














Multi-scale classification decodes the complexity of the human E3 ligome

Received: 11 March 2025

Accepted: 1 December 2025

Published online: 25 December 2025

 Check for updates

Arghya Dutta ^{1,2,3,5}, Alberto Cristiani ^{1,2,5}, Siddhanta V. Nikte ^{1,2,5}, Jonathan Eisert ^{1,2}, Yves Matthes ¹, Borna Markusic ^{1,4}, Cosmin Tudose ¹, Chiara Becht ¹, Varun Jayeshkumar Shah¹, Thorsten Mosler ¹, Koraljka Husnjak¹, Ivan Dikic ^{1,2,4}, Manuel Kaulich ¹ & Ramachandra M. Bhaskara ^{1,2,4} 

E3 ubiquitin ligases are vital enzymes that define the ubiquitin code in cells. Beyond promoting protein degradation to maintain cellular health, they also mediate non-degradative processes like DNA repair, signaling, and immunity. Despite their therapeutic potential, a comprehensive framework for understanding the relationships among diverse E3 ligases is lacking. Here, we classify the “human E3 ligome”—an extensive set of catalytic human E3s—by integrating multi-layered data, including protein sequences, domain architectures, 3D structures, functions, and expression patterns. Our classification is based on a metric-learning paradigm and uses a weakly supervised hierarchical framework to capture authentic relationships across E3 families and subfamilies. It extends the categorization of E3s into RING, HECT, and RBR classes, including non-canonical mechanisms, successfully explains their functional segregation, distinguishes between multi-subunit complexes and standalone enzymes, and maps E3s to substrates and potential drug interactions. Our analysis provides a global view of E3 biology, opening strategies for drugging E3-substrate networks, including drug repurposing and designing specific E3 handles.

Cells constantly modulate their proteomes in response to physiological and environmental changes. The timely removal and turnover of cellular proteins is integral to protein homeostasis¹. In eukaryotes, individual proteins, complexes, and large assemblies are degraded via either autophagy or the ubiquitin-proteasome system (UPS)². In mammalian cells, ~80% of the cellular proteome is degraded through the UPS¹. In this pathway, the designated protein cargo is tagged with ubiquitin (Ub) molecules through a series of enzymatic reactions, marking them for degradation by the proteasome³. Following the action of E1 and E2 enzymes, the E3 ligase brings both the E2–ubiquitin complex and the substrate protein in proximity, allowing the transfer of Ub from the E2 enzyme to a lysine residue on the target protein^{4,5}. This process is often repeated (poly-ubiquitination), resulting in

substrates with distinct types of Ub-chains. In UPS, for instance, K48-linked Ub-chains are recognized by Ub-binding domains (UBDs) on 19S proteasomal particles, initiating the degradation of substrates¹. In autophagy, ubiquitination often serves as a necessary condition for identifying substrates, conferring specificity⁶. Cargo components, damaged organelles, and intracellular pathogens targeted for degradation are often ubiquitinated. Further, autophagy receptors are enriched in UBDs to recognize modified cargo components⁷ or themselves strongly ubiquitinated to trigger aggregation of protein assemblies in the cytosol and organellar membranes^{8,9}, thus enhancing autophagic flux.

E3 ubiquitin ligases confer substrate specificity for ubiquitination. They recognize distinct targets, operate in diverse cellular locations,

¹Institute of Biochemistry II, Faculty of Medicine, Goethe University, Theodor-Stern-Kai 7, Frankfurt am Main, Germany. ²Buchmann Institute for Molecular Life Sciences, Goethe University, Max-von-Laue Str. 15, Frankfurt am Main, Germany. ³Department of Physics, SRM University–AP, Amaravati, Andhra Pradesh, India. ⁴IMPRS on Cellular Biophysics, Max-von-Laue Str. 3, Frankfurt am Main, Germany. ⁵These authors contributed equally: Arghya Dutta, Alberto Cristiani, Siddhanta V. Nikte. ✉e-mail: Bhaskara@med.uni-frankfurt.de

and exert spatial control of protein turnover^{10,11}. In addition to controlling homeostatic processes, E3 ligases regulate immunity and inflammation pathways^{12,13}. Given their tissue-specific expressions and association with developmental and metabolic syndromes, including cancer progression, E3 ligases have emerged as promising candidates, particularly for drugging previously undruggable targets¹⁴. In stark contrast to E1 (~10) and E2 enzymes (~50), a substantial number of E3 ligases (~600) have been recognized in humans^{15,16}. This count of putative E3s stems from various investigations: Li et al.¹⁷ identified ~617 potential human E3-encoding genes by conducting a genome-wide search to detect RING (Really Interesting New Gene) finger catalytic domains using hidden Markov models. Subsequently, Deshaies and Joazeiro¹⁸ characterized ~300 RING and U-box E3 ligases, while Medvar et al.¹⁹ documented ~377 E3 ligases, with a primary focus on confirmed catalytic activity. Despite these efforts, many human E3 ligases have been only partially characterized. A significant fraction remains unexplored and hypothetical or unknown²⁰. To date, those studied exhibit extensive heterogeneity in their sequence, domain composition, 3D structure, subcellular localization, and tissue expression, establishing them as one of the most diverse classes of enzymes. Furthermore, several E3 ligases function as multi-subunit complexes with varied substrate specificities modulated by specific receptors, adaptors, and scaffold proteins²¹. The extensive variety and large numbers of E3 ubiquitin ligases create a bottleneck for pattern recognition and large-scale study. Therefore, detailed characterization and analysis of the human E3 ligome—the complete set of E3 ubiquitin ligases encoded by the human genome—is essential for a comprehensive understanding.

The current classification of the E3 ligases—based on the ubiquitin-transfer mechanism—categorizes them into three main classes: RING (Really Interesting New Gene), HECT (Homologous to the E6AP Carboxyl Terminus), and RBR (RING-Between-RING) classes¹⁵. This classification drastically oversimplifies the mechanistic diversity of E3 ligases, compels the grouping of enzymes with hybrid characteristics, and fails to accommodate emerging information on recent and atypical ligases, limiting its overall utility¹⁸. A multi-scale classification of the human E3 ligome offers a unique solution to tackle the complexity and remarkable diversity inherent in these enzymes at various scales. This organized approach can provide more accurate and functional groupings, crucial for a nuanced understanding of different E3 ligase families. Further, emerging patterns detected help trace evolutionary relationships more effectively, revealing conserved elements and adaptive changes that are not evident. Furthermore, mapping essential information such as functional diversity, substrate-specificities, and druggability onto the classification provides a global view, guiding specific and directed investigations to fill in the missing information.

Here, we systematically catalog all E3 ubiquitin ligases to build a comprehensive and manually curated human E3 ligome. We then encode the relationships between high-confidence E3 ligases using multiple distance measures at various granular layers spanning the molecular- and the systems-level organization. By amalgamating selected distance measures from multiple layers into an optimized emergent distance metric, we group all human E3 ligases into distinct families and subfamilies. Our classification delineates features and patterns specific to E3 ligase families, providing insights into their organization. By combining CRISPR-Cas9 dropout screens and proteomic analysis with functional enrichment analysis, we identify essential E3s differentially regulated under stress conditions. We demonstrate the utility of this unbiased classification by mapping the existing state of knowledge on E3 ligase domain architecture, 3D structure, function, substrate networks, and small molecule interactions to gain generic and family-specific insights. The multiscale classification framework developed here embodies canonical and atypical E3 mechanisms largely reflecting the ubiquitin code, offering a

comprehensive roadmap to navigate the vast landscape of E3 ligase biology, laying the groundwork for future therapeutic applications.

Results

Assembly of the human E3 ligome

To comprehensively identify all E3 ligases in the human genome, we conducted a census using datasets from previously published studies and public repositories. By visualizing their overlaps, we found that all existing datasets were largely inconsistent (Fig. 1a and Supplementary Fig. 1a). Most strikingly, only 99 proteins were consistently categorized as human E3 ligases from all eight datasets. The low overlap in these datasets reflects the diverse approaches and often variable definitions used to collate E3 systems (Supplementary Table 1). We resolved these conflicts by explicitly defining the catalytic components of E3 systems, i.e., polypeptide sequences containing one or more catalytic domains ($C = \{d_i\}$, see “Methods”). Using this objective criterion ($\{X_i \in \bigcup_{n=1}^8 \{d_i \in C\}\}$; Supplementary Table 2) facilitated proper annotation and targeted analysis of E3s. We found that 462 polypeptide sequences, across all datasets ($\bigcup_{n=1}^8 A_n = 1448$), contain at least one catalytic domain constituting the curated E3 ligome (Fig. 1b and Supplementary Fig. 1b).

To substantiate our curation process, we defined a consensus score for each protein based on its presence in various source datasets (Fig. 1c). We found that the HECT and RBR classes of E3 ligases showed high agreement across datasets (confidence score ≥ 0.6 ; orange and purple bars). The RING class (green bars) had a broad distribution of consensus scores indicative of annotation challenges. However, the most significant discrepancy among the datasets (confidence score ≤ 0.25) was due to misannotated proteins. E1, E2, and other non-catalytic components of E3 systems, such as receptors, scaffolds, and adaptor proteins, were often merged with E3 ligases, accounting for an additional 298 proteins, leaving 514 unclassified and 174 with no domain annotations (Fig. 1b). Furthermore, several proteins obtained from UniProt and BioGRID using keyword-based searches (Supplementary Fig. 1c) were not included by others, have low consensus scores, and remain unclassified and unannotated, excluding 688 proteins from the curated E3 ligome (Fig. 1c, black bars). Our approach thus minimized false positives and true negatives, includes high-confidence catalytically active E3s and accounts for pseudo-E3 ligases and other E3s with untested catalytic activities, providing a detailed assessment of the completeness of the human E3 ligome (see Supplementary Note 1).

To quantify the diversity of the human E3 ligome, we mapped the sequence, structure, and functional features of individual E3s corresponding to well-known E3 classes (RING, HECT, and RBR). We found that the length distribution of the E3s is broad, ranging from 100 to 5000 residues (mean size = 635 residues; Fig. 1d). The average fractional coverage of E3s annotated with unique domains is 37%, 42%, and 53% for RING, HECT, and RBR classes, respectively (Fig. 1e). Furthermore, on average, the RING, HECT, and RBR domains span 23%, 31%, and 39% of their total lengths, respectively (Fig. 1f). By mapping information from the Protein Data Bank (PDB), we found 2259 distinct structures representing RING, HECT, and RBR-containing proteins (2001+168+90), providing partial structural information for 51% (208+21+8 unique UniProt records) of the E3 ligome (Fig. 1g). Analysis of AlphaFold models revealed that for most E3s, the coverage of structured domains is high, and the amount of intrinsic disorder is generally low (pLDDT ≤ 50 covering only $\leq 10\%$ E3 length; Supplementary Fig. 1d). We quantified the functional diversity of the E3 ligome by retrieving the unique Gene Ontology (GO) annotations corresponding to Biological Processes (BP), Cellular Component (CC), and Molecular Function (MF). We annotated 96–100% of the E3s with unique GO terms (Fig. 1h). The number of distinct GO terms captured the diversity of functional assignments attributed to the three E3 classes.

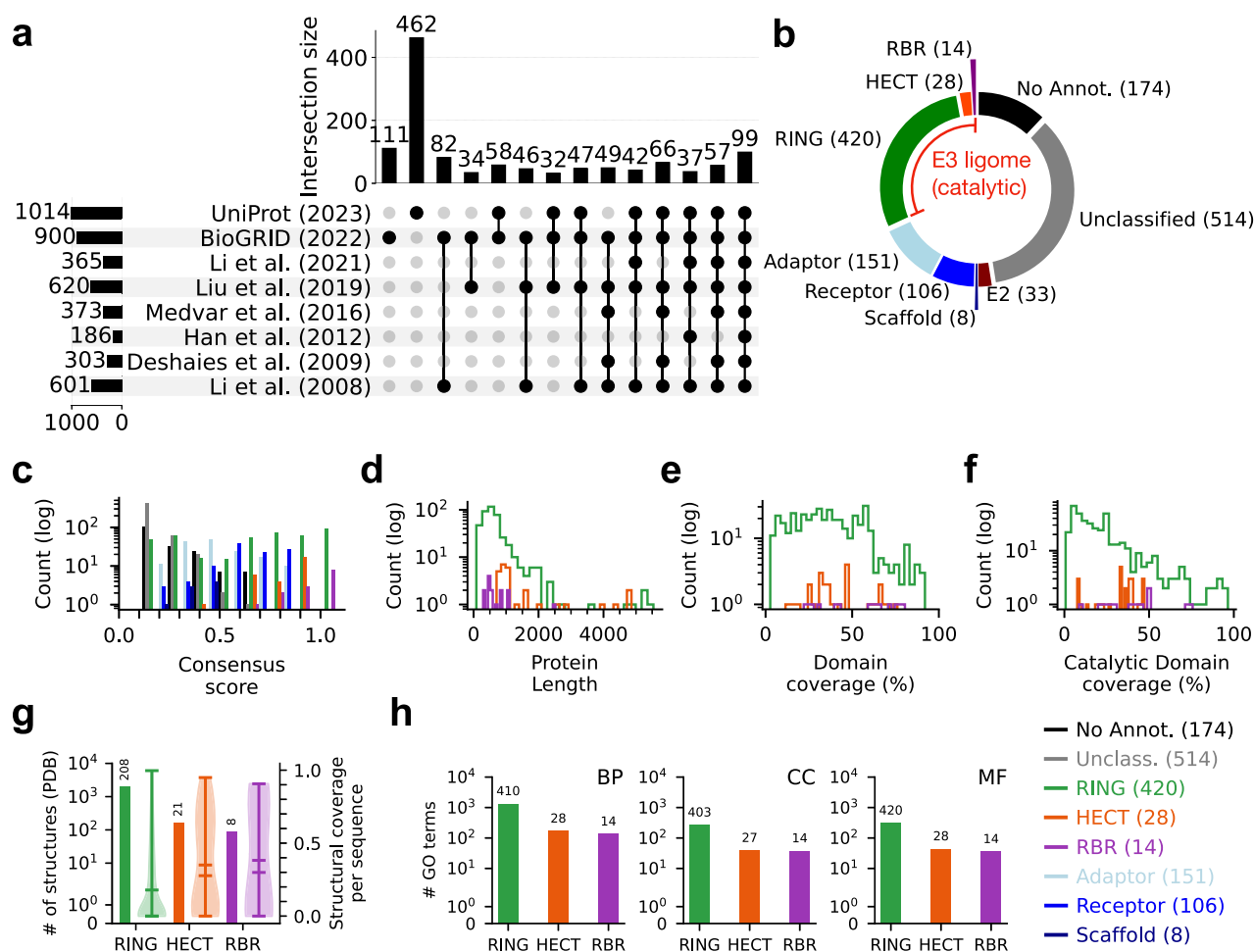


Fig. 1 | Diversity of the human E3 ligome. **a** A visualization showing the intersections of eight E3 ligases datasets (A_1, \dots, A_8) obtained from existing literature and public repositories. The matrix layout for all intersections of individual datasets is sorted by size. Filled circles and their corresponding bars indicate sets that are part of the intersection and their sizes, respectively. Individual proteins (X_i) from the all eight datasets $\bigcup_{n=1}^8 A_n = 1448$ annotated with one or more domains, d_i , belonging to a set of well-studied catalytic components of E3 enzymes ($C = \{d_c\}$) were compiled to form the high-confidence E3 ligome, $\{X_i \in \bigcup_{n=1}^8 A_n | \exists d_i \in C\}$. **b** Pie chart showing the extent of protein annotations and filtering to identify the catalytic and non-catalytic components of the human E3 ligome. **c** Distribution of consensus scores for all annotated protein classes reflects cross-dataset

reproducibility on E3 ligase catalytic components. The distribution of (d) protein lengths and annotation coverage for (e) all domains and (f) catalytic domains highlights the heterogeneity of the E3 ligome. **g** Distribution of structural coverage of the E3 ligome at class-level. Barplots (left axis) display the number of available PDB structures for $n = 208$ RING, $n = 21$ HECT, and $n = 8$ RBR proteins. Violin plots (right axis, min, max, median, and mean values with mirrored density estimates on either side) represent distributions of fractional coverage for $n = 2001$ RING, $n = 168$ HECT, and $n = 90$ RBR structures. **h** The total number of unique GO terms associated with E3 classes indicates their functional vista under biological process BP, cellular component CC, and molecular function MF ontologies. n -values on bars indicate unique proteins with GO terms.

