**Fig. 3E**). Designs against GABARAP, FBXL12, USP38, and TRAF2 each targeted a conserved binding interface within their respective targets. The highly performing degrader targeting FBXL12 overlaps with the C-terminus of FBXL12, which is normally responsible for engaging native substrates for degradation. This observation suggests that the conformational flexibility of the C-terminus of FBXL12 may permit binder engagement and substrate redirection, consistent with the intended design principle of redirecting FBXL12 to degrade the protein recruited by the target handle.

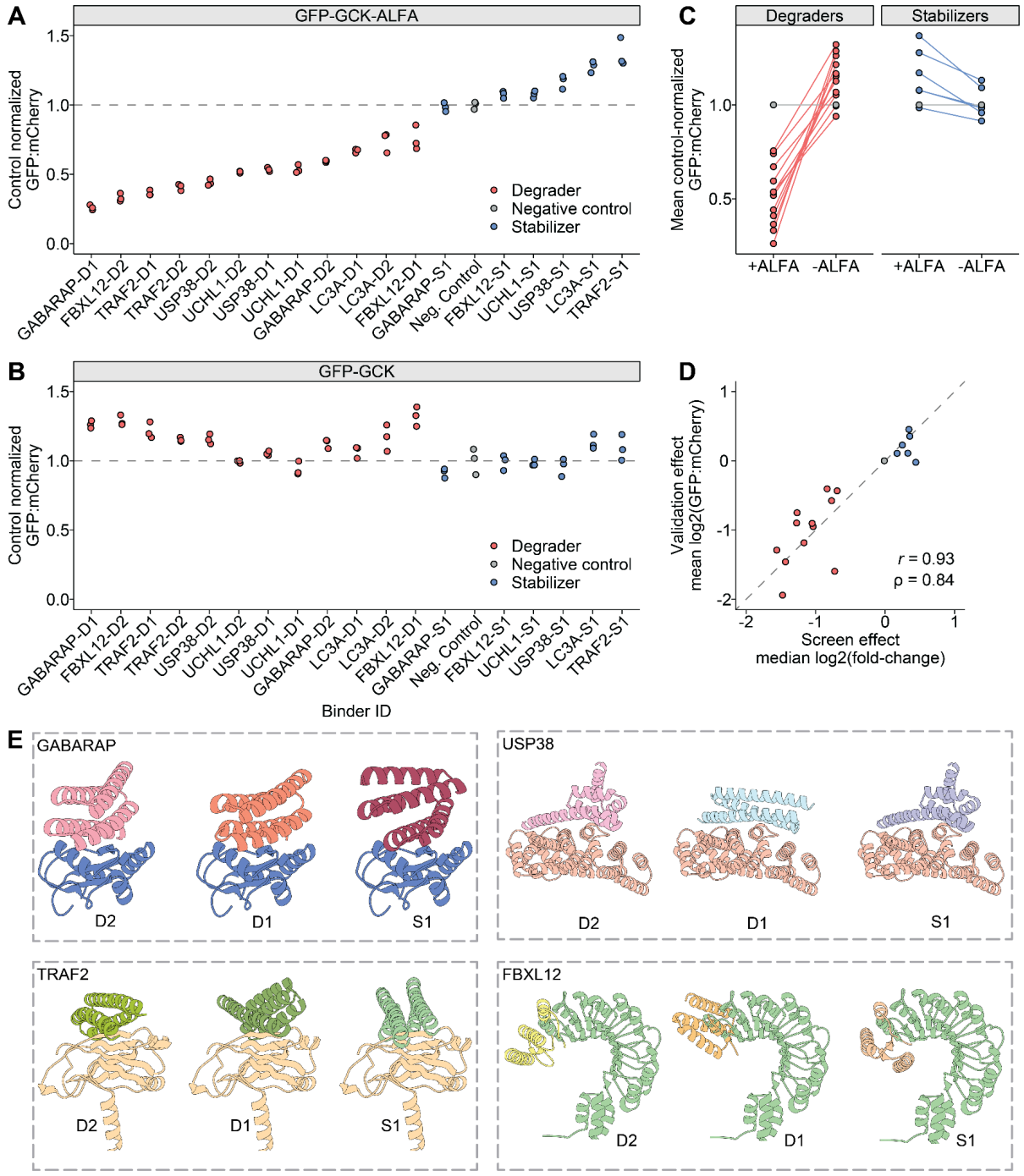
A
GFP-GCK-ALFA
Control normalized GFP:mCherry
Degrader
Negative control
Stabilizer
GABARAP-D1
FBXL12-D2
TRAF2-D1
TRAF2-D2
USP38-D2
UCHL1-D2
USP38-D1
UCHL1-D1
GABARAP-D2
LC3A-D1
LC3A-D2
FBXL12-D1
GABARAP-S1
Neg. Control
FBXL12-S1
UCHL1-S1
USP38-S1
LC3A-S1
TRAF2-S1
B
GFP-GCK
Binder ID
C
Degraders
Stabilizers
Mean control-normalized GFP:mCherry
+ALFA
-ALFA
D
Validation effect mean log2(GFP:mCherry)
Screen effect median log2(fold-change)
r = 0.93
ρ = 0.84
E
GABARAP
USP38
TRAF2
FBXL12
D2
D1
S1

Figure 3. Singleton validation of functional designed effector handles.

A) Dot plot of control-normalized GFP:mCherry ratios measured in cells expressing a EGFP-GCK-ALFA reporter and indicated candidate effector handle designs. Each point represents a biological replicate and the dashed line at 1 indicates the mean of the negative-control effector handle (grey points). Effector handles are ordered along the x-axis by mean normalized GFP:mCherry ratio, Dx and Sx (where x is a numeral) indicate degraders and stabilizers, respectively. **B)** Same as in panel A but measured in cells expressing a EGFP-GCK reporter lacking the ALFA tag. Effector handles in panel B maintain the same left-to-right ordering as in panel A. **C)** Slope plot faceted by effector class showing mean GFP:mCherry ratios measured for each binder in each cell line. Binder measurements in each line are connected with a line to highlight the effect of deleting the ALFA tag from the reporter construct. **D)** Scatterplot comparing measured effects from the ~10,000-plex screen and LABEL-seq (x-axis) vs. the singleton validations (y-axis). **E)** Visualizations of predicted structural models for indicated effector-binder pairs. Identifiers in each panel (i.e. D2, D1, S1) are matched to identifiers in panels A and B.

3.3.5 Targeting an endogenous protein with designed effector handles

We next asked whether designed effector handles could mediate degradation of endogenous proteins by replacing the ALFA-targeting NbALFA handle with a binder specific to an intracellular target. We focused on the anti-apoptotic BCL-2 family protein MCL1, which has a short half-life due to UPS-mediated degradation, is overexpressed in many cancers, and has previously been targeted by high-affinity designed binders[55].

The top degrader design for each of six effectors (**Table S7**) was cloned into a lentiviral backbone as a bivalent fusion with an anti-MCL1 binder. MDA-MB-231 breast cancer cells were transduced with individual MCL1-degrader fusions, and endogenous MCL1 protein levels were assessed by Western blot 72 hrs post-transduction (**Fig. S6**). MCL1 was expressed at low levels in untransduced cells and in cells expressing a GFP-specific nanobody (NbGFP) (**Fig. S6**, lanes 1-2). Notably, MCL1 levels were elevated in cells expressing a negative control construct in which the MCL1 binder was fused to a non-functional effector handle from the LABEL-seq screen. This upregulation is consistent with prior studies showing that MCL1 inhibitors or binders can stabilize MCL1 by blocking native ubiquitination, leading to accumulation of non-functional MCL1 that is unable to sequester pro-apoptoptic factors[77–79].

Accordingly, MCL1 levels were normalized to those observed with the negative control MCL1-non-functional binder fusion. Among the tested constructs, GABARAP and LC3A effector handles showed the strongest effects, reducing MCL1 protein levels to 50% and 27% of control levels, respectively (**Fig. S6**). These findings mirror the strong degradation effects of autophagy-related effector handles observed in our ~10,000-plex experiment and singleton validation assays. In contrast, fusions incorporating UPS-related effectors showed only modest or stabilizing effects on endogenous MCL1, despite their activity in the GCK–ALFA reporter system, suggesting target- or cell line-specific dependencies. Testing several effector handles for each of these autophagy effector confirmed that the D1 designs for GABARAP and LC3A (*i.e.* the ones exhibiting the strongest effects in both the multiplex discovery and singleton validation experiments) also yielded the strongest activity against endogenous MCL1 (**Fig. 4A**).

To evaluate the functional consequences of MCL1 degradation, we quantified apoptosis and cell death in MDA-MB-231 cells expressing GABARAP-targeting fusions. Consistent with their degradation effects, GABARAP-D1—which produced the strongest degradation by Western blot—also induced increased apoptosis and cell death, whereas non-degrading GABARAP-targeting fusions had no effect (**Fig. 4B**; **Fig. S7**). Previous studies have shown that the functional (anti-apoptotic) pool of MCL1, rather than total abundance, determines the apoptotic threshold[80]. The bifunctional proteome editor targeting MCL1 and GABARAP may therefore act through dual mechanisms: promoting degradation while simultaneously preventing sequestration of pro-apoptotic

factors, thereby neutralizing MCL1's protective function even when residual protein remains. These results indicate that degradation of the anti-apoptotic pool of MCL1 triggers apoptosis and cell death, consistent with prior findings from PROTAC-based degraders[79,81]. Collectively, these data demonstrate that de novo designed effector handles can drive degradation of an endogenous oncoprotein and elicit functional phenotypes in human cancer cells.