Metric learning for classification of the human E3 ligome

To study the organization and relationships of proteins within the human E3 ligome, we attempted to classify these enzymes using multiple sequence alignment (MSA) followed by phylogenetic tree construction. However, we obtained a low-quality MSA with numerous gaps (Supplementary Fig. 2a), primarily due to (i) high sequence divergence, (ii) numerous proteins with uneven length distributions, (iii) inadequate alignment of conserved, catalytic domains, and (iv) an extensive repertoire of domain architectures (Supplementary Fig. 2b).

To capture the complex relationships within the human E3 ligome, we used a machine-learning approach to learn an emergent distance measure. Using a linear sum model, we combined multiple distance measures with optimal weights to reproduce class-level organization (partial ground truth) in hierarchical clustering (Fig. 2a). We first computed twelve pairwise distance matrices for all E3 ligase pairs (d_{PQ}^i where $i = \{1, \dots, 12\}$, for all P and Q \in E3 ligome; $12 \times 462 C_2$ distances) across distinct granular layers: primary sequence, domain

architecture, 3D structure, function, subcellular localization and expressions (see “Methods”). These distances between ligase pairs are widely distributed and capture their relationships across distinct molecular- and systems-level hierarchies (Fig. 2b). Interestingly, most distance measurements showed low correlations (Fig. 2c), suggesting that they capture largely orthogonal information from the distinct granularity layers. Only the three domain architecture-based distances which quantify domain composition (d_{PQ}^{lac}), domain order (d_{PQ}^{Gky}), and domain duplication (d_{PQ}^{Dup}) are highly correlated (Pearson $r \geq 0.5$). Further, the 3D structure-based distance measure (d_{PQ}^{Str}) is also positively correlated with domain composition and duplication distances (Pearson $r \geq 0.5$).

Next, to learn an emergent distance measure, D_{PQ} , we combined four individual distances (d_{PQ}^i), representative of E3 sequence, domain composition, structural, and functional level organization, with their appropriate weights ($w_i \in \{0.05, \dots, 0.95\}$ in 0.1 intervals).

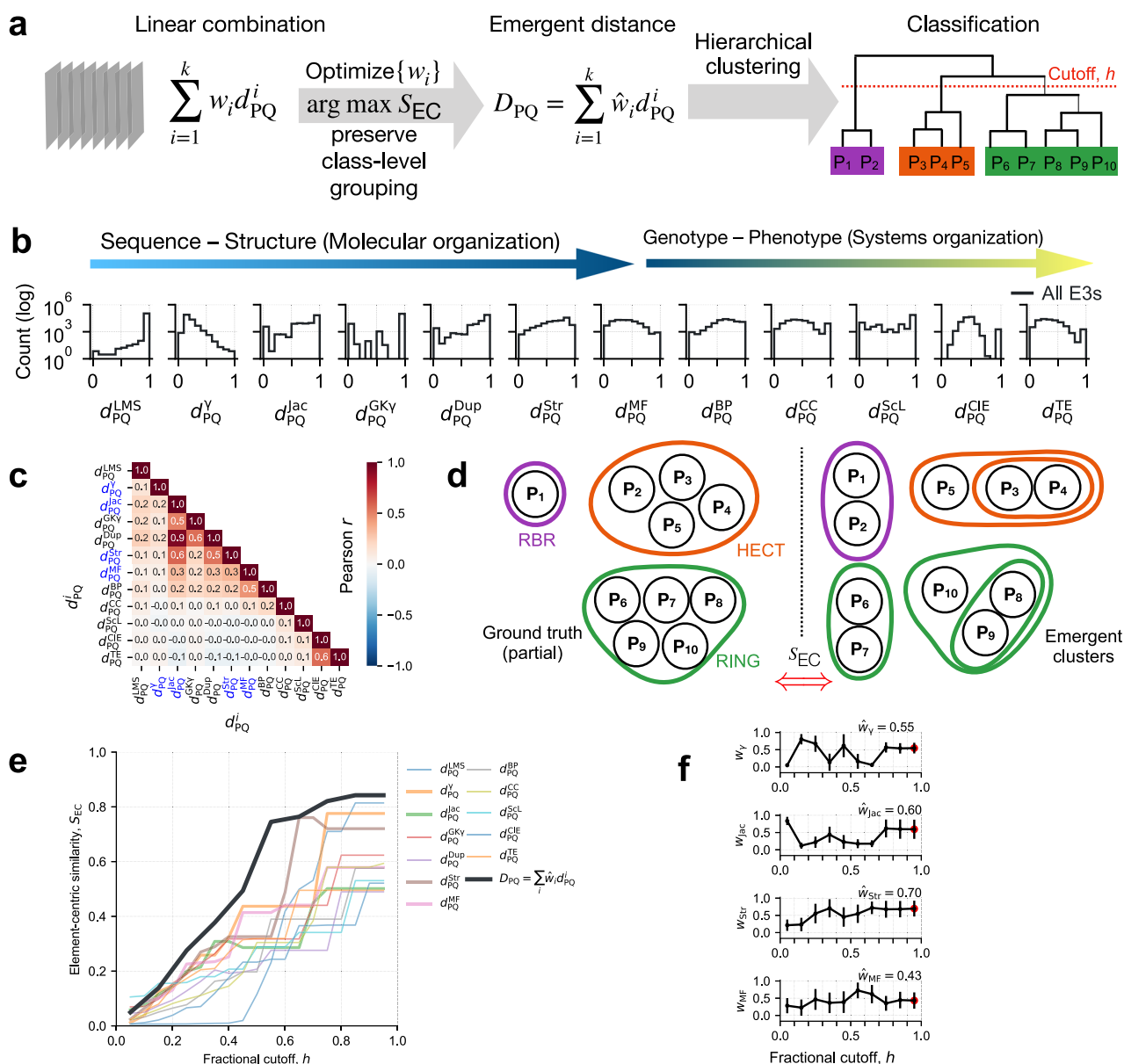


Fig. 2 | Metric learning for E3 ligases. a Schematic of the metric learning process. **b** Distribution of various pairwise distance measures spanning the molecular and systems level organization. **c** Pearson correlation of distance measures indicate orthogonality, mostly $r \in (-0.3, 0.3)$. Distances based on sequence alignment, domain composition, 3D structure (catalytic), and molecular function (marked in blue) are combined into an emergent distance (D_{PQ}) with appropriate weights. **d** By

maximizing element-centric similarity, a measure of the overlap of emergent hierarchical clusters (right) with the ground truth (left) (**e**) evaluates individual metrics and their linear combinations. **f** Regression weights (mean \pm S.D.) corresponding to the four relevant distances as a function of fractional tree cutoff h . 100 clusters with largest S_{EC} were sampled at each value of h to estimate the mean and S.D.

These distances capture intrinsic molecular attributes and their relationships spanning the molecular scale. By uniformly sampling the weights, we constructed 10^5 combination measures as a function of the hyperparameter (fractional tree cutoff, h , between 0.05 and 0.95). By simultaneously maximizing element-centric similarity²² of the emergent hierarchical clusters resulting from combined measures, with partial ground truth (weakly-supervised scheme, Fig. 2d), we optimized an emergent distance measure (D_{PQ}) with appropriate weights (\hat{w}_i). We found that the linear combination of distances provided clusters with high element-centric similarity S_{EC} compared to clusters obtained from individual distances (Fig. 2e, black curve vs. colored).

Normalized Mutual Information (NMI) and Fowlkes–Mallows Index (FMI) compare clustering assignments (various distance-

based vs. ground truth), but they are sensitive to cluster count (determined by tree cutoff, h ; Supplementary Fig. 3a). Therefore, optimized weights \hat{w}_i were obtained by averaging one hundred realizations of hierarchical clustering with maximum S_{EC} ²². The weights corresponding to maximum S_{EC} initially varied and then plateaued (at $h \geq 0.75$; Fig. 2f), resulting in the construction of an optimized emergent distance measure, D_{PQ} (Eq. (1)). We found that the relative influence of 3D structure, domain composition, and sequence alignment ($\hat{w}_i \geq 0.5$) was more significant on the final learned metric and its ability to reproduce class labels accurately. Compared to the emergent distance measure, we found variable tree topologies with poor overlap and highly entangled trees for all four individual distances (Supplementary

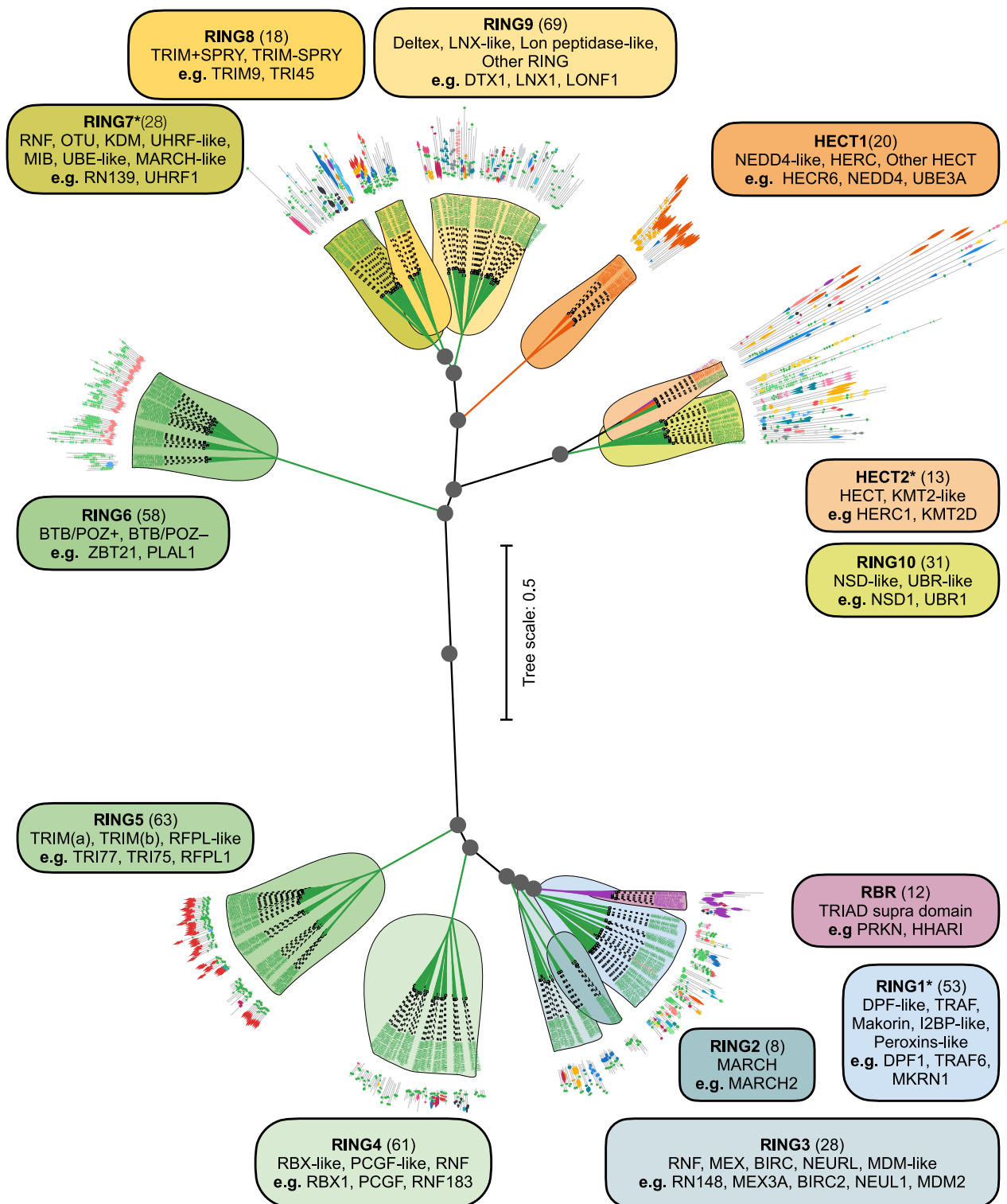


Fig. 3 | Classification of the human E3 ligome. Unrooted hierarchical tree computed using the optimized emergent distance metric D_{PQ} (scaled branch lengths). The RBR (purple), HECT (orange), and RING classes (blue/green/yellow) are partitioned at $h = 0.25$ into 1, 2, and 10 families, respectively. Each cluster is defined by

shared sequence, domain-architectural (mapped), structural, and functional elements. Boxes show family information, i.e., family name, size, and subfamilies, with representative examples. Grey-filled circles denote bifurcation nodes with $\geq 95\%$ bootstrap support, and * denotes families with a few class-level outliers (3/13).