3.3.6 Perturbation of mitochondrial organization

Beyond modulating individual proteins, we wondered whether designed binders could be used to redirect or reorganize subcellular structures. To test this, we performed live-cell imaging of HEK293T cells co-transfected with plasmids encoding ALFA–GFP–FIS1 and NbALFA–effector handle fusions, and visualized mitochondrial networks by confocal microscopy. FIS1 (mitochondrial fission 1 protein) localizes to the outer mitochondrial membrane; thus, ALFA–GFP–FIS1 allows visualization of mitochondria via the GFP signal (**Fig. 4C**, GFP channel, green). GFP fluorescence co-localized with mitochondria stained with MitoTracker (**Fig. 4C**, MitoTracker, magenta), confirming correct targeting.

Strikingly, cells co-expressing NbALFA fusions with GABARAP or LC3A effector handles displayed aggregated mitochondrial networks (**Fig. 4C**, top rows), in sharp contrast to the dispersed, reticular morphology seen in control cells co-expressing NbALFA alone or in BFP-only controls (**Fig. 4C**, bottom rows). Given the established roles of GABARAP and LC3A in autophagy, these aggregates likely reflect synthetic tethering or sequestration of mitochondria into autophagosomal compartments[57,58]. Although the precise mechanism remains to be determined, the binder-dependent clustering of mitochondria demonstrates that designed effector handles can act beyond single-protein modulation to reprogram organelle-scale organization, illustrating the potential for programmable subcellular remodeling.

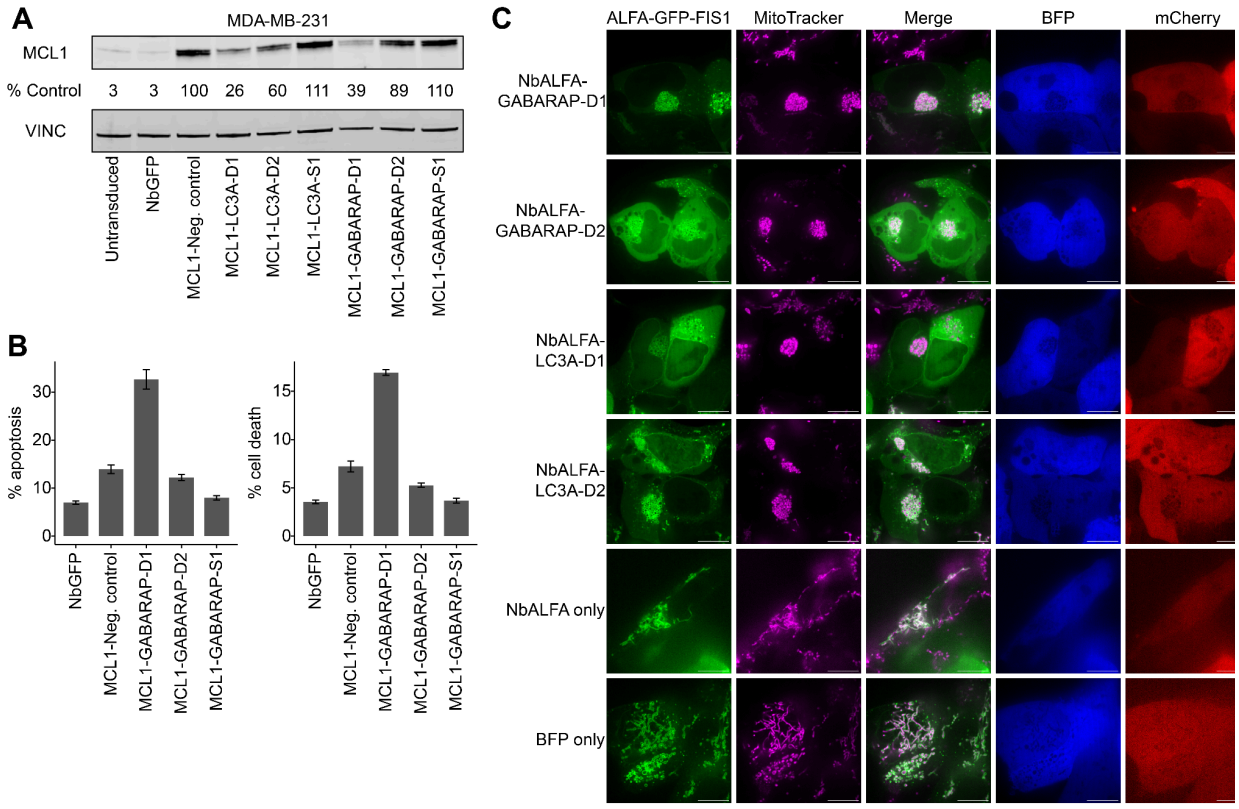

Figure 4. Designed autophagy effectors mediate endogenous protein degradation, promote cancer cell apoptosis, and perturb mitochondrial networks.

A) Western blot analysis of MCL1 levels in MDA-MB-231 cells, 72 hours after transduction with indicated MCL1-effector binder fusions, as well as untransduced and GFP-nanobody transduced controls. **B)** Percentage of MDA-MB-231 cells positive for apoptotic (left) or cell death markers (right) as measured by flow cytometry, 72 hours after transduction with indicated MCL1-GABARAP binder fusions or controls. Error bars indicate ± SEM (n = 3). **C)** Confocal microscopy images of ALFA-GFP-FIS1 (green) and mitochondria (magenta). Images were taken 24 hours after co-transfection of HEK293T cells with plasmids encoding ALFA-GFP-FIS1 and indicated MCL1-effector binder fusions or controls. mCherry and BFP are transfection markers for the ALFA-GFP-FIS1 and MCL1-effector binder fusion plasmids respectively.

3.4 DISCUSSION

Here we describe a multiplex framework for quantifying the intracellular activities of *de novo* designed proteome editors. In this proof-of-concept, we evaluated ~10,000 designed binders of components of the UPS and autophagy pathways as effector handles within bifunctional proteome editors. Remarkably, the majority of designed binders reproducibly altered the abundance of a target protein, although most effects were modest in magnitude. We validated a subset of the most active “degraders” and “stabilizers” in an orthogonal protein abundance assay, and found these exhibited strong concordance with the measurements obtained in the multiplex experiment, both in terms of directionality and magnitude of effects.

To date, the field of targeted protein degradation has been dominated by small molecules such as PROTACs. However, their development is constrained by the limited availability of well-characterized effector ligands and the requirement for chemical tractability of target proteins. By integrating *de novo* protein design with a multiplex functional screen based on LABEL-seq, we establish a general framework that circumvents these constraints. Computational design enables generation of bespoke binders to virtually any effector, target, or interface, while the multiplex assay provides a scalable means of measuring and optimizing their activities within living cells.

For this initial demonstration, we employed an ALFA-tagged GCK reporter to evaluate designed effector handles. Even for binders targeting the same effector interface, we observed a broad spectrum of effects on reporter abundance, suggesting that attributes such as binding affinity, dissociation rate, intracellular stability, or expression may modulate efficacy. Further studies will be necessary to elucidate the precise mechanisms and effector dependencies of our designed binders, both to validate their mode of action and to guide future iterations of design. Although we designed our study with the expectation that recruiting certain effectors would result in either degradation or stabilization, our results suggest a continuum of effects rather than a binary outcome. Furthermore, as our experiment produced quantitative measurements for hundreds to thousands of designs per effector, investigations of the biophysical features that determine the directionality and magnitude of these outcomes may be enabled by the data reported here.

GCK’s long half-life likely limited our dynamic range for detecting turnover, and more generally, the portability of designed effector handles across different targets remains an open question. Consistent with this, effector handles that recruited autophagy effectors (*e.g.* LC3A, GABARAP) successfully degraded the endogenous oncoprotein MCL1 when paired with an anti-MCL1 binder, whereas those targeting the UPS did not. These findings suggest that the efficacy of bifunctional proteome editors will depend on both the effector and the context of the recruited target.

Looking forward, this framework opens several avenues for scaling and diversification:

First, *expanding effector diversity* (**Fig. 5A**). The present study targeted only seven effectors, but the same approach could be extended to hundreds of components across degradation, stabilization, and trafficking pathways. Such expansion would yield a rich toolkit of effector handles for proteome editing across cell types and physiological contexts.

Second, *systematic design and validation of target handles for the entire human proteome* (**Fig. 5B**). By fixing a validated effector handle and instead varying the target handle, large-scale intracellular screens could identify binders for virtually every human protein). In principle, this would create a comprehensive “phone book” of functional binder pairs—enabling the programmable recruitment of any protein to any effector, location, or function—analogous to how base-pairing rules empower genome editing.

Third, *mechanistic and modeling advances* (**Fig. 5C**)**.** Unlike traditional *in vitro* binding assays, our assay measures designs in the native cellular milieu and yields quantitative functional data for thousands of designs in a single experiment. These datasets, coupled with structural models, provide fertile ground for understanding how designed interactions translate to function within cells and for improving next-generation generative models for binder design.

Fourth, *extending proteome editing beyond abundance* (**Fig. 5D**). Beyond modulating protein levels, bifunctional proteome editors could be adapted to control a broader range of protein and cellular phenotypes, *e.g.* as we have shown here by perturbing mitochondrial networks. The same multiplex framework could also incorporate conditional regulation—via small molecules, light, or engineered feedback—to achieve tunable and context-dependent control of intracellular processes. Together, these extensions would transform proteome editors from degraders or stabilizers into general-purpose actuators for programmably reshaping cellular function.

Collectively, these advances point toward a future in which protein design and multiplex functional screening converge into a programmable toolkit for remodeling the intracellular proteome and the cell itself.