Figs. 3b–e).

$$D_{PQ} = 0.43d_{PQ}^{MF} + 0.55d_{PQ}^V + 0.60d_{PQ}^{lac} + 0.70d_{PQ}^{Str}. \quad (1)$$

Organization of the human E3 ligome

Using the optimized emergent distance metric, D_{PQ} (Eq. (1)), we constructed a scaled hierarchical tree classifying the human E3 ligome (Fig. 3 and Supplementary Fig. 4a). To assess the validity of nodes, branch stability, and the robustness of our classification, we resampled

the emergent distance matrix ($n = 500$) and assigned bootstrap support at each branch point (Fig. 3, grey circles). The bootstrap support for all nodes beyond tree cutoff, $h > 0.15$, is 95–100%, indicating a stable branch pattern (Supplementary Fig. 4b) with a fixed tree topology. At $h \leq 0.15$, the bootstrap support for the nodes dropped drastically. This allowed us to use a tree cutoff threshold, $h = 0.25$, to parse the dendrogram and obtain robust and stable clusters with clear family and subfamily patterns while preserving RING-, HECT-, and RBR-class segregation.

We identified thirteen distinct clusters or E3 families ($h = 0.25$). At the class level, the E3 ligome is well segregated into ten RING families (Fig. 3, blue to green colors; clockwise arrangement from RING1 to RING10), two HECT (Fig. 3, top-branch; orange), and one RBR family (Fig. 3, bottom-branch; purple). Each E3 family is subdivided into one or more subfamilies (Fig. 3, boxes) with distinct patterns. Mapping domain architecture information onto the individual leaves aids recognition of well-preserved sequence and domain features, consistent with family and subfamily grouping, a pattern more evident in the unscaled circular dendrogram of the E3 ligome (Supplementary Fig. 4a). Further, few heterogeneous families are grouped more closely and emerge from single branches (bootstrap support ≈ 90 –95%; Supplementary Fig. 4b) hinting at divergence of plausible superfamilies: (i) RBR and RING1–3 branch (small E3s), (ii) RING7–9 branch (medium E3s), and (iii) HECT2–RING10 branch (large E3s). This organization stems from the central node that bifurcates the E3 ligome into two groups characterized by average protein size (Fig. 3). The bottom branch displays six families with smaller E3s, while the top branch groups seven larger E3 families.

The E3 family organization reflects mechanistic differences (Supplementary Fig. 4c and Supplementary Table 3). The RING E3s mediate the direct transfer of Ub to the substrate, while the RBR and HECT E3s enable ubiquitin transfer via a two-step mechanism. The RBR-containing E3s form a homogeneous cluster, highlighting their conserved sequence and the TRIAD supra domain. Similarly, HECT-domain-containing E3s are organized into two clusters/families, HECT1 and HECT2. The HECT1 family is homogeneous and includes three subfamilies: NEDD4-like, HERC, and other HECT E3s. The HECT2 family contains a pure HECT E3 subfamily and an outlier subfamily containing large multi-domain RING-type E3s (> 2000 residues) often with atypical mechanisms (e.g., MYCBP2, RNF213, see Supplementary Note 2). The most abundant RING-domain-containing E3s are organized into 10 families, each characterized by further grouping related proteins into distinct subfamilies with shared sequence elements, domain architectures, and structural features. For instance, the RING2 family comprises membrane-associated RING-CH-type domain (MARCH) E3 ligases (Fig. 3, bottom-right). This family includes all small MARCH E3 ligases characterized by their transmembrane domains and sequence lengths below 500 amino acids. TRIM E3 ligases are exclusively limited to two distinct families, RING5 and RING8, and feature the SPRY domain (Fig. 3, bottom-left). E3 ligases containing BTB/POZ and Zn-finger domain repeats are grouped into the RING6 family (Fig. 3, upper-left).

Although our emergent metric largely maximizes pure and homogeneous clusters (e.g., RBR, RING2, RING5, RING6, RING8, and HECT1), heterogeneity often arises at the subfamily level, resulting in sub-groupings of E3s with varied and unique domain architectures. Isolated proteins (singletons) in the RING1, RING7, RING8, and RING9 families form distinct subfamily groupings, complicating pattern detection. Only RING1, RING7, and HECT2 families display occasional class-level outliers (Supplementary Table 3). Supplementary Notes 3 to 15 describe each family structure in detail with information on subfamily branching, characteristic features, and distinct patterns, along with outliers providing a nuanced description (Supplementary Figs. 5–17 and Supplementary Notes 3–15).

Functional segregation of the human E3 ligome

To evaluate the human E3 ligome, we conducted a CRISPR-Cas9 dropout screen of UPS genes, using cellular fitness as the main phenotype. We identified 53 catalytic and 32 non-catalytic E3 components to be essential for cell fitness ($\text{FDR} \leq 0.05$ and $|\log_2(\text{FC})| \geq 1.0$; Fig. 4a, b). Notably, these essential E3s were predominantly enriched in RING1/4/7, and RING9 families, suggesting critical biological roles (Fig. 4a). Several E2 enzymes and adaptors were also essential, reinforcing the importance of the ubiquitin conjugation and multi-subunit E3s (Fig. 4b). Overall, our CRISPR screen replicates correlated well with DepMap data (Pearson $r \geq 0.5$; Supplementary Fig. 18a). GO analysis of 53 essential E3s showed significant enrichment for nuclear components and DNA damage, replication, and repair processes (Fig. 4c), indicating their roles in genome integrity and nuclear regulation fundamental to cell survival. These findings point to essential E3 components crucial for cell viability.

To understand the functional diversity of the human E3 ligome, we filtered high-confidence GO terms and mapped them onto our classification, enabling us to draw functional clusters and visualize their networks across all three ontologies. This allowed recognition of generic and family-specific functions (enriched, $-\log_{10}(p) \geq 2$). At the BP level, as expected, the network analysis revealed prominent core functional sub-clusters associated with all terms containing “ubiquitination (Ub)” (Fig. 4e, Top). These BPs are shared across all families, indicating their generality. Additional core clusters relate to innate immunity, host-driven viral restriction, NF- κ B regulation, and IL-17 signaling (Fig. 4d). Further, cooperative diverse non-degradative functions such as regulation of gene expression, protein stability, cell growth, and ERAD pathway are enriched within the E3 ligome (see Supplementary Note 16).

Examining family-specific GO enrichment uncovered functional specialization supported by experimental evidence (Fig. 4e, colored triangles; Supplementary Table 4). For instance, the RBR family members, RNF14, RNF144A, and PRKN, demonstrated specificity for K6-linked-Ub (Fig. 4f, left). K6-linked chains flag stalled RNA-protein cross-linked complexes (RNF14), DNA-sensing adaptor STING for activation of interferon signaling (RNF144), and damaged mitochondria for clearance (PRKN)^{23–25}. Similarly, TRIM E3s (RING5) were significantly enriched in antiviral innate immune response (Fig. 4f, right). They regulate pattern recognition receptor activity in cells, such as RIG-1 and MDA5-mediated responses^{26,27}.

To test family-specific functions, we performed whole-cell proteomics in response to cellular perturbations (EBSS and CPT treatment) and monitored the differential expression of E3 ligases. The cellular responses to EBSS and CPT treatments are broad and multifaceted, impinging on a wide range of BP functional clusters (Fig. 4e, blue and orange triangles). 18/34 implicated E3s have direct evidence linking them to starvation response pathways (Supplementary Table 4). Notably, MGRN9 and BRCA1 were also essential in our CRISPR screens (Fig. 4g, left). Our analysis revealed a differential expression (p value ≤ 0.05) of the E3 ligases TRIM27/32, and UBR1 in addition to key autophagy regulators SQSTM1, CALCOCO2, GABARAPL2, and MAP1LC3B2, highlighting a coordinated modulation of the autophagic machinery during starvation (Fig. 4g, right colored vs. blue). In contrast, the observed up-regulation of EIF4EBP1, a translational repressor regulated by mTOR signaling²⁸, and GDF15, a stress-responsive cytokine involved in metabolic adaptation²⁹, indicates activation of complementary stress response pathways that may support cellular survival. TRIM27/32 regulate autophagy initiation and selective degradation pathways by ubiquitinating essential autophagy proteins such as ULK1 and p62, thereby promoting autophagosome formation and cargo recognition^{30–32}.

Similarly, experimental evidence links 29/87 implicated E3s directly to DNA damage response (DDR) (Supplementary Table 4), of which 12 E3s turned out to be essential (e.g., TRAF3/7, MDM2; Fig. 4h,

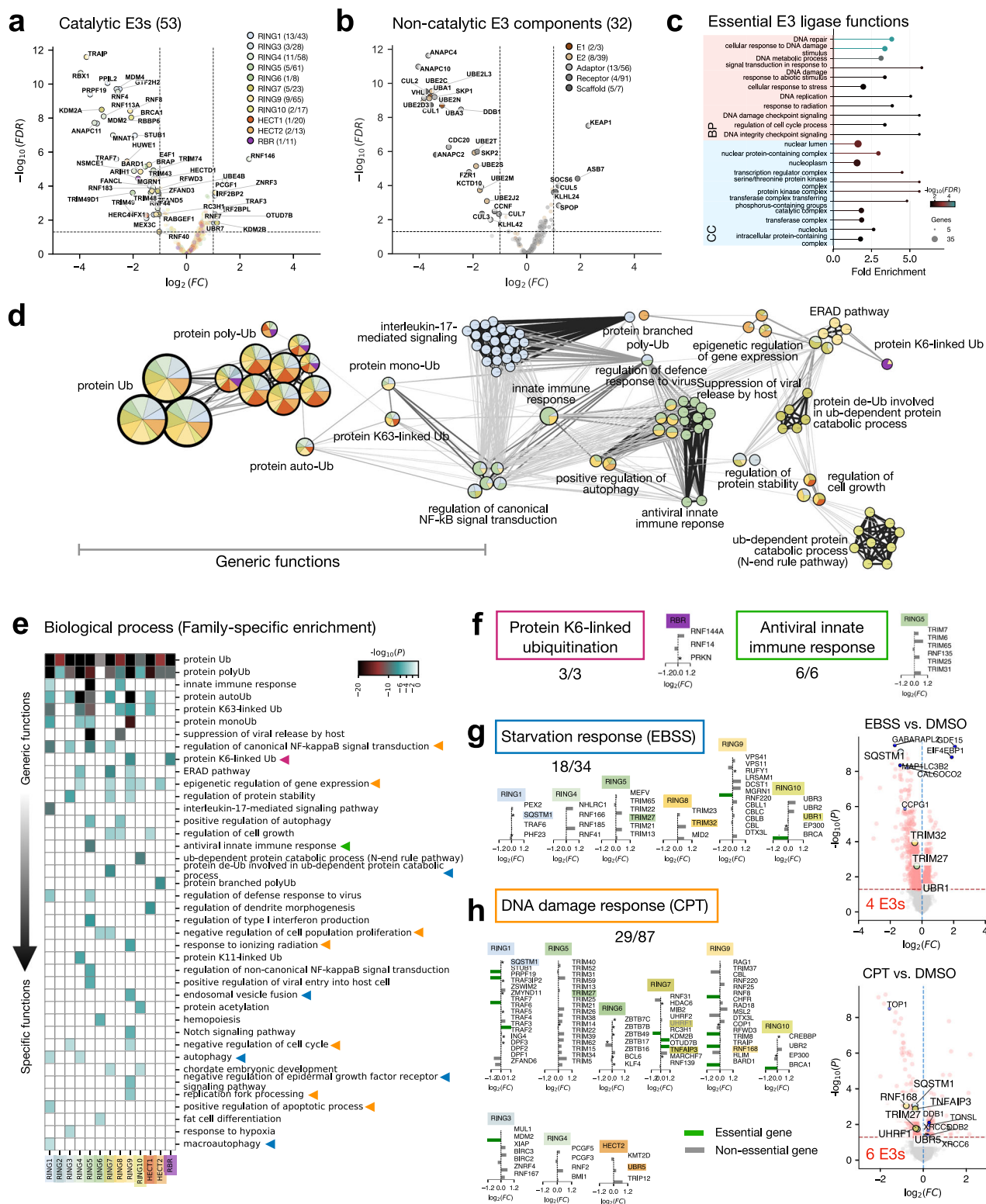


Fig. 4 | Functional segregation of the E3 ligome. Volcano plots of Gene essentiality analysis derived from CRISPR screens for (a) catalytic and (b) non-catalytic components of the E3 ligome. c GO enrichment analysis for essential catalytic E3s. d The functional landscape of the E3 ligome (biological processes) is captured by the network of GO annotation clusters. Individual nodes representing GO clusters (20 labeled) are drawn as pie charts (sizeproportional to # of E3s; colored by family enrichment) connected by distinct edges (κ -similarity ≥ 0.3). e The heatmap displays all functional clusters corresponding to family-specific enrichment of E3s (p value estimated using hypergeometric test (two-sided), discrete color scale for p value ≤ 0.01 ; white otherwise). Colored triangles show examples of family specific

enrichment for (f) K6-linked ubiquitination (purple) and antiviral innate immune response (green), (g) starvation response under 6h EBSS treatment (blue), and (h) DNA damage response under 4h 100 nM CPT treatment (orange). For panels f–h gene essentiality data $\log_2(FC)$ or DepMap Gene Effect scores (*) are plotted for individual E3s. The ratio denotes the fraction of E3s with experimental evidence (PMIDs) for GO functions directly. g, h panels also show volcano plots of proteomic analysis, highlighting significantly up-regulated and down-regulated proteins (red scatter; adjusted p values were obtained using Benjamini-Hochberg method in two-sided moderated t-tests) with overlapping E3s (colored) and control proteins (blue filled circles).

(RING9), and unmethylated CpG binding (RING7). Other non-catalytic E3 domains mediate PPI interactions with actin, SH3-domains, Kinases, and proteases. Distinct subcellular localization of E3s exerts spatial control of protein-Ub (Supplementary Fig. 18c). Most E3 ligases are cytosolic, which form an essential part of the “ubiquitin ligase complexes” (CC: Generic). We find enriched E3 families for distinct CCs: CD40 receptor, GID, and nBAF complexes (RING1); early endosomes and lytic vacuoles (RING2). In total, we identified 17 unique cellular components with distinct E3-specific enrichment patterns (Supplementary Figs. 18e, 19a).

GO annotations from author/curator statements and electronic methods—despite varying confidence levels—provide a rich, abundant dataset of key testable hypotheses on E3 systems (Supplementary Figs. 18f, 19a, and Supplementary Note 17). Publication counts for each protein-specific annotation highlight knowledge gaps and pinpoint well-studied and underexplored E3 systems (Supplementary Fig. 19b).

Interaction landscape of the human E3 ligome

E3 ligases can operate as standalone or multi-subunit enzymes. Complex E3s consist of scaffolds, adaptors, and substrate receptors that determine specificity, stability, and regulation²¹. For instance, RBX1 binds scaffolds (CUL1–CUL5) and anchors the E2 enzyme to form the catalytic core for Ub transfer (Fig. 5a). Its interactions with various cullins, adaptors, and receptors enable ~250 CRL configurations, providing modular regulation and substrate specificity. By contrast, standalone E3 ligases, like MDM2, c-CBL, PARKIN, or SMURF1/2, either have specialized domains or undergo specific PTMs that recognize substrates and facilitate E2 binding and ubiquitin transfer. For example, HECTD3 operates via a two-step ubiquitin transfer mechanism (Fig. 5b). However, substrate binding occurs through specific motifs within the non-HECT regions that presumably recognize particular degrons (sequence motifs, distinct PTMs, or unique structural elements).

Previous annotations^{38,39} report 6 complex, 329 standalone, and several unclassified E3s. We extended this annotation by curating non-catalytic components and cataloging their direct interactions with E3s (Fig. 5c). Multi-subunit complex structures are only resolved for four E3s (RBX1/2, ARI1, and APC11). While partial complexes are resolved for 12 E3s (e.g., APC11, ARI1/2), we found several binary direct interactions between E3s and non-catalytic subunits, re-annotating 75 E3s operating as complexes (Fig. 5d, black), leaving 277 standalone³⁸ and 110 unclassified E3s (Fig. 5d, red). RING8 family displayed many complex E3s (50%), followed by RING1 (26%), while RING2 and HECT2 families displayed entirely standalone E3s (Fig. 5e, Supplementary Table 5). Consistent with our findings, we observe that MARCH-type E3s (RING2) operate in the membrane environment primarily as standalone enzymes. HECT2 proteins (e.g., HECTD3) possess multiple domains for adaptor, receptor, and scaffolding, explaining their standalone function.

Next, we assembled the E3–substrate interaction (ESI) network by integrating data from known ESIs ($n = 2012$), direct PPIs ($n = 5844$), indirect PPIs ($n = 6530$), and predicted ESIs ($n = 64802$; Fig. 5f, see Methods). Integrating these data and verifying their ubiquitination status resulted in excluding false positives (E3-associated) and

improving the annotation of likely substrates (Supplementary Figs. 20a–c). This enabled mapping $\approx 62\%$ ($n = 7691$ substrates) of the ubiquitinated human proteome (Fig. 5g).

Analysis of the E3–substrate network revealed distinct specificity patterns. Using well-known ESIs, we observed that most E3 ligases have a single substrate ($\sim 10^2$), fewer target multiple ($\sim 10^1$), and only a handful possess extensive substrate sets (Supplementary Fig. 20d). Most substrates are targeted by E3s belonging to two or more families ($n = 5520$ Promiscuous substrates, $n = 72749$ interactions; Fig. 5h, bottom; Supplementary Table 6). We also identified substrates that are potentially ubiquitinated by two or more E3s belonging to the same E3 family ($n = 804$ Family-specific substrates, $n = 3292$ interactions; Fig. 5h, middle) and substrates uniquely targeted by specific E3 ligases ($n = 1367$ E3-specific substrates, $n = 1367$ interactions; Fig. 5h, top).

For instance, the E3 ligase SMUF1 specifically targets TBX6 for degradation during cell differentiation⁴⁰. Similarly, MARCH5 specifically targets FIS1 for ubiquitination (Fig. 5i) to regulate mitochondrial fission⁴¹. Both NEDD4 and ITCH belong to the HECT family and ubiquitinate MART1 to exert complementary functions for the sorting and degradation⁴², and PACS2 is ubiquitinated by BIRC2 and BIRC3, members of the RING3 family (Fig. 5i), conferring TRAIL resistance to hepatobiliary cancer cell lines⁴³. CDN1A, a central regulator of cell cycle and DNA damage response, is ubiquitously targeted by diverse E3 ligases, linking multiple signaling pathways to replication checkpoints (Fig. 5i). Overall, integrating the ESI network obtained here with ubiquitination and degron mapping efforts provides a powerful framework to uncover additional E3–substrate relationships, validate candidate ligases, and reveal functional redundancies or specificity (Supplementary Note 17).

Druggability map of the human E3 ligome

To learn likely avenues of proximity-based therapeutics and leverage the relationships within the human E3 ligome, we mapped existing E3 handles derived from known Proteolysis Targeting Chimeras (PRO-TACs) and E3 binders to individual E3s and their families (Supplementary Fig. 21a, Supplementary Table 7). Only 16 proteins (9 catalytic E3s and 7 adaptors) are directly targeted by existing E3 handles (Fig. 6a, top). A large fraction of the designed E3 handles are specific to adaptor proteins (e.g., VHL, CRBN), and only a very select few directly target the catalytic E3s (e.g., XIAP, MDM2/4/7, BIRC2/3/7; Supplementary Fig. 21b). Nearest neighbor analysis using our E3 ligome identified five closely related proteins (BIRC8, RN166/181/141, and UBR2; Fig. 6a, top; grey boxes). Given their high structural similarity (often paralogs), the same E3 handles could be repurposed to target them. Mapping small-molecule E3 binders gave us a potential set of compounds targeting 25 additional E3s and 15 non-catalytic components, thus identifying unexplored targets and avenues for lead development for the rational design of E3 handles (Fig. 6a, bottom; red labeled).

Next, we mapped the chemical landscape of E3 handles and E3 binders using the Uniform Manifold Approximation and Projection (UMAP) embedding of high-dimensional 2048-bit Morgan fingerprints. By clustering them, we visualized their molecular similarities. Comparison with other dimensionality reduction methods, such as PCA and

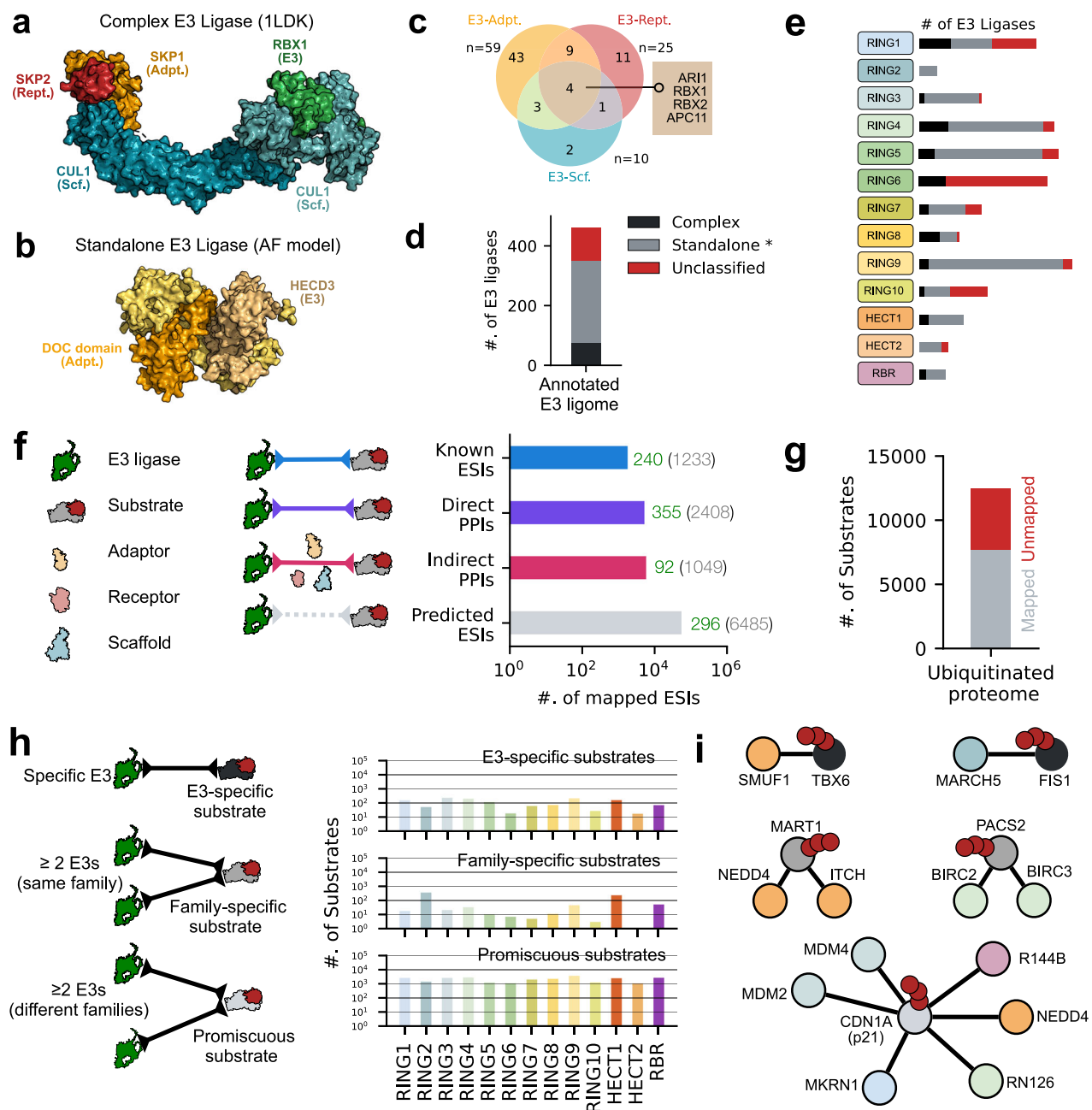


Fig. 5 | Protein-protein interactions of the E3 ligome. Representative examples of E3 ligases functioning as a (a) multi-subunit protein complex (CRL) or (b) a standalone enzyme (HECD3). c Venn diagram of pairwise interactions of adaptors, receptors, and scaffold proteins with E3s. d Annotation of 462 E3 ligases into complex, standalone, or unclassified modes of action. e Family-wise mapping of data from (d). f Pairwise E3-substrate interactions for all E3s obtained by

integrating data from known ESIs, mapped transient direct and indirect PPIs, and predicted ESIs. g Mapping of the ubiquitinated proteome with E3s ($\approx 62\%$, $n = 12464$). h Schematic showing substrate categorization into E3-specific, family-specific, and promiscuous classes (left) and their relative distributions mapped onto E3 families (right). i Representative examples for the three types of ESI networks.

t-SNE, did not provide optimal clusters (Supplementary Figs. 21c–d). We detected 20 chemically distinct clusters within the UMAP space (Fig. 6b, indexed, distinct colors) by fast search and identification of density peaks⁴⁴. 12 compound clusters contain only E3 binders, while the remaining 8 also contain E3 handles targeting specific proteins (Supplementary Fig. 22). The top 6 clusters occupy a large central subspace, indicating cluster heterogeneity (relatively large chemical space). These compound clusters predominantly target proteins from RING3 and RING7 and adaptors (9123 molecules, 67%). E.g., the cluster #2 contains 1222 ligands forming 5 dense sub-clusters representing distinct chemo-types that selectively target MDM2/4, IRAK4, XIAP,

HDAC6, and BIRC2, respectively (Fig. 6c). Further, the compounds from clusters #1–6 target a large number of proteins and display less specificity (Fig. 6d; $LP_{ij} \leq 3$, low-binding likelihood). The clusters #9–20 are relatively more homogeneous, occupy small peripheral regions of the UMAP-space, and display high selectivity (1–3 targets/cluster with $LP_{ij} \geq 3$, cyan squares).

3D structures of compound-E3 interactions are limited. Our mapping showed that only 231/14615 pair-wise interactions are resolved in the PDB (Fig. 6e, covering 212/13620 compounds). These structures mainly feature well-studied E3 handles and binders, highlighting clear examples for drug repurposing. In cluster #13, the

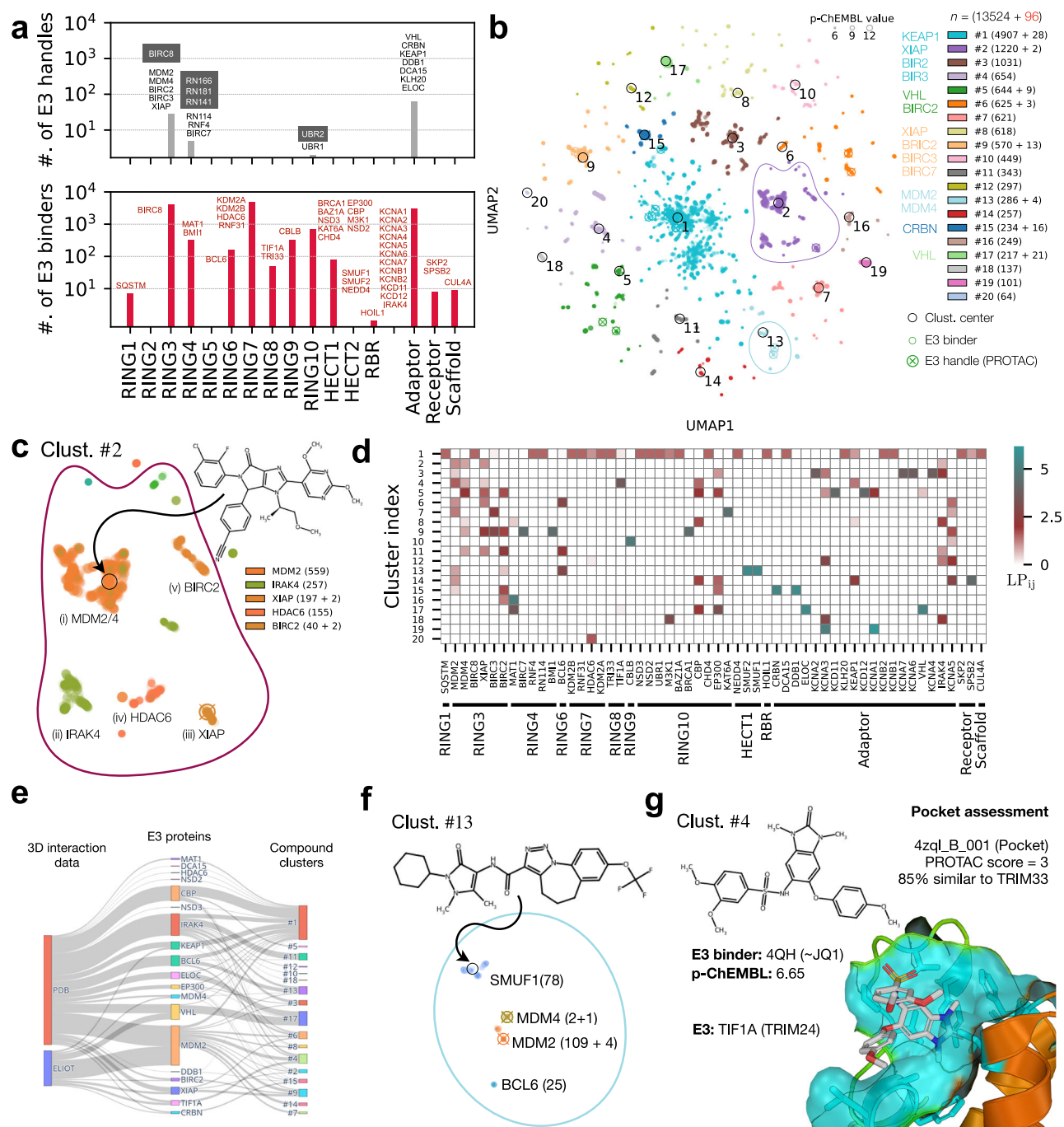


Fig. 6 | Druggability map of the E3 ligase. **a** Distribution of known E3 handles (extracted from PROTACs, top) and expanded set of E3 binders (potential lead compounds, bottom) targeting E3 families. Individual proteins uniquely targeted by E3 handles ($n = 16$, black) and E3 binders ($n = 40$, red) are displayed for each family. Grey-filled boxes (top) show closely related protein targets for E3 handle/PROTAC repurposing. **b** Reduced 2D UMAP chemical space of E3 handles ($n = 96$) and E3 binders ($n = 13524$); sizeproportional to p-ChEMBL value. Compound clusters (colored) within UMAP space represent distinct chemical structures (Cluster centers indexed #1–20) are identified by local density peaks (see Supplementary Figs. 21). **c** Magnified view of cluster #2 showing dense sub-clusters of compounds targeting multiple proteins. **d** Log-transformed propensities, LP_{ij} of individual