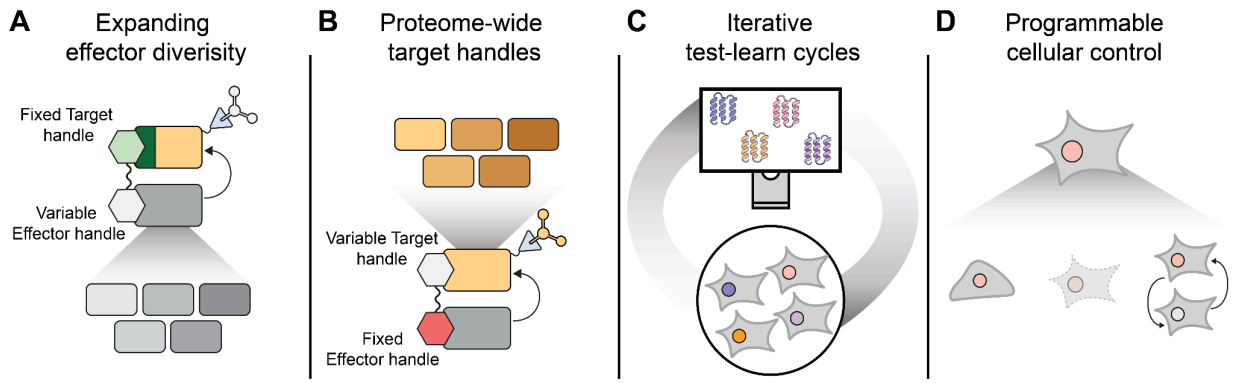

Figure 5. Potential applications of multiplex cellular screens of de novo designed proteins

A) A fixed target handle (green) can be fused to a library of designed effector handles with the goal of enabling diverse effector outcomes to be programmed. **B)** A fixed effector handle (red) can be fused to a library of designed target handles with the goal of broadening the set of recruitable proteins to the entire human proteome. **C)** Iterative cycles of effector or target handle design and experimental characterization may yield data which serves the improvement of the design of intracellularly functional binders. **D)** Installation of conditional control of protein editors may enable programmable cellular control. Potential applications include programming inducible changes to target protein abundance or localization, cell state, cell death, and cell-to-cell communication.

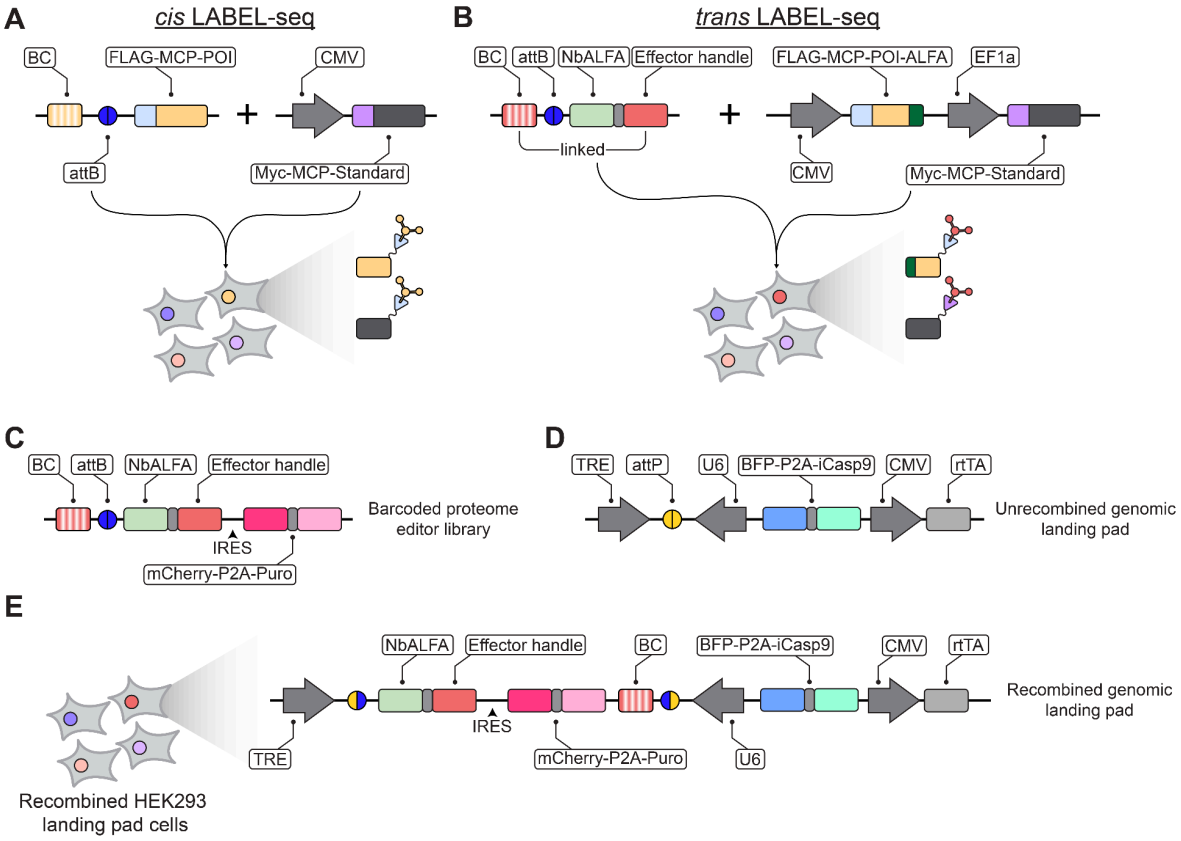
A
cis LABEL-seq
BC
FLAG-MCP-POI
CMV
attB
Myc-MCP-Standard
B
trans LABEL-seq
BC
attB
NbALFA
Effector handle
FLAG-MCP-POI-ALFA
EF1a
linked
CMV
Myc-MCP-Standard
C
BC
attB
NbALFA
Effector handle
IRES
mCherry-P2A-Puro
Barcoded proteome editor library
D
TRE
attP
U6
BFP-P2A-iCasp9
CMV
rtTA
Unrecombined genomic landing pad
E
NbALFA
Effector handle
BC
BFP-P2A-iCasp9
CMV
rtTA
IRES
TRE
mCherry-P2A-Puro
U6
Recombined genomic landing pad
Recombined HEK293 landing pad cells

Figure S1. Schematic of genomic landing pad and cis versus trans LABEL-seq assays

A) Diagram of *cis* LABEL-seq in which the barcode (BC) and FLAG-MCP-POI protein are encoded from the same DNA molecules. **B)** Diagram of *trans* LABEL-seq in which barcode (BC) and FLAG-MCP-POI-ALFA protein are encoded from distinct DNA molecules. **C)** A barcoded proteome editor library prior to recombination at the landing pad. The barcoded proteome editor library also has a mCherry-P2A-Puro cassette that is translated from an upstream internal ribosome entry site (IRES). **D)** Diagram showing the single-copy genomic landing pad prior to recombination, as well as a barcoded element library. The landing pad contains an attP site flanked by convergent tetracycline response element (TRE) and U6 promoters, a BFP-P2A-iCasp9 cassette, and a CMV-driven reverse tetracycline transactivator (rtTA). A barcoded element library with an attB site can be recombined with the genomic attP site via Bxb1-mediated recombination. **E)** Diagram showing the single-copy genomic landing pad following recombination, which places the library element and IRES-mCherry-P2A-Puro cassette under control of the TRE promoter and the circular RNA barcode under control of the U6 promoter.