compound clusters capture binding likelihood. **e** Sankey plot showing the map between PDB (3D interaction), ELIOT (pocketome), individual E3 proteins (19/56), and their compound clusters (covering 212/13620) from our small-molecule interaction analysis. **f** Magnified view of cluster #13 showing the proximity of E3 handles targeting MDM2/4 to SMUF1/BCL6 binders, inferred from compound similarity and clustering in the reduced UMAP representation. **g** Example of a potential lead compound identified from cluster #4. Ligand 4QH (similar to JQ1) binds to TIF1A bromodomain (PDB code: 4ZQL) and can be developed into a specific E3 handle. Binding site analysis (from ELIOT) indicates a favorable PROTAC score and high similarity to the TRIM33 bromodomain pocket (RING8 member).

MDM2/4-specific E3 handles could be repurposed to target SMUF1 (HECT) or BCL6 (RING6). The presence of SMUF1 binders or BCL6 binders, forming dense sub-clusters nearby, reflects their chemical similarity (Fig. 6f). Further, analysis of surface pockets of E3s could

provide structural context and reveal similarity patterns for the development of lead compounds into E3 handles. To this end, we integrated the ELIOT data⁴⁵, which includes detailed pocket analysis for 73 compounds bound to 8 E3 ligases from our small-molecule

interaction dataset ($n = 74$, Fig. 6e). Interestingly, in cluster #4, we found a compound, 4HQ, bound to the bromodomain of TIF1A (Fig. 6g). Given its moderate affinity and favorable ELIOT pocket assessment, it could be developed into a specific E3 handle for either TIF1A or homologous TRI33 (85% pocket similarity). Moreover, 4HQ is similar to JQ1, a well-known ligand targeting the bromodomain of human BRD4 (PDB code: 3mxf). Further, integration and comparison of sub-cellular location, cell- and tissue-specific expression patterns of E3s will be essential for effective drugging of the human E3 ligome (Supplementary Fig. 23). In summary, this approach enables rational and family-informed PROTAC design by linking chemical tractability with context-aware E3-substrate relationships (Supplementary Note 17).

Discussion

Navigating the vast and complex landscape of E3 ligase biology requires a comprehensive approach. Despite decades of dedicated investigation, the intricate diversity and functional complexity of E3 ubiquitin ligases continue to pose a significant challenge. In decoding this complexity, we first curated and filtered E3 ligases, ensuring data accuracy, consistency, and relevance for all downstream analyses. By assigning confidence scores to each ligase and employing stringent inclusion criteria, we remove false positives and improve annotation, providing a high-quality and comprehensive human E3 ligome. Ultimately, this simplification facilitated the identification of key catalytic components and paved the way for applying machine learning and algorithmic approaches to E3 systems.

The human E3 ligome exhibits remarkable heterogeneity, evident in its diverse sequence, domain architectures, structures, and functions. This diversity is shaped by not only the evolutionary forces influencing domain shuffling and genetic rearrangements but also biophysical forces influencing molecular recognition and spatio-temporal regulation of enzymatic reactions, leading to specialization and adaptation⁴⁶. To effectively categorize E3 ligases, we require overarching organizational principles delineating broad evolutionary clans and functionally distinct subgroups within the E3 ligome. Hierarchical classification captures organizational principles, achieves higher prediction accuracy, and can handle previously uncharacterized data and class imbalances more effectively⁴⁷. These methods enable a more precise and context-aware organization of proteins, facilitating the recognition of salient and unique features⁴⁸. However, its performance heavily depends on choosing an appropriate metric reflecting authentic relationships.

Assessments of similarity and distance are critical components of human cognitive function and constitute a foundational element in developing machine learning and applying data mining techniques⁴⁹. Using a weakly supervised learning paradigm, we optimized a linear metric that is simple, scalable, and straightforward to interpret with broad applicability. We bridged the molecular scale from protein sequence, domain architecture, 3D structure, and molecular function, resulting in a unique measure capable of detecting subtle shifts, reproducing class-level grouping of E3s, and improving family and subfamily definitions.

We present a multi-scale classification model enabling a comprehensive analysis of the E3 systems. We classify the E3 ligome into 13 distinct E3 families characterized by shared domains, comparable architectures, and similar 3D structures. By moving beyond traditional taxonomic methods and subjective, ad hoc classifications, our multi-scale model provides a systematic and objective framework. Although not explicitly dependent on any individual distance measure, it is strongly associated with shared structural similarities and domain architectures, providing exceptional resolution into functional specialization and mechanistic action of E3s. The largest RING class, divided into 10 families, shows remarkable diversity. For instance, MARCH, TRIM, and BTB groupings reflect domain architecture. The

RBR class exhibits striking homogeneity, suggesting strong evolutionary conservation⁵⁰. The HECT class is split into two individual families (HECT1 and HECT2), consistent with the previous classification⁵¹. These organizational insights lead to interesting additional hypotheses, revealing unexplored roles for existing E3s.

The functional diversity of the E3 ligome serves as a central hub for cellular homeostasis, extending far beyond the classical proteasomal degradation. CRISPR-Cas9 dropout screens identified a substantial number of catalytic E3 ligases and non-catalytic components as essential for cellular viability. They predominantly contribute to critical cellular processes, including DNA replication, repair, and maintenance of genome stability. Given the limitations of experimental data on E3 functions and the inherent challenges in quantifying them, GO terms serve as functional proxies. We find that the generic functions of E3 ligases are conserved across all families, encompassing protein ubiquitination, modification, and degradation (BPs); localization to E3 complexes or the cytosol (CC); and catalytic activity driving ubiquitin transfer (MF). More importantly, our classification captures functional segregation of E3 systems, providing significant insights into the distinct biochemical and functional mechanisms regulated by individual families. In test cases, we highlight how the RBR family promotes K6-ubiquitination²³, and RING5 E3s are involved in innate immune responses²⁶. By integrating known experimental evidence, gene essentiality data, and proteomic profiling of HeLa cells, we mapped E3 systems involved in global starvation³⁰ and DNA damage response^{52,53}. Furthermore, our quality analysis identifies additional testable hypotheses (from automated GO terms) and catalogs under-characterized ligases.

Mapping the protein interaction landscape of the whole E3 ligome is challenging. We integrate disparate datasets to build enzyme-substrate network maps for each ligase family. We found that RING1/3/8 and RBR members contain more multi-subunit E3 complexes, while RING2 and HECT2 are likely to operate in a standalone manner, directly recruiting substrates. Further, we could classify substrate molecules into E3-specific, family-specific, and promiscuous substrates, providing foundational data for understanding the molecular principles of substrate recognition. Recognition of shared patterns in substrates can point to a better understanding of individual and group-specificity. Further, orthogonal data on subcellular localization of E3s and substrates, their tissue-specific expression patterns, explain the context-dependent ESIs and the prevalence of promiscuous substrates.

Targeted protein degradation via PROTACs is a promising therapeutic strategy to target previously undruggable proteome⁵⁴. Despite its potential, progress in targeting unexplored E3s and the rational design of specific E3 handles has been gradual. Most often, PROTACs and glue-like compounds exploit ligands against well-known adaptor proteins like CRBN- and VHL-dependent modalities to target CRLs for the specific degradation of substrates. Only a few E3s have been directly targeted using PROTACs^{55,56}. By leveraging the E3 ligome structure, we extend the map of E3 handles, increasing the likelihood of repurposing existing PROTACs to target closely related E3s in a family-specific manner. Further, by mapping an expanded set of E3 binders and associating them with unexplored E3s, we build a curated set of lead compounds with unique chemical signatures for further rational design of specific E3 handles. Although 2D-based UMAP visualizations offer a computationally efficient overview of chemical diversity and aid in hypothesis generation, they may fail to capture activity cliffs or bio-isosteric relationships. Complementing the druggability map generated here through physics-based approaches, including 3D conformational sampling and HTP docking screens of ligands against the druggable E3 pocketome⁴⁵ will propel the rational design of specific E3 handles. Furthermore, exploiting the emergent relationships offered by the E3 ligome, in combination with other metrics capturing their

localization, cell-, tissue-expression, enriched ESI networks, and a list of already targeted and unexplored E3 binders, allows an efficient drugging strategy for unexplored targets in health and disease.

In conclusion, the multi-scale classification framework developed here provides a comprehensive global view of the human E3 ligome. Mapping disparate multi-modal and multi-resolution data onto the ligome structure, such as functions, interactions, and druggability, provides a systems-level understanding, enabling high-throughput screening and profiling. The metric learning paradigm developed here is simple and transferable to other areas of data-driven biology. More specifically, the human E3 ligome offers clear actionable guidelines and a unified framework for systematic annotation and interpretation of ubiquitination data, enabling integration with high-throughput proteomics and genetic screens, including spatial proteomics to identify functionally relevant E3s, uncover pathway-specific patterns, and guide rational design of targeted degradation strategies. We anticipate that the data and insights presented here will stimulate further research into E3 systems and drive the development of innovative therapeutics.

Methods

Assembly of the human E3 ligome

We collected eight individual human E3 ligase datasets (A_1, \dots, A_8) including previously published reports^{17–19} and public repositories: E3Net³⁹, UbiHub³⁸, UbiNet 2.0⁵⁷, UniProt (retrieved on 2023-02-13 with keyword “e3 ubiquitin-protein ligase”)⁵⁸, and BioGRID (retrieved on 2022-01-26)⁵⁹ compiled using multiple distinct criteria. We merged all of them to form an initial dataset ($|\bigcup_{n=1}^8 A_n| = 1448$), visualized the overlap of individual resources using UpSet plot⁶⁰, and assigned a consensus score to each entry based on its presence/absence among the source datasets. We then compiled a list of distinct, well-studied E3 catalytic domains from InterPro⁶¹ corresponding to RING, HECT, and RBR classes from all published sources ($C = \{d_i\}$). Using the presence of characteristic catalytic domain(s) d_i within each polypeptide, we identified and filtered 1448 proteins corresponding to all catalytic subunits of E3 ligases, $\{X_i \in \bigcup_{n=1}^8 A_n | \exists d_i \in C\}$. This was followed by manual curation based on InterPro domain descriptions of possible catalytic activity (E2-binding and Ub transfer) to obtain the final refined set of 462 E3 ligases (E3 ligome).

In multisubunit E3 ubiquitin-ligases—most prominently the Cullin-RING ligases (CRLs)—three functionally distinct classes of subunits cooperate to bring an E2-Ub in proximity to a specific substrate. Large, rigid, central scaffold proteins (e.g., the Cullin family, Cul1–Cul5) organize the ligase complex by simultaneously binding the catalytic RING-finger subunit and the docking sites for adaptors or receptors. Adaptor proteins bridge modules that link a scaffold’s N-terminal docking surface to a separate substrate-receptor. They often recognize a conserved motif on the receptor subunit (e.g., F-box, SOCS-box, WD40 proteins). The receptor proteins are specificity determinants that directly recognize and bind degron motifs on substrates. They define which substrates get ubiquitinated (e.g., Skp2, Keap1, VHL). We independently annotate and classify each of the three classes: 151 adaptors, 106 receptors, and 8 scaffold proteins, and use their PPIs to map distinct substrates for multi-subunit E3s.

Multi-scale distance measures

We encoded the pair-wise relationship of E3 ligases by computing twelve distinct distances (d_{PQ}) spanning several granularity levels: primary sequence, domain architecture, tertiary structure, function, subcellular location, and cell line/tissue expression. All the distance measures were scaled between [0, 1] for comparison and even combination.