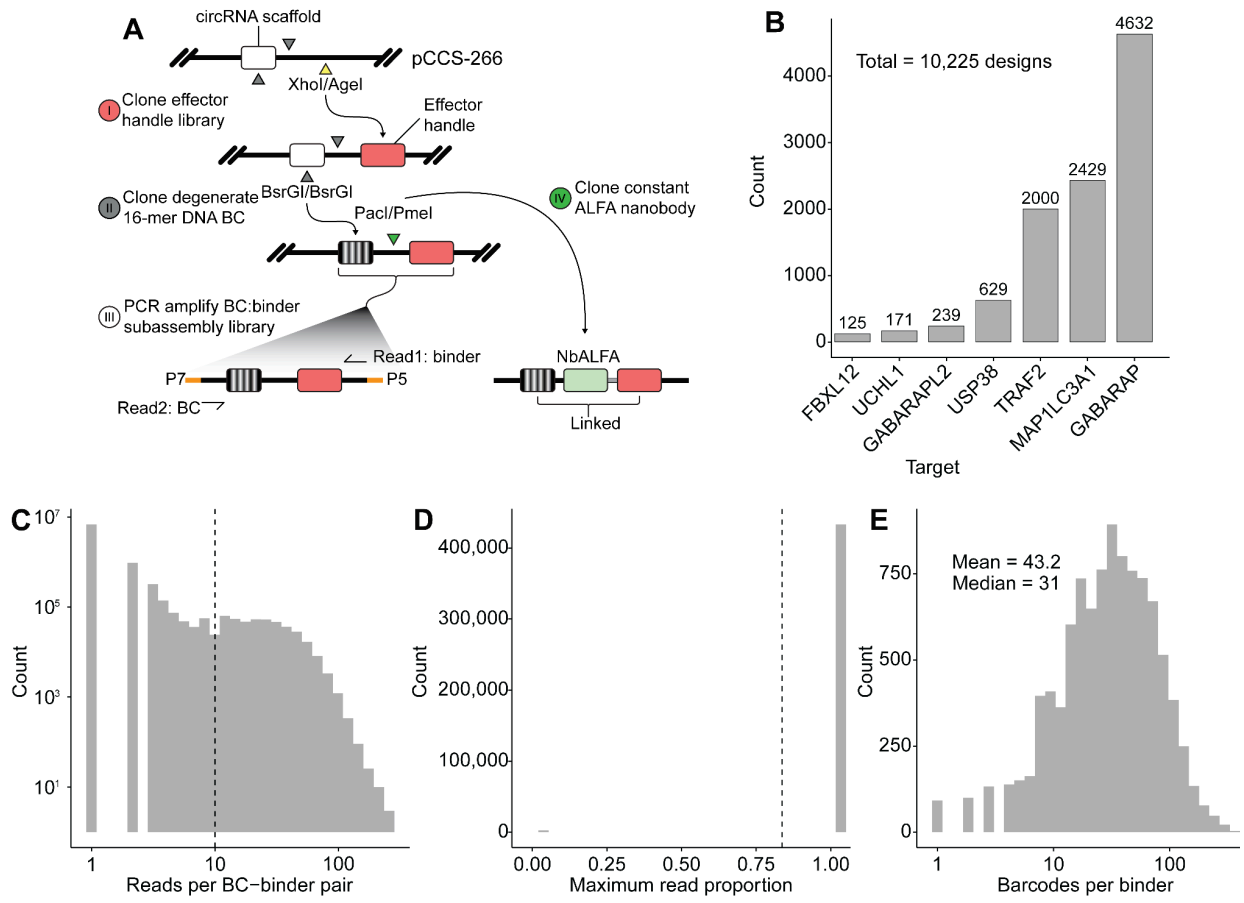

Figure S2. Plasmid library composition and characterization

A) Schematic detailing how the barcoded binder library was cloned, including subassembly. **B)** Barplot showing the number of designs generated for each target effector. **C)** Histogram showing the distribution of the number of sequencing reads per barcode-binder pair. The vertical dashed line indicates the threshold above which barcode-binder pairs were considered valid. **D)** Histogram showing the maximum read proportion of individual barcodes. For an individual barcode, the proportion of reads associated with any binder was computed to ascertain the fidelity of barcode binder association. The vast majority of barcodes were associated with a single binder. **E)** Histogram showing the distribution of the number of barcodes associated with each binder.

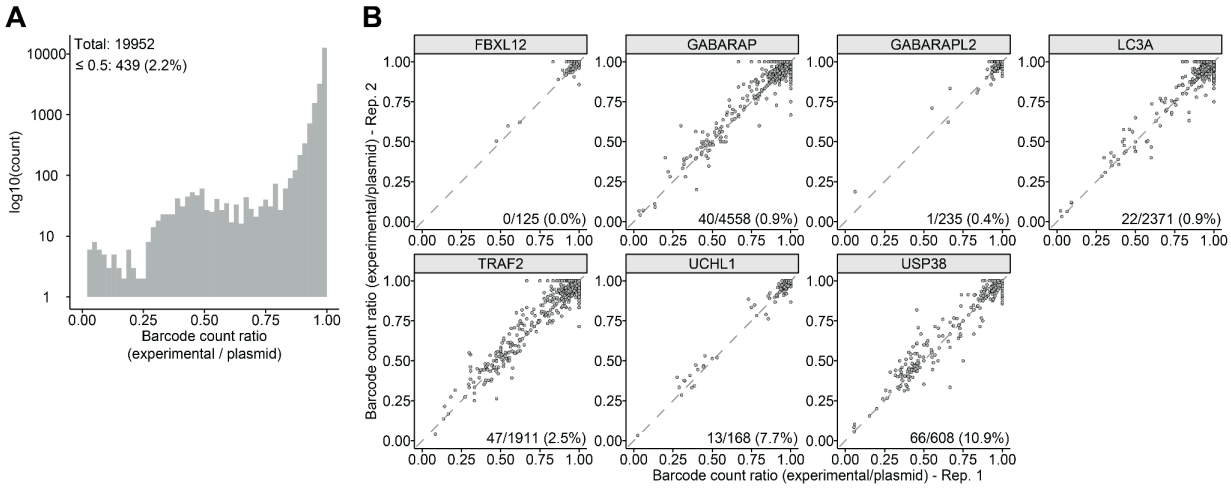

Figure S3. Comparison of plasmid and experimental barcodes

A) Histogram showing distribution of the ratio of experimentally recovered barcodes to plasmid library barcodes for each binder across two transfection replicates. The total number of binders represented in the plot is indicated, as well as the number and percentage of binders with a ratio ≤ 0.5 in either of the two replicates. The distribution suggests that some binders are "dropping out" of the experiment when expressed in cells. **B)** Scatter plots comparing the ratio of experimentally recovered barcodes to plasmid library barcodes for each binder between the two transfection replicates. Each point represents an individual binder. For each effector target, the number of binders with a ratio ≤ 0.5 in both replicates is indicated, as well as the percentage. The patterns suggest that the drop-out phenomenon is reproducible between replicates, and moreover is more likely to occur with binders designed against some effectors (*e.g.* USP38; 10.9%) than others (*e.g.* FBXL12; 0%).

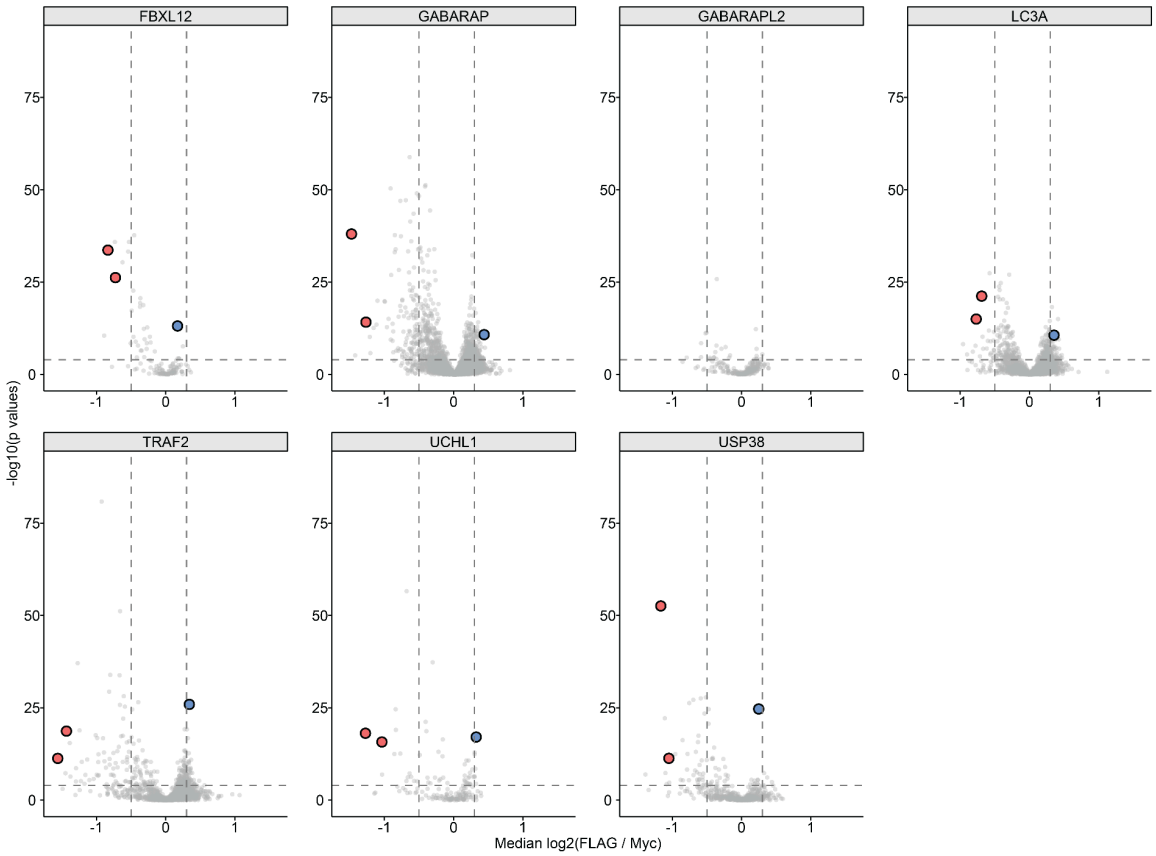

Figure S4. Assessment of hit rate per effector and selection of effector handles for validation experiments.

Volcano plots comparing the distribution of effect sizes (x-axes) and significance (y-axes) faceted by effector. Degrader (red points) and stabilizer (blue points) effector handles selected for validation are indicated.

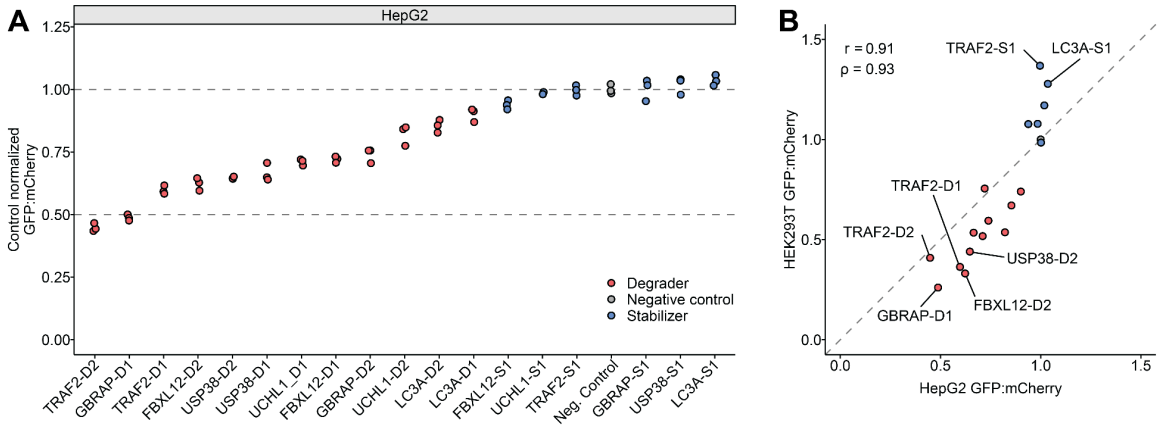

Figure S5. Binder-induced GCK degradation validates in an alternative cell line.