At the sequence level, we used an alignment-free local matching score-based (LMS) distance and an alignment-based γ distance between protein pairs using the canonical isoform sequences. The LMS distance d_{PQ}^{LMS} between two proteins P and Q is given by

$$d_{PQ}^{\text{LMS}} = 1 - \frac{2\text{LMS}(P, Q)}{\text{LMS}(P, P) + \text{LMS}(Q, Q)}, \quad (2)$$

where $\text{LMS}(P, Q) = \sum_{i \in \{P', Q'\}} M[i, i]$ captures the extent of local similarity by summing BLOSUM62 substitution scores for overlapping 5-residue fragment pairs $\{P', Q'\}$ from proteins P and Q^{62,63}. The pairwise γ distance measures the evolutionary distance between the globally aligned sequences of two proteins, P and Q, where p_{PQ} is the fraction of alignment positions with residue substitutions and indels, and $a = 2^{64}$.

$$d_{PQ}^{\gamma} = a[(1 - p_{PQ})^{-1/a} - 1], \quad (3)$$

To quantify the preservation of domain architectures among all protein pairs, we computed three distances: Jaccard, Goodman–Kruskal γ , and domain duplication distances, using domain annotations obtained from the InterPro database⁶¹ (Nov 2022). The Jaccard distance^{65,66} represents the compositional similarity of protein domains. It is the ratio of the number of shared (N'_{PQ}) and unique domains (N'_P, N'_Q) between proteins P and Q,

$$d_{PQ}^{\text{jac}} = 1 - \frac{N'_{PQ}}{N'_P + N'_Q - N'_{PQ}}. \quad (4)$$

The Goodman–Kruskal γ distance compares the order of domain arrangements between two proteins, P and Q, and is computed as

$$d_{PQ}^{\text{GKy}} = 1 - \frac{1 + \gamma_{PQ}}{2}, \quad (5)$$

where $\gamma_{PQ} = (N_{PQ}^S - N_{PQ}^R) / (N_{PQ}^S + N_{PQ}^R)$ with N_{PQ}^S and N_{PQ}^R denoting the same- and reversed-ordered pairs of proteins P and Q, respectively^{66,67}. Finally, the domain duplication distance⁶⁶ compares the overlap of tandem domain repeats and is given by

$$d_{PQ}^{\text{Dup}} = 1 - \exp \left[- \sum_{i=1}^{N'_P + N'_Q} \frac{|N_i^P - N_i^Q|}{S} \right], \quad (6)$$

$$\text{where } S = \sum_{i=1}^{N'_P + N'_Q} \max(N_i^P, N_i^Q);$$

N'_P and N'_Q are unique domains in proteins P and Q with N_i^P and N_i^Q repeats, respectively.

To compute distances between structures of pairs of ligases, we used AlphaFold2 models (version 4)⁶⁸. We restricted comparisons to contiguous protein segments containing all catalytic domains for each protein to avoid comparing flexible regions of the full-length structures. We computed the TM-score as implemented in US-align⁶⁹. The TM-score between the 3D structures of proteins P and Q is given by,

$$\text{TM-score}(P, Q) = \max \left[\frac{1}{L_P} \sum_i^{L_{\text{ali}}} \frac{1}{1 + \left(\frac{d_i}{d_0(L_P)} \right)^2} \right], \quad (7)$$

where L_P is the length of protein P, L_{ali} is the number of common residues between aligned proteins P and Q, and $d_0(L_P) = (1.24 \sqrt[3]{L_P - 15}) - 1.8^{69}$. To account for the inherent

asymmetry in the TM similarity scores due to normalization by reference protein length L_P , we computed the structural distance between protein structures P and Q by averaging their TM similarities as

$$d_{PQ}^{\text{Str}} = 1 - \frac{\text{TM} - \text{score}(P, Q) + \text{TM} - \text{score}(Q, P)}{2}. \quad (8)$$

Functional distances among the protein pairs P and Q were captured using semantic similarities of annotated GO terms corresponding to the three GO ontologies—molecular functions, biological processes, and cellular components—using the package GOGO⁷⁰. The protein to GO term mappings were retrieved (in Feb. 2023) from the Open Biological and Biomedical Ontology Foundry and the Gene Ontology resource^{48,71}. For each annotated GO term x , we obtained a directed acyclic graph $\text{DAG}_x = (x, T_x, E_x)$ with nodes T_x and edges E_x . We defined the semantic contribution, following Wang et al.⁷², $S_x(t)$ of a GO term t to the target term x as

$$S_x(t) = \begin{cases} 1 & \text{if } t = x, \\ \max(w_e S_x(t') \mid t' \in \text{children}(t)) & \text{if } t \neq x. \end{cases}$$

Further, the semantic similarity between two GO terms x and y , represented by two graphs DAG_x and DAG_y , is defined as

$$\text{Sim}_{\text{Wang}}(x, y) = \frac{\sum_{t \in T_x \cap T_y} S_x(t) + S_y(t)}{\sum_{t \in T_x} S_x(t) + \sum_{t \in T_y} S_y(t)}.$$

By extension, the semantic similarity between a single GO term x and a set of GO terms $\text{GO}_y = \{y_1, y_2, \dots, y_k\}$ is defined as the maximum semantic similarity between x and any of the terms in Y :

$$\text{Sim}(x, \text{GO}_y) = \max_{1 \leq i \leq k} \text{Sim}_{\text{Wang}}(x, y_i).$$

Finally, the semantic distance between proteins P and Q, annotated with sets of GO terms $\text{GO}_P = \{p_1, p_2, \dots, p_m\}$ and $\text{GO}_Q = \{q_1, q_2, \dots, q_n\}$, respectively, is calculated as

$$\begin{aligned} d_{PQ}^{\text{Sem}} &= 1 - \text{Sim}(\text{GO}_P, \text{GO}_Q) \\ &= 1 - \frac{\sum_{1 \leq i \leq m} \text{Sim}(p_i, \text{GO}_Q) + \sum_{1 \leq j \leq n} \text{Sim}(q_j, \text{GO}_P)}{m + n} \end{aligned} \quad (9)$$

Using Eq. (9), we computed three semantic distances d_{PQ}^{BP} , d_{PQ}^{CC} , and d_{PQ}^{MF} for the three different GO ontologies.

To compute the subcellular localization distance d_{PQ}^{SCL} , each protein's main and auxiliary subcellular locations were mapped from the Human Protein Atlas⁷³ and used to construct a location vector with weights 1 and 0.3, respectively. We then computed d_{PQ}^{SCL} using the cosine similarity between the location vectors of proteins P and Q as

$$d_{PQ}^{\text{SCL}} = 1 - \frac{\mathbf{P} \cdot \mathbf{Q}}{\|\mathbf{P}\| \|\mathbf{Q}\|}. \quad (10)$$

Finally, we computed the tissue (d_{PQ}^{TE}) and cell line co-expression (d_{PQ}^{CLE}) distances from the tissue and cell line expression profiles of the proteins P and Q. We retrieved expression data from the Human Protein Atlas⁷³, transcripts per millions of mRNA levels from the 253 human tissues of RNA HPA tissue gene dataset and 1055 cell lines of RNA HPA cell line gene dataset, respectively. Both distances were calculated using Spearman's rank correlation coefficient $r_{S,PQ}$ as

$$d_{PQ}^{\text{TE}} = 1 - \frac{1 + r_{S,PQ}^{\text{TE}}}{2} \text{ and } \quad (11)$$

$$\begin{aligned} d_{PQ}^{\text{CLE}} &= 1 - \frac{1 + r_{S,PQ}^{\text{CLE}}}{2}, \\ \text{where } r_{S,PQ} &= \frac{\text{cov}(R(P), R(Q))}{\sigma_{R(P)} \sigma_{R(Q)}}. \end{aligned} \quad (12)$$

Metric optimization, clustering, bootstrapping, and classification

We combined the pairwise gamma (d_{PQ}^{γ}), Jaccard (d_{PQ}^{Jacc}), structural (d_{PQ}^{Str}), and semantic molecular function (d_{PQ}^{MF}) distances to capture all orthogonal information from the four significant hierarchies—sequence, domain architecture, 3D structure, and molecular function—into a single metric spanning the entire molecular scale. We used a weighted-sum model of these four distances, $D_{PQ} = \sum_{i=1}^4 w_i d_{PQ}^i$, by uniformly sampling the weights as a function of tree cutoff, h , a hyperparameter. Optimized weights, \hat{w}_i were obtained by maximizing the element-centric similarity index²², which represents the similarity between clusters derived from parsing the emergent dendrogram, at evenly spaced cutoffs, $h \in (0, 1)$ derived from the combined distance and the class-level grouping of E3s into RING, HECT, and RBR classes (partial ground truth). At each cutoff h , we sampled $\sim 10^4$ emergent distance matrices ($\sum_i w_i d_i$), obtained their emergent hierarchical clusters, and computed S_{EC} for each one of them. We chose 100 emergent metrics with the highest S_{EC} for each h and computed the averages and standard deviations of their corresponding weights. The stabilized weights \hat{w}_i at $h \geq 0.9$ corresponding to the maximum S_{EC} were chosen to construct the optimized distance measure. Dendrograms were computed from hierarchical clustering of individual and combined distance matrices using Ward's minimum variance method⁷⁴ as implemented in SciPy. The emergent metric was resampled 500 times by swapping protein labels to compute bootstrap support at each bifurcation node. Unrooted trees with scaled distances were drawn and annotated with domain architectures of individual E3 leaves using iTol⁷⁵. The final tree was parsed at tree cutoff $h = 0.25$ to produce optimal emergent clusters (E3 families). Each family was manually analyzed for shared sequence and domain-architectural features to identify subfamilies and outliers.

Gene essentiality and CRISPR-Cas9 dropout screens

A pooled multiplexed CRISPR-Cas9 library targeting 822 genes of the human ubiquitin-proteasome system (UPS) was generated according to previously reported methods^{76,77}. Each UPS gene was targeted by 2 gRNAs, each expressed from a human 7SK promoter of a dual gRNA-containing lentiviral plasmid, co-expressing puromycin. The second gRNA cassette expressed an AAVS1-targeting gRNA as a neutral control. NGS verified successful cloning, and lentiviral particles were obtained from VectorBuilder.

Library preparation was done using RPE1 cells obtained from ATCC stably expressing Cas9. Cell lines were routinely tested for Mycoplasma contamination. They were transduced with infectious lentiviral particles of the UPS library at 50X coverage and a multiplicity of infection (MOI) of 0.5 to ensure predominantly single integrations per cell. Following infection, cells were subjected to two rounds of puromycin selection (each round for 48 h). Throughout the screen, cells were maintained at a minimum coverage of 50-fold per sgRNA to preserve representation of the library. After ~ 10 cell doublings (≈ 14 days), cells were harvested and genomic DNA was extracted following previously established protocols⁷⁶.

NGS preparation was performed in two steps. First, 900 μg (50X coverage) of genomic DNA was used in a 50 μL PCR-1 reaction with primers amplifying both sgRNA library cassettes. PCR-1 had the following composition: 25 μL of NEBNext® Ultra™ II Q5® Master Mix (NEB,

M0544L), 2 μ L of primer mix (0.4 μ M each, IDT), 2.25 μ g genomic DNA, and H₂O to 50 μ L. PCR cycling conditions were as follows: 1 min at 98 °C, followed by 15 s at 98 °C, 20 s at 65 °C, and 60 s at 72 °C for 20 cycles, with a final extension for 60 s at 72 °C using an Eppendorf Mastercycler® X50. PCR-1 products from the same sample were then pooled. For each sample, a second PCR reaction was performed to add P5 and P7 Illumina adaptors, including demultiplexing barcodes. PCR-2 primers also included eight different stagger lengths, all of which were pooled in equimolar amounts. PCR-2 reaction mix was as follows: 25 μ L of NEBNext® Ultra™ II Q5® Master Mix (NEB, M0544L), 1 μ L of P5 primer and 1 μ L P7 primer, including barcodes (0.2 μ M each, SIGMA), 10% input from PCR1 and H₂O to 50 μ L. PCR cycling conditions were as follows: 1 min at 98 °C, followed by 15 s at 98 °C, 20 s at 68 °C, and 30 s at 72 °C for 11 cycles, with a final extension for 60 s at 72 °C using an Eppendorf Mastercycler® X50. Barcoded PCR-2 products were then pooled and gel-purified according to the Zymoclean Gel DNA Recovery Kit (Zymo, D4001). NGS Sequencing was performed at Genewiz. Read Counting of the library and both biological replicates of end time points were done with ReCo⁷⁸. For each replicate, we performed $\log_2(CPM + 1)$ transformation for each gRNA to obtain normalized read counts.

$$NRC_i = \log_2 \left(\frac{gRNA_i \times n \times 100}{\sum_{i=1}^n gRNA_i} + 1 \right) \quad (13)$$

Where $gRNA_i$ = RC of gRNA targeting gene i ; n = total number of guides in the library. We computed log-fold changes (LFCs) as the differences between the NRC_i of both gRNAs. Gene-level LFCs were obtained by averaging the LFC values of guides targeting the same gene. Using Limma⁷⁹, we fit a linear model for each gene, considering the different gRNAs as technical replicates and the biological replicates as blocks. We used the `Correlation` function to estimate the intra-block correlation. Then, we calculated p values using empirical Bayes (eBayes) followed by correction for multiple testing using a Benjamini-Hochberg correction. We defined genes with significant contribution to cell fitness if their corresponding $|LFC_i| \geq 1$ and $FDR \leq 0.05$.

Identifying generic and specific functions of the E3 ligome

GO enrichment analysis for E3 ligases corresponding to individual 13 families was performed using Metascape⁸⁰, which implements a hierarchical clustering approach based on κ -similarity ≥ 0.3 ⁸⁰. The resulting networks of GO terms at the biological process, cellular component, and molecular function ontologies were rendered using Cytoscape. Nodes were colored and drawn as pie charts to reflect E3 family contribution (number of proteins) and enrichment. Individual GO terms were considered significantly enriched within a ligase family if the enrichment factor, $C_{obs}/C_{exp} \geq 2$, a minimum of 3 proteins corresponding to the family are annotated explicitly with the corresponding GO terms, and a p value ≤ 0.01 . Within each resulting GO cluster, the GO term with the lowest p value was selected as the cluster label for visualization. Heatmaps showing the enriched GO clusters for each family were drawn to highlight the functional specialization of individual E3 families. We ensured high-confidence evidence codes (Experimental, Phylogenetic, and Computational evidence) for enriched GO terms. GO enrichment analysis for essential E3 ligases (background all E3 ligases present in UPS CRISPR dataset) was performed using ShinyGO v0.82⁸¹, employing a Hypergeometric test with Benjamini-Hochberg correction, $FDR \leq 0.05$.