A) Dot plot of control-normalized GFP:mCherry ratios measured in HepG2 cells expressing a GFP-GCK-ALFA reporter and indicated candidate effector designs. Each point represents a transfection replicate and the dashed line at 1 indicates the mean of the negative-control effector (grey points). Effectors are ordered along the x-axis by mean normalized GFP:mCherry ratio. **B)** Scatterplot comparing measured effects from the singleton validation in HepG2 cells (x-axis) vs. HEK293T cells (y-axis). Pearson (r) and Spearman (ρ) correlation values are shown.

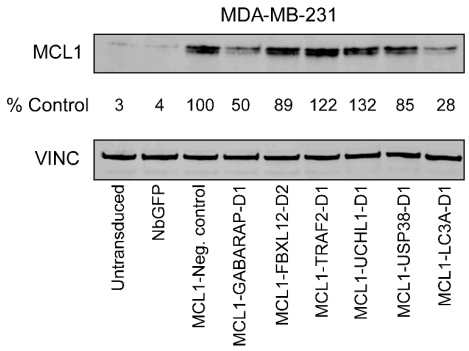

Figure S6. Assessment of effects of bifunctional MCL1-effector binder fusions on endogenous MCL1 levels.

Western blot analysis of MCL1 levels in MDA-MB-231 cells, 72 hrs after transduction with indicated MCL1-effector binder fusions. Untransduced and GFP-nanobody (NbGFP) transduced cells serve as negative controls.

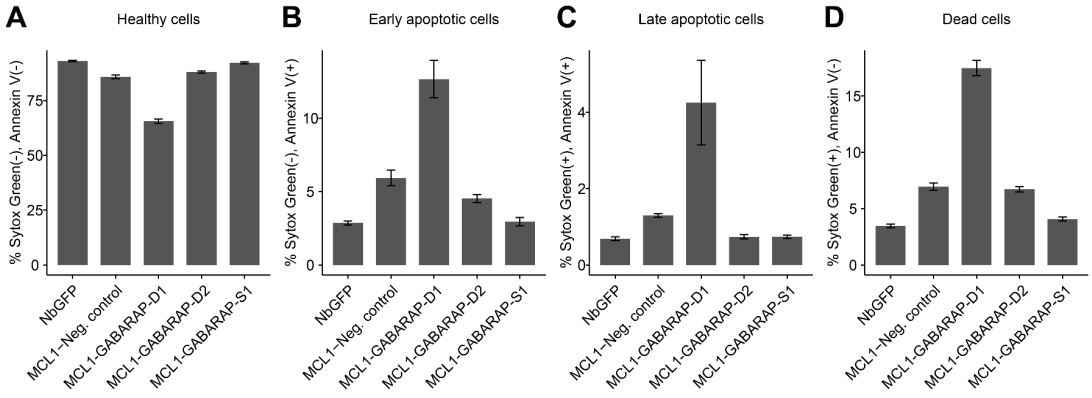

Figure S7. Characterization of apoptosis induction by bifunctional MCL1-GABARAP binder fusions

A-D) Bar plots showing the percentage (y-axis) of cells in four states in the presence of various MCL1-effector binder fusions (x-axis). States were defined as: healthy (**A**), early apoptotic (**B**), late apoptotic (**C**), and dead (**D**) based on the indicated flow cytometry parameters (y-axis labels). Bars indicate the mean of three transduction replicates, error bars indicate ± SEM.

3.5 MATERIALS & METHODS

3.5.1 Cell lines and cell culture

HEK293T (CRL-3216), MDA-MB-231 (HTB-26), and HepG2 (HB-8065) cell lines were purchased from ATCC. circRNA-protein landing pad (RPLP) HEK293 cells were generated as described previously[52]. HEK293T, MDA-MB-231, HepG2, and HEK293 RPLP cells were cultured in DMEM (Gibco). All media were supplemented with 10% FBS (Hyclone) and 1% penicillin-streptomycin (Gibco). All cells were grown with 5% CO_2 at 37 °C.

3.5.2 Binder design

To design de novo binders, we generated 200 backbones using RFdiffusion and generated 10 sequences per backbone with ProteinMPNN[66]. We used AlphaFold2[40] (AF2) with an initial guess and target templating to filter designs[82]. Design models that were predicted from AF2 with pae_interaction <20 were subjected to partial diffusion[65] for backbone optimization, followed by two rounds of ProteinMPNN and AF2 filtering. Designs were selected by filtering on pae_interaction <8 and pLDDT >88.

3.5.3 Binder library cloning

Barcoded binder libraries were cloned in a three step process (Fig. S1B). First, oligo libraries (Twist) encoding designed effector binders were PCR amplified and Gibson assembled (NEBuilder HiFi DNA Assembly, NEB) into XhoI/AgeI digested pCCS-266 plasmid backbone (Fig. S1B, step I). The Gibson assembly was cleaned (Clean and Concentrate kit, Zymo) and electroporated into Escherichia coli (NEB, C3020) which were then cultured at 30 °C for 20 hours at which point plasmid DNA was extracted (ZymoPURE II Plasmid Midiprep Kit). 1% of the electroporation was serially diluted and plated onto LB agar plates with ampicillin to quantify the number of transformants and ensure adequate library complexity.

Second, the library was barcoded via addition of a degenerate 16-mer DNA barcode to the backbone (Fig. S1B, step II). An oligonucleotide containing an internal stretch of 16 random bases was synthesized (IDT), PCR amplified to add Gibson homology handles, and Gibson assembled into BsrGI-digested backbone from Step I. As before, the Gibson assembly was cleaned and electroporated into Escherichia coli, and 1% of the electroporation was serially diluted and plated onto LB agar plates with ampicillin to quantify the number of transformants and ensure adequate library complexity. The barcode complexity was bottlenecked at this step by splitting various volumes of electroporated cells to different flasks (e.g. 1 mL of cells was split to 5 different flasks containing 25 µL, 75 µL, 150 µL, 200 µL, or 300µL) and these flasks were then cultured at 30 °C for 20 hours. The following morning, the number of transformants in the entire transformed volume was calculated and used to calculate the number of transformants inoculated into each flask. Plasmid DNA was extracted from the flask inoculated with

the volume of transformed cells that most closely aligned to the target number of barcodes.

Finally, DNA encoding the ALFA nanobody (NbALFA) was cloned into the library (Fig. S1B, step IV). A gBlock containing the NbALFA coding sequence was synthesized (IDT), PCR amplified, and Gibson assembled into PacI/PmeI digested barcoded binder library from Step II. 1% of the electroporation was serially diluted and plated onto LB agar plates with ampicillin to quantify the number of transformants and ensure adequate library complexity.

3.5.4 Binder-barcode subassembly

3.5.4.1 Preparation of subassembly library

Binders were associated to their corresponding barcodes via subassembly (Figure S1B, step III). Plasmid library DNA containing designed binders and degenerate 16-mer barcodes was subjected to two rounds of PCR. The first round of PCR utilized a forward primer that hybridizes upstream of the DNA barcode and a reverse primer that hybridizes downstream of the designed binder sequences. This first round PCR product was then subjected to a second round of PCR to introduce sequencing indices as well as P5 and P7 adapters for Illumina sequencing. The final amplicon was sequenced on a NextSeq2000 Illumina sequencer with Read 1 capturing the binder sequence and Read 2 capturing the barcode sequence.

3.5.4.2 Computational analysis of subassembly sequencing data

We implemented a Python script to pair each 16 nucleotide DNA barcode with its cognate designed binder sequence from demultiplexed, paired-end FASTQ files. The script requires as input a reference table of binder-encoding DNA sequences and their associated identifiers. From this reference, we first isolate the last 100 nucleotides of each binder sequence which is then reverse complemented. All possible single nucleotide changes of all reverse complemented sequences are then computed and stored for later reference.

Next, binder sequences are extracted from Read 1 by first finding a constant 10 nucleotide string immediately preceding the binder and subsequently extracting the next 100 nucleotides. These 100 nucleotides represent the last 100 nucleotides of the binder sequence and should correspond to a binder contained in the reverse complemented sequences processed previously. The extracted nucleotides are queried against the previously computed reference, and reads that match a single binder (or a single

nucleotide variant of a single binder) are labelled with that binders corresponding identifier.

Barcodes are then extracted from Read 2 in a similar manner. Reads are parsed for a constant 10 nucleotide string immediately preceding the 16 nucleotide barcode. The subsequent 16 nucleotides representing the barcode sequence are then extracted.

Each paired read is written to a file containing a single barcode-binder pair per row. Reads for each barcode-binder pair are subsequently summed, and the distribution visualized as a histogram (Fig. S1C). From the histogram, we set a threshold (i.e. 10 reads per barcode-binder pair) above which we consider barcode-binders pairs to be true. As a second filter, we compute the proportion of all unique associations for each barcode (e.g. 80% of barcode ATGC reads map to binder 1, 15% to binder 2, etc…). We take the maximum of the proportion value for each barcode and visualize it as a histogram (Fig. S1D) which shows that the vast majority of barcodes have a high proportion of their reads mapped to an individual binder. Barcodes with a maximum read proportion ≥ 80% are retained. These barcode-binder pairs (i.e. those with ≥ 10 reads & ≥ 80% of reads associated with a single binder) are then saved as a lookup table.