Cell culture, sample preparation, mass spectrometry data collection, and analysis

HeLa cells were either challenged with (i) global starvation (EBSS treatment triggering nutrient starvation and activating autophagy) or (ii) DNA damage (CPT treatment inhibiting topoisomerase I to induce single-ended DNA double-strand breaks). Cells (ATCC) were cultured

in DMEM supplemented with 10% FBS and 100 I.U./mL penicillin-streptomycin in a 5% CO₂ atmosphere at 37 °C. Starvation response was induced by incubating cells in EBSS medium (Gibco) for 6h. For triggering DNA damage response, cells were treated with 100 nM Camptothecin for 4h before sample preparation. Cells were regularly tested for Mycoplasma contamination using the Mycoplasma PCR Detection Kit (Sigma). Cells were lysed in Lysis Buffer (2% SDS, 50mM Tris pH 8.5, 10 mM TCEP, 40 mM CAA, supplemented with protease inhibitor cocktail and phosphatase inhibitors), sonicated at 4 °C in a Bioruptor Pico2 (30/30, 10 cycles), and boiled at 95 °C. Proteins were precipitated using methanol-chloroform and digested with 1:50 w/w LysC (Wako Chemicals) and 1:100 w/w trypsin (Promega) overnight at 37 °C. De-salted peptides were dried and resuspended in TMT-labeling buffer (200 mM EPPS pH 8.2, 20% acetonitrile) before being subjected to TMT labeling with a 1:2.5 peptide TMT ratio (w/w) for 1 h at room temperature. The labeling reaction was quenched by the addition of 0.5% hydroxylamine final concentration. Successful TMT labeling was verified by mixing equimolar ratios of peptides and subjecting the mix to single-shot LC-MS/MS analysis. Peptides were fractionated using high-pH liquid chromatography on a micro-flow HPLC (Dionex U3000 RSLC, Thermo Scientific). Pooled fractions were dried in a vacuum concentrator and resuspended in 2% ACN, 0.1% TFA for LC-MS analysis.

Tryptic peptides were analyzed on an Orbitrap Fusion Lumos coupled to an easy nLC 1200 (ThermoFisher Scientific) using a 35 cm long, 75 μ m ID fused-silica column packed in-house with 1.9 μ m C18 particles (Reprosil pur, Dr. Maisch), and kept at 50 °C using an integrated column oven (Sonation). Peptides were eluted from 8% to 28% Buffer B (80% ACN, 0.1% FA) over 75 minutes, followed by a step-wise increase to 90% Buffer B in 21 minutes, which was held for another 9 minutes. A synchronous precursor selection (SPS) multi-notch MS3 method was used⁸². Full scan MS spectra (350-1400 m/z): resolution of 120,000 at m/z 200, maximum injection time of 100 ms, and AGC target value of 4×10^5 . MS2 scans (of precursors with charge state between 2-6): maximum injection time of 50 ms, AGC target value of 15000, CID fragmentation with a normalized collision energy (NCE) of 35%. MS3: 10 most intense MS2 fragment ions with an isolation window of 0.7 Th (MS) and 2 m/z (MS2), HCD fragmentation, NCE 50%, resolution of 50000 at m/z 200, scan range of 100-500 m/z, AGC target value of 150000, and a maximum injection time of 86 ms. Repeated sequencing of already acquired precursors was limited by setting a dynamic exclusion of 60 seconds and 7 ppm, and advanced peak determination was deactivated. All spectra were acquired in centroid mode.

MS raw data were analyzed using FragPipe v2.1.1, with MSFragger v.4.0⁸³ and Philosopher v.5.1.0⁸⁴. Acquired spectra were searched against the human reference proteome (Taxonomy ID 9606) downloaded from UniProt (07/25/2024; 20,418 entries) with a precursor mass tolerance of 20 ppm and fragment mass tolerance of 20 ppm. Identifications were filtered to obtain false discovery rates (FDR) below 1% for both peptide spectrum matches (minimum peptide length of 7) and proteins using a target-decoy strategy. For all searches, carbamidomethylated cysteine was set as a fixed modification and oxidation of methionine and N-terminal protein acetylation as variable modifications with allowing up to 3 modifications per peptide. Strict trypsin cleavage was set as the protein digestion rule. Label-free quantification was performed using IonQuant v.1.10.27⁸⁵. The differential expression analysis was performed within FragPipe analyst⁸⁶ using Limma⁷⁹, and derived p values were corrected for multiple comparisons using Benjamini-Hochberg. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier [PXD067015](https://doi.org/10.1038/s41467-025-67450-9).

Integrating PPI and ESI datasets

To identify E3 ligases likely functioning in complex mode, we combined data from PDB (<https://www.rcsb.org/>) and IntAct⁸⁷. Using the refined lists of proteins corresponding to the E3 ligome, E1, E2,

adaptors, receptors, and scaffold proteins (UbiHub and manually curated lists), we retrieved all the PDB structures (as of Jul. 2025) involving E3-adaptors, E3-receptors, and E3-scaffold protein complexes. Additionally, PPIs obtained between E3-adaptor, E3-receptor, and E3-scaffold proteins filtered ("experimentally validated" PPIs, MI:0045 ≥ 0.5) for high confidence entries. E3s interacting, or in a resolved structure, with at least one receptor, adaptor, or scaffold protein, were re-annotated as complex E3s. To assemble the E3-substrate interaction map, we integrated multiple data sources, including experimentally validated enzyme-substrate interactions (known ESIs) from UbiNet 2.0⁵⁷ and UbiBrowser⁵⁸, a set of predicted ESIs from UbiBrowser (top 1% of predictions), physically interacting protein pairs (PPIs) from the IntAct database (mapped PPIs), and indirect PPIs involving ligases and potential substrates mediated by adaptor, receptor, or scaffold proteins from IntAct (indirect PPIs).

To obtain a cut-off for filtering PPIs, we first detected a subset of the well-known ESIs (experimentally curated from UbiNet; blue) overlapping with the direct PPIs obtained from IntAct. The PSI-MI confidence scores for these overlapping ESIs ($n = 239$) display a median value of 0.55. This encouraged us to use a threshold value ≥ 0.5 for filtering PPIs. This cutoff strikes a balance between including enough interactions for meaningful analysis while excluding lower-confidence edges that disproportionately contribute noise. Further, major PPI resources (IntAct/IMEx, Reactome, Open Targets) typically use 0.4–0.45 as a medium-confidence baseline. Rounding that to 0.5 provides us with a threshold that sits squarely in the "medium-to-high" confidence range, increasing the proportion of PPIs detected using diverse approaches⁵⁹.

Known ESIs and the PPIs dataset were enriched using substrates detected mainly by pull-down experiments, followed by two-hybrid techniques. A map of the ubiquitinated human proteome was obtained by cross-checking the ubiquitination status and mapping ubiquitination sites for each identified substrate from dbPTM⁹⁰ and PhosphoSitePlus⁹¹. All substrates were categorized based on their interactions with E3 ligases: those paired with a single, unique E3 ligase were classified as E3-specific; those associated with multiple E3 ligases from the same family were designated as family-specific; and those linked to two or more E3 ligases from different families were labeled promiscuous.

Mapping small molecule interaction data

A unified dataset of E3 handles (corresponding to all publicly documented PROTACs) and E3 binders targeting specific E3 ligases, adaptors, receptors, and scaffold proteins was obtained by combining data from PROTACpedia (<https://protacpedia.weizmann.ac.il>), and PROTAC-DB 3.0⁹² and ChEMBL v34⁹³. All small molecules were uniquely identified by their chemical structure, represented using the canonical SMILES format, and mapped to their target proteins and E3 families. Information from ChEMBL v34 was gathered using an SQL query combining compound data, experimental data, and target protein information, and filtered using data from binding assays using p-ChEMBL values.

$$p - \text{ChEMBL} = \log_{10}(\text{Activity}), \quad (14)$$

where activity is given by IC₅₀, EC₅₀, K_i, K_d, or some measure of potency in molar units. It enables comparison across different bioactivity types. p-ChEMBL ≥ 6.0 is a commonly used threshold in early-stage screening as a baseline for biologically relevant activity ($\leq 1 \times 10^{-6}$ M)⁹⁴. 2048-bit Morgan fingerprint⁹⁵ for each small molecule was obtained using RDKit (2048-bit array, radius = 3, <http://www.rdkit.org>). Dimensionality reduction was performed using PCA, t-SNE, and UMAP using the Python Scikit-learn package (default parameters) and visualized along a 2D subspace, projections showing the highest variance. Further, clustering of small molecules was performed by fast search and

identification of local density peaks⁴⁴ using a scaled Euclidean distance measure defined on the corresponding normalized 2D subspaces. Using threshold values for local density ($\rho = 20$) and nearest neighboring peak ($\delta = 0.1$) demonstrated the superiority of UMAP embedding and resulted in 20 small molecule clusters (indexed by i , targeting proteins, j , with population $n_i = \sum_j n_{ij}$). For each cluster, i , representative small molecules were automatically determined from local density peaks, and their distinct set of targeted proteins j from the E3 ligome were also mapped. We then quantified the binding likelihood of individual clusters to target distinct proteins by computing log-transformed propensities, LP _{ij} .

$$\text{LP}_{ij} = \log_2 \left[\left(\frac{n_{ij}}{\sum_j n_{ij}} \right) \times \left(\frac{\sum_i \sum_j n_{ij}}{\sum_i n_i} \right) \right] \quad (15)$$

Compound-E3 interaction structures were retrieved from the PDB by converting compound SMILES to InChI keys with RDKit, querying the EBI UniChem API for all small-molecule PDB entries, and merging them with target-protein PDB entries from our dataset, retaining only entries containing both. E3 pocketome data from ELIOT⁴⁵ was then mapped to these compound-E3 pairs.

Reporting summary

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

Data availability

All data supporting the findings are provided in the paper and the Supplementary Information File. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier **PXD067015**. All data generated, including source data for all figures, and additional Supplementary data files (1–10), are deposited in Zenodo (<https://doi.org/10.5281/zenodo.17771730>) under a CC BY-NC-SA 4.0 license. A website on "The Human E3 Ligome" to browse the processed data is also hosted on (<https://e3-ligome-91adc4.gitlab.io/index.html>) under CC BY-NC-SA 4.0 license.

Code availability

The codes used to develop and analyse "The Human E3 ligome" and perform this study are deposited in the public Zenodo repository (<https://doi.org/10.5281/zenodo.17771730>) under a CC BY-NC-SA 4.0 license.

References

- Collins, G. A. & Goldberg, A. L. The logic of the 26S proteasome. *Cell* **169**, 792–806 (2017).
- Dikic, I. Proteasomal and autophagic degradation systems. *Annu. Rev. Biochem.* **86**, 193–224 (2017).
- Berndsen, C. E. & Wolberger, C. New insights into ubiquitin E3 ligase mechanism. *Nat. Struct. Mol. Biol.* **21**, 301–307 (2014).
- Komander, D. & Rape, M. The ubiquitin code. *Annu. Rev. Biochem.* **81**, 203–229 (2012).
- Zheng, N. & Shabek, N. Ubiquitin ligases: structure, function, and regulation. *Annu. Rev. Biochem.* **86**, 129–157 (2017).
- Xie, Y. et al. Posttranslational modification of autophagy-related proteins in macroautophagy. *Autophagy* **11**, 28–45 (2014).
- Cristiani, A., Dutta, A., Poveda-Cuevas, S. A., Kern, A. & Bhaskara, R. M. Identification of potential selective autophagy receptors from protein-content profiling of autophagosomes. *J. Cell. Biochem.* **n/a** (2023).
- González, A. et al. Ubiquitination regulates ER-phagy and remodeling of endoplasmic reticulum. *Nature* **618**, 394–401 (2023).
- Foronda, H. et al. Heteromeric clusters of ubiquitinated ER-shaping proteins drive ER-phagy. *Nature* **618**, 402–410 (2023).