3.5.5 LABEL-seq screen

3.5.5.1 Library recombination

Barcoded binder libraries were recombined into RPLP HEK293T cells at a genomically integrated landing pad as previously described[52]. On day 0, four 15 cm tissue culture dishes were individually seeded with 18×10^6 RPLP HEK293T cells in complete DMEM. On day 1, cells were transfected with 20 µg of barcoded binder library (Fig. 1C) and 3 µg pCAG-NLS-HA-Bxb1 (Addgene #51271) using Lipofectamine 3000 according to the manufacturers instructions, and the media was replaced with complete DMEM on day 2. Expression of genomically integrated binders was induced on day 3 by addition of doxycycline (Sigma-Aldrich #D9891) to a final concentration of 2 µg mL^{-1}. Recombined cells were then selected on day 5 by treating cells with 1 nM AP-1903 (Selleckchem S9726) for 6 hours, after which media was replaced with complete DMEM supplemented with 2 µg mL^{-1} doxycycline. On day 7, 2 µg mL^{-1} puromycin (Thermo A1113803) was added to the media to maintain selection for recombined cells and cells were expanded until day 13 at which point cells were trypsinized, pelleted, resuspended in CELLBANKER 1 (Amsbio 11884) and frozen at -80 °C.

3.5.5.2 Immunoprecipitation of reporter proteins and their associated circular RNA barcodes

Frozen aliquots of RPLP HEK293T recombined with the barcoded binder library were thawed and expanded in complete DMEM supplemented with 2 µg mL^{-1} doxycycline. Six 15 cm tissue culture dishes were individually seeded with 20×10^6

recombined RPLP HEK293T cells expressing the barcoded binder library. These cells were transfected with 30 µg of dual POI/Standard RBP expression plasmid 24 hours post-seeding using Lipofectamine 3000 according to the manufacturer's instructions. Cells were harvested and reporter proteins immunoprecipitated 48 hours post-transfection.

A 5x concentrated modified RIPA buffer (5xModRIPA; 250 mM Tris, 750 mM NaCl, 20mM $MgCl_2$, and 25% glycerol) was prepared before immunoprecipitation and stored at 4 °C. The following buffers were prepared fresh on the day of immunoprecipitation:

Resuspension buffer: 1xModRIPA, 2 mM PMSF (Millipore Sigma 10837091001), 1x protease inhibitor cocktail (Thermo 78430), 20 units mL^{-1} SUPERaseIn RNase Inhibitor (Thermo AM2696), 300 nM synthetic decoy RNA.

Lysis buffer: 1xModRIPA, 1x protease inhibitor cocktail (Thermo 78430), 20 units mL^{-1} SUPERaseIn RNase Inhibitor (Thermo AM2696), 0.2% Igepal CA-630 (Sigma-Aldrich I8896-50ML).

Wash buffer: 1xModRIPA, 0.1% Igepal CA-630 (Sigma-Aldrich I8896-50ML).

The six 15 cm tissue culture plates were trypsinized and combined into two replicates each consisting of cells from 3 individual plates. Replicate pools were counted and $9x10^7$ cells per replicate were pelleted by centrifugation at 500×g for 5 minutes after which supernatant was removed and the samples transferred to ice. Pellets were resuspended in 4.5 mL of ice cold resuspension buffer and then lysed by addition of 4.5 mL of lysis buffer followed by a 10 minute incubation on ice. Cell lysates were then transferred to 2 mL tubes and insoluble material cleared by centrifugation at 17,000×g for 10 minutes at 4 °C.

For each replicate, 300 µL of Pierce anti-c-Myc magnetic beads (Thermo 88842) and 300 µL of anti-FLAG M2 magnetic beads (Sigma M8823) were prepared by collecting beads on a magnet, removing supernatant, and resuspending with 1.5 mL (i.e. 5x bead volume) of wash buffer. The bead & wash buffer solution was removed from the magnet, inverted to mix, and then washed an additional 3 times with a 5x bead volume of wash buffer. After the final wash, beads were resuspended in 300 µL of wash buffer.

Lysates from each replicate were split between FLAG (3 mL lysate) and Myc (6 mL lysate) immunoprecipitations, and 300 µL of washed beads were added to the corresponding lysate aliquot. Samples were then rotated end-over-end at 4 °C for 3 hours in a cold room, after which beads were collected on a magnet, washed 3 times with 3 mL of wash buffer (i.e. 10x bead volume). After the final wash, beads were collected on a magnet, supernatant was removed, and beads were resuspended in 150 µL of PBS followed by 450 µL of TRIzol LS reagent (Invitrogen 10296010). Resuspended beads were then frozen at -20 °C.

3.5.5.3 RNA isolation and generation of DNA sequencing library

Frozen beads were thawed on ice, beads collected on a magnet, and supernatant was transferred to a clean tube. Beads were washed with an additional 200 µL of TRIzol LS prediluted 1:4 with PBS (e.g. 250 uL of PBS + 750 uL Trizol-LS) and that 200 µL was transferred to the same tube. RNA was isolated from the supernatant using a Direct-zol RNA Miniprep (Zymo R2050) according to the manufacturer's instructions. RNA was eluted in 50 µL of nuclease-free water. 45 µL of eluted RNA was mixed with 2 µL of TURBO DNase (Invitrogen AM2238), 5 µL of 10X Turbo DNase buffer, and 3 µL of nuclease free water and incubated at 37 °C for 30 minutes. RNA was cleaned using a Monarch Spin RNA Cleanup Kit (NEB T2040S) and eluted in 25 uL of nuclease free water.

The entire volume of eluted RNA was then reverse transcribed into cDNA using SuperScript IV Reverse Transcriptase (Invitrogen 18090010). First, 25 µL of cleaned RNA was mixed with 10 uL 1 uM oCCS-1267 reverse transcription primer, 5 uL 10 mM dNTPs, and 25 uL nuclease free water. This mixture was incubated at 65 °C for 5 minutes, and then transferred to ice. Reverse transcriptase reaction mix consisting of 20 µL of 5x SSIV buffer, 5 µL 100 mM DTT, 5 µL nuclease free water, and 5 µL of SuperScript IV Reverse Transcriptase were combined together and then mixed with 65 µL of previously annealed RNA and RT primer. Samples were incubated at 55 °C for 30 minutes, followed by an incubation at 80 °C for 10 minutes and then at 4°C.

To convert the cDNA into an Illumina sequencing library, the entire 100 µL volume of cDNA was spread across eight 50 µL PCR reactions each consisting of 0.25 uL 100 µM oCCS-1266 forward primer, 0.25 µL 100 µM oCCS-1268 reverse primer, 12.5 µL of RT reaction product, 0.25 µL of 100X SYBR Green (ThermoFisher S7567), 25 µL of KAPA HiFi HotStart ReadyMix (Roche KK2602), and 11.75 µL of nuclease free water. Cycling parameters were 3 minutes at 95 °C followed by 15 cycles of 20 seconds at 98 °C, 15 seconds at 75 °C and 20 seconds at 72 °C. Reactions were monitored on a realtime qPCR machine (CFX Opus, Bio-Rad) and terminated while the fluorescence signal was in the log phase. The 8 PCR reactions for each sample were pooled, cleaned using 1.5x AMPure XP beads (Beckman Coulter Life Sciences A63882), and eluted in 240 µL of Buffer EB (Qiagen 19086). Illumina sequencing adapters and indices were added to each sample in a second PCR reaction consisting of 20 µL 1st round PCR product, 100 µL of 2x NEBNext master mix, 10 µL 10 µM P5 indexing primer (for example oCCS-P5-1), 10 µL of 10 µM P7 indexing primer (for example oCCS-P7-1), and 60 µL of nuclease-free water and cycling parameters of 30 seconds at 98 °C followed by 5 cycles of 10 seconds at 98 °C, 30 seconds at 63 °C and 15 seconds at 72 °C.

Reverse transcription of a circular RNA produces concatameric cDNA products which result in laddered PCR amplicons[83]. Indexed PCR reactions were run on a 6%

polyacrylamide gel (Invitrogen EC62652BOX) at 200V for 35 minutes, after which the gel was stained with SYBR Gold (Invitrogen S11494). Bands were visualized on a light box and the smallest band in the laddered product was excised from the gel. DNA was purified from the gel following a previously published protocol[84].

Purified products were quantified on an Agilent 4200 TapeStation, pooled, and sequenced on a Illumina NextSeq 2000 sequencer.

3.5.5.4 Computational analysis of barcodes

Barcode-containing sequencing reads were processed with custom Python scripts to extract barcodes and assign corresponding binder identities. Demultiplexed paired-end FASTQ files were then processed as follows. The barcode sequence was extracted from Read 1 as the nucleotide stretch between predefined flanking sequences, followed by computation of its reverse complement. This barcode was then queried against the lookup table of barcode-binder pairs (see “Computational analysis of subassembly sequencing data” section) to retrieve the corresponding binder identifier. The sample name, read identifier, extracted barcode, extracted barcode reverse complement, binder assignment, and mean Phred quality across the entire read were recorded. The dataset was written to a CSV file containing a single read and its associated information per row.

Downstream analysis of the binder–barcode sequencing output was performed in R. The CSV file containing read-level annotations was imported, and filtered to 1) retain reverse-complement barcodes of the expected length, 2) exclude barcodes lacking an associated binder, and 3) exclude . Remaining reads were collapsed to unique combinations of barcode, binder ID, pulldown (i.e. FLAG or Myc), and replicate with an associated read count.

To compare results between replicates, per-sample read proportions were calculated from summed read counts of each barcode-binder pair in each pulldown in each replicate. Barcodes were retained only if they were detected at a non-negligible proportion in both replicates (per-barcode proportion $\geq 10^{-7}$ in both replicate 1 and replicate 2). Only binders with ≥ 3 barcodes in each replicate were retained and a small pseudocount (10^{-6}) was added to each read proportion. Within each replicate, differences in the FLAG and Myc read proportions were tested by a paired Wilcoxon signed-rank test utilizing all barcodes belonging to the same binder. P values were corrected for multiple-hypothesis testing using the Benjamini–Hochberg method.

We then pooled all read proportions and recomputed significance in order to increase our statistical power. Similar to the between-replicate test, barcodes with a frequency $\geq 10^{-7}$ in both replicates were retained and binders with ≥ 3 barcodes in each replicate were retained and a small pseudocount (10^{-6}) was added to each read proportion. Differences in the FLAG and Myc read proportions were again tested by a

paired Wilcoxon signed-rank test utilizing all barcodes belonging to the same binder and P values corrected for multiple-hypothesis testing using the Benjamini–Hochberg method.

3.5.6 Singleton validation of screen effects with a fluorescent protein abundance reporter

A stable, polyclonal HEK293T cell line harboring an EF1a-driven reverse tetracycline transactivator (rtTA) and a doxycycline-inducible TRE-3xFLAG-GFP-GCK-ALFA-IRES-mCherry plasmid was generated by piggyBac transposition. A negative control line was harboring an EF1a-driven rtTA and a reporter in which the ALFA tag was deleted (i.e. 3xFLAG-GFP-GCK-IRES-mCherry) was generated in parallel. 2×10^5 of these cells were seeded in wells of a 24 well plate and transfected 24 hours later with 500 ng of a plasmid encoding a NbALFA-binder fusion and mTagBFP2 driven by CMV and EF1a promoters respectively. Reporter expression was induced 24 hours post transfection via addition of 100 ng/mL doxycycline to the culture media. 72 hours post-transfection (48 hours post-induction) cells were trypsinized, pelleted at 500×g for 5 minutes, resuspended in FACS buffer (PBS supplemented with 10% FBS), and analyzed on a Attune Flow Cytometer (Thermo). These results are presented in Figure 3A-3B.

Binder-induced changes in the GFP:mCherry ratio was also measured in HepG2 cells with several changes. Here, HepG2 cells were seeded (1.1×10^5 cells per well of a 24 well plate) and co-transfected with 500 ng each of constitutive binder and reporter plasmids (i.e. CMV-NbALFA-binder-EF1a-mTagBFP and CMV-3xFLAG-GFP-GCK-IRES-mCherry, respectively).

3.5.6 Analysis of fluorescent protein abundance reporter flow cytometry data

Flow cytometry experiments were analyzed using FlowJo. Cells were gated from all events (forward scatter area versus side scatter area) followed by gating on single cells (forward scatter area versus forward scatter height). From the population of single cells, mCherry positive cells were gated, followed by mTagBFP2 positive cells. A GFP:mCherry ratio parameter was derived and applied to this population.

3.5.7 Assessment of endogenous MCL1 protein degradation by Western blotting

To prepare lentiviral particles encoding MCL1-degrader binder fusions, 1.2×10^6 HEK293T cells were seeded in individual wells of a 6 well plate one day before transfection with pMDLg/pRRE (Addgene #12251), pRSV-Rev (Addgene #12253), pMD2G (Addgene #12259) and 750 ng of binder-encoding lentiviral vector. Transfection media was changed 24 hours post transfection. 48 hours post-transfection, virus containing supernatant was collected, centrifuged at 500×g for 5 minutes, and filtered through a 0.45 µm filter.

3×10^5 MDA-MB-231 cells were seeded in 2 mL of complete DMEM in individual wells of a 6 well plate 24 hours before transduction. To transduce cells, 1 mL of DMEM was removed from each well and replaced with 1 mL of freshly prepared and filtered lentiviral supernatant. Media were replaced with 2 mL of complete DMEM 24 hours post-transduction. 72 hours post-transduction, cells were washed 3 times with DPBS and lysed with RIPA Lysis Buffer supplemented with protease inhibitor, phosphatase inhibitor, and benzonase on ice for 30 minutes. Cells were scraped and spun down at 21,000×g for 15 minutes at 4 °C. The supernatant was collected and the concentrations were normalized through a BCA assay. 45 µg of lysates were loaded onto 4-12% Bis-Tris Gel and separated by SDS-PAGE. The gel was then transferred to a nitrocellulose membrane, and blocked with Intercept Blocking Buffer (LICOR) for 1 hour at room temperature. The blots were stained with primary antibody overnight at 4 °C, washed 3 times with TBS-T, and stained with secondary antibody for 1 hour at room temperature. A LICOR Odyssey CLx Imager was used to image and quantify the blot.

3.5.8 Apoptosis Assay

1.1×10^5 MDA-MB-231 cells were seeded in 0.3 mL of complete DMEM in individual wells of a 48 well plate 24 hours before transduction. The following day, 150 µL of DMEM was removed from each well and replaced with 150 µL of freshly prepared lentiviral supernatant. Media was replaced with 300 µL of complete DMEM 24 hours post-transduction. 72 hours post-transduction, cells were trypsinized, quenched with media, and moved to a 96-well-U-bottom plate. Cells were spun down and washed once with DPBS. Each well was incubated with 50 µl Annexin Buffer (10 mM HEPES, 140 mM NaCl, and 2.5 mM CaCl2, pH 7.4) and 2.5 µl of Annexin V AF647 Conjugate (Thermo Fischer A23204) for 15 minutes at rt. Afterwards, 150 µl of Annexin buffer with Sytox Green was added, and cells were analyzed by Attune Flow Cytometer.

3.5.9 Confocal microscopy

3×10^4 HEK293T cells were seeded in 96-well glass bottom plate (Cellvis #P96-1.5H-N) and transfected 24 hours later with 25 ng of ALFA-GFP-FIS1-IRES-mCherry and 75 ng of NbALFA-degrader-EF1a-mTagBFP plasmids. Media was replaced 24 hours post-transfection with complete DMEM supplemented with 125 nM MitoTracker Deep Red FM dye (ThermoFisher Scientific #M22426) and cells were incubated in a tissue culture incubator at 37°C. After staining for 30 minutes, media was replaced with complete DMEM and confocal imaging was performed using a spinning-disk confocal microscope. As described previously [85], a Yokogawa CSU-W1 SoRa spinning disc confocal device is attached to a Nikon ECLIPSE Ti2 microscope. Excitation light was emitted from lasers housed inside of a Nikon LUNF 405/488/561/640NM 1F commercial launch. Emission light was directed by a quadband dichroic mirror (Semrock, Di01-T405/488/568/647-13x15x0.5) and filtered by one of four single-bandpass filters (DAPI, Chroma, ET455/50M; ATTO 488, Chroma,

ET525/36M; ATTO 565, Chroma, ET605/50M; Alexa Fluor 647, Chroma, ET705/72M) and focused onto an Andor Sona 4.2B-11 camera. Specific settings for this work were as follows: Excitation light was emitted for 405 nm, 488 nm, 561 nm, or 640 nm lasers at 30%, 35%, 30%, and 35% of maximal intensity, respectively. The exposure times of 600, 200, 200, and 200 ms were used for 405 nm, 488 nm, 561 nm, and 640 nm, respectively. A 100x N.A. 1.49 Apo oil immersion objective lens and a 2.8x lens in the SoRA unit were used for super-resolution imaging, resulting in an effective pixel size of 39.3 nm. Z-stacks were captured for all channels at a 0.3 µm step size for 51 steps using the “Z then channel” sequence. Nikon Elements AR 5.20.00 was used to acquire microscopy images on the Nikon Ti2 system. Images were processed in ImageJ (Fiji).

3.6 References for Chapter 3

1. Békés, M., Langley, D. R. & Crews, C. M. PROTAC targeted protein degraders: the past is prologue. Nat. Rev. Drug Discov. 21, 181–200 (2022).
2. Simon, J. J., Fowler, D. M. & Maly, D. J. Multiplexed profiling of intracellular protein abundance, activity, interactions and druggability with LABEL-seq. Nat. Methods 21, 2094–2106 (2024).
3. Glickman, M. H. & Ciechanover, A. The Ubiquitin-Proteasome Proteolytic Pathway: Destruction for the Sake of Construction. Physiol. Rev. 82, 373–428 (2002).
4. Glick, D., Barth, S. & Macleod, K. F. Autophagy: cellular and molecular mechanisms. J. Pathol. 221, 3–12 (2010).
5. Berger, S. et al. Computationally designed high specificity inhibitors delineate the roles of BCL2 family proteins in cancer. Elife 5, (2016).
6. Wang, H., Guo, M., Wei, H. & Chen, Y. Targeting MCL-1 in cancer: current status and perspectives. J. Hematol. Oncol. 14, 67 (2021).
7. Lorentzen, K. C., Prescott, A. R. & Ganley, I. G. Artificial targeting of autophagy components to mitochondria reveals both conventional and unconventional mitophagy pathways. Autophagy 21, 315–337 (2025).
8. Jiang, Z., Kuo, Y.-H. & Arkin, M. R. Autophagy receptor-inspired antibody-fusion proteins for targeted intracellular degradation. J. Am. Chem. Soc. 145, 23939–23947 (2023).
9. Pacesa, M., Pelea, O. & Jinek, M. Past, present, and future of CRISPR genome editing technologies. Cell 187, 1076–1100 (2024).
10. Stanton, B. Z., Chory, E. J. & Crabtree, G. R. Chemically induced proximity in biology and medicine. Science 359, (2018).
11. Zhao, L., Zhao, J., Zhong, K., Tong, A. & Jia, D. Targeted protein degradation: mechanisms, strategies and application. Signal Transduct. Target. Ther. 7, 113 (2022).
12. Liu, Y. et al. Expanding PROTACtable genome universe of E3 ligases. Nat. Commun. 14, 6509 (2023).
13. Kannt, A. & Đikić, I. Expanding the arsenal of E3 ubiquitin ligases for proximity-induced protein degradation. Cell Chem. Biol. 28, 1014–1031 (2021).
14. Ishida, T. & Ciulli, A. E3 ligase ligands for PROTACs: How they were found and how to discover new ones. SLAS Discov. 26, 484–502 (2021).
15. Watson, J. L. et al. De novo design of protein structure and function with RFdiffusion. Nature 620, 1089–1100 (2023).
16. Dauparas, J. et al. Robust deep learning-based protein sequence design using ProteinMPNN. Science 378, 49–56 (2022).
17. Boswell, C. W. et al. Genetically encoded affinity reagents are a toolkit for visualizing and manipulating endogenous protein function in vivo. Nat. Commun. 16, 5503 (2025).
18. Caussinus, E., Kanca, O. & Affolter, M. Fluorescent fusion protein knockout mediated by anti-GFP nanobody. Nat. Struct. Mol. Biol. 19, 117–121 (2011).
19. Poirson, J. et al. Proteome-scale discovery of protein degradation and stabilization effectors. Nature 628, 878–886 (2024).

20. Clift, D. et al. A method for the acute and rapid degradation of endogenous proteins. Cell 171, 1692–1706.e18 (2017).
21. Matreyek, K. A. et al. Multiplex assessment of protein variant abundance by massively parallel sequencing. Nat. Genet. 50, 874–882 (2018).
22. Kinney, J. B., Murugan, A., Callan, C. G., Jr & Cox, E. C. Using deep sequencing to characterize the biophysical mechanism of a transcriptional regulatory sequence. Proc. Natl. Acad. Sci. U. S. A. 107, 9158–9163 (2010).
23. Bertrand, E. et al. Localization of ASH1 mRNA particles in living yeast. Mol. Cell 2, 437–445 (1998).
24. Götzke, H. et al. The ALFA-tag is a highly versatile tool for nanobody-based bioscience applications. Nat. Commun. 10, 4403 (2019).
25. Hiatt, J. B., Patwardhan, R. P., Turner, E. H., Lee, C. & Shendure, J. Parallel, tag-directed assembly of locally derived short sequence reads. Nat. Methods 7, 119–122 (2010).
26. Kircher, M. et al. Saturation mutagenesis of twenty disease-associated regulatory elements at single base-pair resolution. Nat. Commun. 10, 1–15 (2019).
27. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589 (2021).
28. Tantawy, S. I. et al. Mechanisms of MCL-1 protein stability induced by MCL-1 antagonists in B-cell malignancies. Clin. Cancer Res. 29, 446–457 (2023).
29. Tantawy, S. I., Timofeeva, N., Hernandez, A., Sarkar, A. & Gandhi, V. Decoding the mechanism behind MCL-1 inhibitors: A pathway to understanding MCL-1 protein stability. Oncotarget 14, 653–655 (2023).
30. Papatzimas, J. W. et al. From inhibition to degradation: Targeting the antiapoptotic protein myeloid cell leukemia 1 (MCL1). J. Med. Chem. 62, 5522–5540 (2019).
31. Kotschy, A. et al. The MCL1 inhibitor S63845 is tolerable and effective in diverse cancer models. Nature 538, 477–482 (2016).
32. Wang, Z. et al. Proteolysis targeting chimeras for the selective degradation of Mcl-1/Bcl-2 derived from nonselective target binding ligands. J. Med. Chem. 62, 8152–8163 (2019).
33. Bennett, N. R. et al. Improving de novo protein binder design with deep learning. Nat. Commun. 14, 2625 (2023).
34. Lalanne, J.-B. et al. Multiplex profiling of developmental cis-regulatory elements with quantitative single-cell expression reporters. Nat. Methods 1–11 (2024) doi:10.1038/s41592-024-02260-3.
35. Johnson, G. E. & Li, G.-W. Genome-wide quantitation of protein synthesis rates in bacteria. Methods Enzymol. 612, 225–249 (2018).
36. Attar, S. et al. Efficient and highly amplified imaging of nucleic acid targets in cellular and histopathological samples with pSABER. Nat. Methods 22, 156–165 (2025).

Chapter 4
DISCUSSION AND FUTURE DIRECTIONS

4.1 Summary of Main Findings

In this thesis, I have developed and applied multiplex experimental methods to profile and program intracellular protein degradation. Central to this work is the convergence of high-throughput experimental genomics with deep learning-based protein modeling to systematically interrogate and manipulate intracellular protein stability.

In Chapter 2, I presented COMET, a framework for mapping the endogenous wiring of the ubiquitin-proteasome system. By coupling combinatorial genetics with flow cytometry and deep sequencing, I moved beyond the traditional "one-by-one" analysis of E3-substrate interactions. This work revealed that the proteolytic landscape is highly complex with most E3-substrate mappings occurring as many-to-many networks rather than simple one-to-one relationships. Further, I demonstrated that deep-learning-based structural prediction (AlphaFold-Multimer) can serve as a scalable, orthogonal validation tool for experimentally nominated interactions.

In Chapter 3, I shifted my focus from mapping endogenous protein degradation networks to programming synthetic ones. Here, I integrated computational protein design with a multiplex protein abundance assay (LABEL-seq) to characterize thousands of de novo designed "proximity handles." This approach succeeded in identifying binder-effector pairs that could induce degradation or stabilization of target proteins, including endogenous oncogenes like MCL1. Furthermore, I showed that binders designed to target autophagy pathway proteins were capable of inducing mitochondrial clustering, demonstrating control of phenotypes beyond simple control of protein abundance.

While Chapter 2 and Chapter 3 address related but distinct biological questions, they are united by a shared methodological philosophy. Both chapters tackle the critical bottleneck of functional characterization by using individual cells as containers within which to conduct many parallel protein abundance experiments.

4.2 Toward a "Phone Book" of Protein Editors

A broader goal suggested by this work is the creation of a comprehensive "phone book" for proteome editing. A universal library of parts that allows any target protein to be coupled to any cellular effector would have clear use cases as basic biology tools and potential therapeutic applications.

The work in Chapter 3 focused on a modest set of seven effectors. A natural extension is to scale this approach across many different protein degradation effectors. In addition, it would be interesting to characterize the functional effect that recruitment of other effectors, such as kinases or phosphatases, may have on substrate abundance. By systematically mapping how different handles behave across diverse pathways, we can better understand the basic biology of these effectors as well as continue to build the “phone book” of protein editing.

Simultaneously, we must address the "target" side of the equation. By fixing a validated effector handle and varying the target binder, large-scale screens could identify binders for virtually every human protein. A proteome wide parts list capable of addressing any arbitrary effector-target pair would empower the modular construction of bifunctional molecules for precise proteome manipulation.

4.3 Inverting the Discovery Paradigm: The Computation-Experiment Loop

Both COMET (Chapter 2) and the de novo design screen (Chapter 3) highlight a fundamental shift in biological discovery. Historically, screens were used to discover hits which were then rationalized *post hoc*. The rapid emergence of deep learning models of biology allows inversion of this framework such that we leverage models to predict or design elements of interest which are then subsequently validated in a high-throughput experiment.

In Chapter 2, I demonstrated that AlphaFold-Multimer could distinguish true E3-substrate pairs from random background, suggesting that future efforts could computationally nominate a predicted E3-substrate interactome. These high-confidence candidate pairs could then be plugged into a COMET experiment and validated at scale. In Chapter 3, I more directly showed this framework by demonstrating that *de novo* designed effector binders could be

Demonstration of either model improvement or training of entirely new models is a key next step. The current implementation is imperfect; degradation readouts are a proxy for binding, and a design may fail to degrade a target for reasons unrelated to binding affinity (e.g., recruitment geometry or lysine availability). To train the next generation of predictive models, we need assays that measure the biochemical interaction directly at scale. Developing variants of the framework described in Chapter 3 that are instead capable of measuring protein-protein interactions (rather than abundance) would provide the high-quality, quantitative training data necessary to transform protein design from a hit-or-miss endeavor into a precise engineering discipline.

4.4 Beyond Abundance: Spatial and Functional Logic

While this thesis largely focused on modulating protein stability, the "proximity handles" developed in Chapter 3 hint at broader capabilities. The observation that

recruiting autophagy receptors to the mitochondrial membrane resulted in organelle clustering suggests that proteome editors can act as spatial organizers.

Future work should explore "edit types" beyond degradation. Combining proteome editing with established transcriptional and post-transcriptional control systems could reasonably enable construction of synthetic circuits that implement logic, feedback, and memory directly at the protein level. For example, one could design "conditional" handles that are only active in the presence of a specific cellular signal or drug, or "logic gates" where degradation of a target occurs only if two distinct effectors are present. Such circuits would allow us to program cellular behavior with the speed and reversibility of post-translational modifications, rather than the slower kinetics of transcriptional regulation.

4.5 The Next Frontier: Multiplex Enzyme Design

Finally, the framework for multiplex characterization of computationally designed proteins is well-suited to tackle problems of greater complexity, such as enzyme design. Unlike protein binders, which require only surface complementarity, enzymes require precise positioning of catalytic residues and transition state stabilization. Consequently, they suffer from a lack of high-quality training data linking sequence to catalytic turnover.

A high-impact extension of this work would be the multiplex testing of designed enzymes. As one example, *de novo* proteases capable of cleaving specific peptide target sequences would allow destruction of a specific protein target in a manner not dependent upon the UPS or autophagy pathway. Cellular assays where the ratio cleaved to uncleaved protein molecules are measured using a high-throughput DNA sequencing readout would allow characterization of *de novo* protease activity across thousands of designs against many different target sequences in a single experiment. Recovering functional designs from such a screen would provide a rich dataset to train structure-based models on the rules of catalysis. If successful, this demonstration would guide the field from a low-throughput, bespoke practice into a scalable, data-driven discipline.

4.6 Concluding Remarks

The transition from biological description to biological engineering is one of the defining shifts of our time. This thesis contributes to that shift by establishing scalable technologies to read and write the code of protein degradation.Treating the cell’s degradation machinery as a modular, programmable system opens new avenues for basic biology and for medicine. Ultimately, if we can routinely and predictably sculpt the proteome in living cells, we gain control over the primary layer where biological function is executed.