10. Mehnert, M., Sommer, T. & Jarosch, E. ERAD ubiquitin ligases. *BioEssays* **32**, 905–913 (2010).
11. Serrano, I., Campos, L. & Rivas, S. Roles of e3 ubiquitin-ligases in nuclear protein homeostasis during plant stress responses. *Front. Plant Sci.* **9** (2018).
12. Mallery, D. L. et al. Antibodies mediate intracellular immunity through tripartite motif-containing 21 (trim21). *Proc. Natl. Acad. Sci.* **107**, 19985–19990 (2010).
13. Vince, J. E. et al. Iap antagonists target ciap1 to induce tnfr α -dependent apoptosis. *Cell* **131**, 682–693 (2007).
14. Cruz Walma, D. A., Chen, Z., Bullock, A. N. & Yamada, K. M. Ubiquitin ligases: guardians of mammalian development. *Nat. Rev. Mol. Cell Biol.* **23**, 350–367 (2022).
15. Morreale, F. E. & Walden, H. Types of ubiquitin ligases. *Cell* **165**, 248–248 (2016).
16. Schapira, M., Calabrese, M. F., Bullock, A. N. & Crews, C. M. Targeted protein degradation: expanding the toolbox. *Nat. Rev. Drug Discov.* **18**, 949–963 (2019).
17. Li, W. et al. Genome-wide and functional annotation of human E3 ubiquitin ligases identifies Mulan, a mitochondrial E3 that regulates the organelle's dynamics and signaling. *PLOS ONE* **3**, 1–14 (2008).
18. Deshaies, R. J. & Joazeiro, C. A. Ring domain e3 ubiquitin ligases. *Annu. Rev. Biochem.* **78**, 399–434 (2009).
19. Medvar, B., Raghuram, V., Pisitkun, T., Sarkar, A. & Knepper, M. A. Comprehensive database of human E3 ubiquitin ligases: application to aquaporin-2 regulation. *Physiol. Genom.* **48**, 502–512 (2016).
20. Liu, Y. et al. Expanding proteome universe of E3 ligases. *Nat. Commun.* **14**, 6509 (2023).
21. Harper, J. W. & Schulman, B. A. Cullin-ring ubiquitin ligase regulatory circuits: a quarter century beyond the f-box hypothesis. *Annu. Rev. Biochem.* **90**, 403–429 (2021).
22. Gates, A. J., Wood, I. B., Hetrick, W. P. & Ahn, Y.-Y. Element-centric clustering comparison unifies overlaps and hierarchy. *Sci. Rep.* **9**, 8574 (2019).
23. Cunningham, C. N. et al. Usp30 and parkin homeostatically regulate atypical ubiquitin chains on mitochondria. *Nat. Cell Biol.* **17**, 160–169 (2015).
24. Suryo Rahmanto, A. et al. K6-linked ubiquitylation marks formaldehyde-induced RNA-protein crosslinks for resolution. *Mol. Cell* **83**, 4272–4289.e10 (2023).
25. Yang, B. et al. Rnf144a promotes antiviral responses by modulating sting ubiquitination. *EMBO Rep.* **24**, e57528 (2023).
26. Gack, M. U. et al. Trim25 ring-finger E3 ubiquitin ligase is essential for rig-i-mediated antiviral activity. *Nature* **446**, 916–920 (2007).
27. Lang, X. et al. Trim65-catalyzed ubiquitination is essential for MDA5-mediated antiviral innate immunity. *J. Exp. Med.* **214**, 459–473 (2016).
28. Thoreen, C. C. et al. A unifying model for mTORC1-mediated regulation of mRNA translation. *Nature* **485**, 109–113 (2012).
29. Asrih, M. et al. Overview of growth differentiation factor 15 in metabolic syndrome. *J. Cell. Mol. Med.* **27**, 1157–1167 (2023).
30. Yang, Y. et al. Trim27 cooperates with Stk38l to inhibit ULK1-mediated autophagy and promote tumorigenesis. *EMBO J.* **41**, e109777 (2022).
31. Garcia-Garcia, J. et al. Trim27 is an autophagy substrate facilitating mitochondria clustering and mitophagy via phosphorylated tbk1. *FEBS J.* **290**, 1096–1116 (2022).
32. Overå, K. S. et al. Trim32, but not its muscular dystrophy-associated mutant, positively regulates and is targeted to autophagic degradation by p62/sqstm1. *J. Cell Sci.* **132**, jcs236596 (2019).
33. Pommier, Y. Topoisomerase I inhibitors: camptothecins and beyond. *Nat. Rev. Cancer* **6**, 789–802 (2006).
34. Doil, C. et al. Rnf168 binds and amplifies ubiquitin conjugates on damaged chromosomes to allow accumulation of repair proteins. *Cell* **136**, 435–446 (2009).
35. Kelliher, J., Ghosal, G. & Leung, J. W. C. New answers to the old riddle: Rnf168 and the DNA damage response pathway. *FEBS J.* **289**, 2467–2480 (2021).
36. Zhang, H. et al. A cell cycle-dependent BRCA1-Uhrf1 cascade regulates DNA double-strand break repair pathway choice. *Nat. Commun.* **7**, 10201 (2016).
37. Yang, Y. et al. PcnA-binding activity separates Rnf168 functions in DNA replication and DNA double-stranded break signaling. *Nucleic Acids Res.* **52**, 13019–13035 (2024).
38. Liu, L. et al. UbiHub: a data hub for the explorers of ubiquitination pathways. *Bioinformatics* **35**, 2882–2884 (2019).
39. Han, Y., Lee, H., Park, J. C. & Yi, G.-S. E3net: a system for exploring e3-mediated regulatory networks of cellular functions. *Mol. Cell. Proteom.* **11**, O111.014076 (2012).
40. Chen, Y.-L. et al. Smad6 inhibits the transcriptional activity of tbx6 by mediating its degradation. *J. Biol. Chem.* **284**, 23481–23490 (2009).
41. Yonashiro, R. et al. A novel mitochondrial ubiquitin ligase plays a critical role in mitochondrial dynamics. *EMBO J.* **25**, 3618–3626 (2006).
42. Lévy, F. et al. Ubiquitylation of a melanosomal protein by HECT-E3 ligases serves as sorting signal for lysosomal degradation. *Mol. Biol. Cell* **16**, 1777–1787 (2005).
43. Guicciardi, M. E. et al. Cellular inhibitor of apoptosis (ciap)-mediated ubiquitination of phosphofurin acidic cluster sorting protein 2 (pacs-2) negatively regulates tumor necrosis factor-related apoptosis-inducing ligand (trail) cytotoxicity. *PLoS ONE* **9**, e92124 (2014).
44. Rodriguez, A. & Laio, A. Clustering by fast search and find of density peaks. *Science* **344**, 1492–1496 (2014).
45. Palomba, T., Baroni, M., Cross, S., Cruciani, G. & Siragusa, L. Eliot: A platform to navigate the e3 pocketome and aid the design of new PROTACS. *Chem. Biol. Drug Des.* **101**, 69–86 (2022).
46. Sikosek, T. & Chan, H. S. Biophysics of protein evolution and evolutionary protein biophysics. *J. R. Soc. Interface* **11**, 20140419 (2014).
47. Rezende, P. M., Xavier, J. S., Ascher, D. B., Fernandes, G. R. & Pires, D. E. V. Evaluating hierarchical machine learning approaches to classify biological databases. *Brief. Bioinform.* **23**, bbac216 (2022).
48. Ashburner, M. et al. Gene ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
49. Blanco-Mallo, E., Morán-Fernández, L., Remeseiro, B. & Bolón-Canedo, V. Do all roads lead to rome? studying distance measures in the context of machine learning. *Pattern Recognit.* **141**, 109646 (2023).
50. Spratt, D. E., Walden, H. & Shaw, G. S. Rbr e3 ubiquitin ligases: new structures, new insights, new questions. *Biochemical J.* **458**, 421–437 (2014).
51. Shah, S. S. & Kumar, S. Adaptors as the regulators of HECT ubiquitin ligases. *Cell Death Differ.* **28**, 455–472 (2021).
52. Ciani, B., Layfield, R., Cavey, J. R., Sheppard, P. W. & Searle, M. S. Structure of the ubiquitin-associated domain of p62 (sqstm1) and implications for mutations that cause Paget's disease of bone. *J. Biol. Chem.* **278**, 37409–37412 (2003).
53. Mancini, M., Magnani, E., Macchi, F. & Bonapace, I. M. The multifunctionality of Uhrf1: epigenome maintenance and preservation of genome integrity. *Nucleic Acids Res.* **49**, 6053–6068 (2021).
54. Rodríguez-Gimeno, A. & Galdeano, C. Drug discovery approaches to target E3 ligases. *ChemBio. Chem.* **26**, e202400656 (2024).
55. Schneckloth, A. R., Pucheault, M., Tae, H. S. & Crews, C. M. Targeted intracellular protein degradation induced by a small molecule: en route to chemical proteomics. *Bioorg. Medicinal Chem. Lett.* **18**, 5904–5908 (2008).
56. Li, Y. et al. Discovery of md-224 as a first-in-class, highly potent, and efficacious proteolysis targeting chimera murine double minute 2 degrader capable of achieving complete and durable tumor regression. *J. Medicinal Chem.* **62**, 448–466 (2018).

57. Li, Z. et al. UbiNet 2.0: a verified, classified, annotated and updated database of E3 ubiquitin ligase–substrate interactions. *Database* **2021**, baab010 (2021).
58. Consortium, T. U. UniProt: the universal protein knowledgebase in 2023. *Nucleic Acids Res.* **51**, D523–D531 (2022).
59. Oughtred, R. et al. The biogrid database: a comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Sci.* **30**, 187–200 (2021).
60. Lex, A., Gehlenborg, N., Strobel, H., Vuilleumot, R. & Pfister, H. Upset: visualization of intersecting sets. *IEEE Trans. Vis. Comput. Graph.* **20**, 1983–1992 (2014).
61. Paysan-Lafosse, T. et al. Interpro in 2022. *Nucleic Acids Res.* **51**, D418–D427 (2023).
62. Martin, J., Anamika, K. & Srinivasan, N. Classification of protein kinases on the basis of both kinase and non-kinase regions. *PLoS one* **5**, e12460 (2010).
63. Bhaskara, R. M. et al. The relationship between classification of multi-domain proteins using an alignment-free approach and their functions: a case study with immunoglobulins. *Mol. Biosyst.* **10**, 1082–1093 (2014).
64. Zhang, J. & Kumar, S. Detection of convergent and parallel evolution at the amino acid sequence level. *Mol. Biol. Evol.* **14**, 527–536 (1997).
65. Levandowsky, M. & Winter, D. Distance between sets. *Nature* **234**, 34–35 (1971).
66. Lin, K., Zhu, L. & Zhang, D.-Y. An initial strategy for comparing proteins at the domain architecture level. *Bioinformatics* **22**, 2081–2086 (2006).
67. Goodman, L. A. & Kruskal, W. H. Measures of association for cross classifications. ii: Further discussion and references. *J. Am. Stat. Assoc.* **54**, 123–163 (1959).
68. Varadi, M. et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, D439–D444 (2021).
69. Zhang, Y. & Skolnick, J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* **33**, 2302–2309 (2005).
70. Zhao, C. & Wang, Z. Gogo: an improved algorithm to measure the semantic similarity between gene ontology terms. *Sci. Rep.* **8**, 1–10 (2018).
71. Consortium, T. G. O. et al. The gene ontology knowledgebase in 2023. *Genetics* **224**, iyad031 (2023).
72. Wang, J. Z., Du, Z., Payattakool, R., Yu, P. S. & Chen, C.-F. A new method to measure the semantic similarity of go terms. *Bioinformatics* **23**, 1274–1281 (2007).
73. Uhlén, M. et al. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
74. Ward, J. H. Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* **58**, 236 (1963).
75. Letunic, I. & Bork, P. Interactive tree of life (itol) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res.* **52**, W78–W82 (2024).
76. Diehl, V. et al. Minimized combinatorial CRISPR screens identify genetic interactions in autophagy. *Nucleic Acids Res.* **49**, 5684–5704 (2021).
77. Wegner, M. et al. Circular synthesized CRISPR/cas gRNAs for functional interrogations in the coding and noncoding genome. *eLife* **8**, e42549 (2019).
78. Wegner, M. & Kaulich, M. Reco: automated NGS read-counting of single and combinatorial CRISPR gRNAs. *Bioinformatics* **39**, btad448 (2023).
79. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47–e47 (2015).
80. Zhou, Y. et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **10**, 1523 (2019).
81. Ge, S. X., Jung, D. & Yao, R. Shinygo: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics* **36**, 2628–2629 (2019).
82. McAlister, G. C. et al. Multinotch ms3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes. *Anal. Chem.* **86**, 7150–7158 (2014).
83. Kong, A. T., Leprevost, F. V., Avtonomov, D. M., Mellacheruvu, D. & Nesvizhskii, A. I. Msfragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nat. Methods* **14**, 513–520 (2017).
84. da Veiga Leprevost, F. et al. Philosopher: a versatile toolkit for shotgun proteomics data analysis. *Nat. Methods* **17**, 869–870 (2020).
85. Yu, F., Haynes, S. E. & Nesvizhskii, A. I. Ionquant enables accurate and sensitive label-free quantification with FDR-controlled match-between-runs. *Mol. Cell. Proteom.* **20**, 100077 (2021).
86. Hsiao, Y. et al. Analysis and visualization of quantitative proteomics data using fragpipe-analyst. *J. Proteome Res.* **23**, 4303–4315 (2024).
87. Orchard, S. et al. The mintact project-intact as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res.* **42**, D358–D363 (2013).
88. Wang, X. et al. Ubibrowser 2.0: a comprehensive resource for proteome-wide known and predicted ubiquitin ligase/deubiquitinase–substrate interactions in eukaryotic species. *Nucleic Acids Res.* **50**, D719–D728 (2021).
89. Porras, P. et al. Towards a unified open access dataset of molecular interactions. *Nat. Commun.* **11**, 6144 (2020).
90. Chung, C.-R. et al. dbptm 2025 update: comprehensive integration of ptms and proteomic data for advanced insights into cancer research. *Nucleic Acids Res.* **53**, D377–D386 (2024).
91. Hornbeck, P. V. et al. Phosphositeplus, 2014: mutations, ptms and recalibrations. *Nucleic Acids Res.* **43**, D512–D520 (2014).
92. Ge, J. et al. Protac-db 3.0: an updated database of protacs with extended pharmacokinetic parameters. *Nucleic Acids Res.* **53**, D1510–D1515 (2024).
93. Zdrazil, B. et al. The ChEMBL database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods. *Nucleic Acids Res.* **52**, D1180–D1192 (2023).
94. Smajić, A., Rami, I., Sosnin, S. & Ecker, G. F. Identifying differences in the performance of machine learning models for off-targets trained on publicly available and proprietary data sets. *Chem. Res. Toxicol.* **36**, 1300–1312 (2023).
95. Morgan, H. L. The generation of a unique machine description for chemical structures—a technique developed at Chemical Abstracts Service. *J. Chem. Doc.* **5**, 107–113 (1965).

Acknowledgements

We thank Stefan Knapp, Gerhard Hummer, Marcel Heinz, and Matthew Shapira, along with all members of the PROXIDRUGS consortium, for their support and constructive discussion. We thank David Krause for system administration and the Center for Supercomputing, Goethe University Frankfurt (GUF), for computing time on the Goethe-HLR cluster. We thank all members of the Centre for Functional Proteomics GUF, in particular, Martin Adrian-Allgood and Bhavesh Parmar. We thank the German Research Foundation, DFG, for funding the LC-MS system (easy nLC1200, Orbitrap Fusion LUMOS) used in this study (FuGG Project-ID: 403765277). This work is primarily funded by PROXIDRUGS, InnoDATA 1.0, and 2.0 projects (O3ZU1109KA and O3ZU2109JA), which are part of the “Clusters4Future” initiative funded by the Federal Ministry of Education and Research (BMBF) to R.M.B. We acknowledge financial support from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No. 875510 to A.C. and the Deutsche

Forschungsgemeinschaft Project-ID 259130777-SFB1177 on Selective Autophagy to I.D. and R.M.B.

Author contributions

Conceptualization: R.M.B. Methodology: A.D., A.C., S.V.N., and R.M.B. Software: A.D., A.C., S.V.N., and R.M.B. Validation: A.D., A.C., S.V.N., Y.M., B.M., C.T., C.B., V.J.S., and T.M. Formal analysis: A.D., A.C., S.V.N., and R.M.B. Investigation: A.D., A.C., S.V.N., J.E., Y.M., B.M., C.T., C.B., V.J.S., T.M., K.H., I.D., M.K., and R.M.B. Resources: I.D., M.K., and R.M.B. Data Curation: A.D., A.C., S.V.N., T.M., Y.M., I.D., M.K., and R.M.B. Writing—original draft: A.D., A.C., S.V.N., and R.M.B. Writing—review & editing: A.D., A.C., S.V.N., V.J.S., T.M., I.D., M.K., and R.M.B. Visualization: A.D., A.C., S.V.N., and R.M.B. Supervision: R.M.B. Funding acquisition: R.M.B.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

R.M.B., M.K., K.H., and I.D. are head scientists at the Frankfurt Competence Center for Emerging Therapeutics (FCET), Goethe Center for (high) technology (Go4Tec), Goethe University, Frankfurt am Main, Germany. The remaining authors declare no competing interests. This manuscript reflects the views of the authors, and neither IMI nor the European Union, EFPIA, nor any associated partners are liable for any use that may be made of the information contained herein. I.D. is a founder/shareholder of Vivlion GmbH and a member of its scientific advisory board. I.D. is also a member of the scientific advisory board of the Boehringer Ingelheim Foundation, the expert committee (for international research leader grants) of the Novo Nordisk Foundation, and the advisory board of Cell and Molecular Cell. I.D. was a founder and consultant of Caraway Therapeutics Inc. M.K. is a co-founder, shareholder, and chief officer of Vivlion GmbH.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-67450-9>.

Correspondence and requests for materials should be addressed to Ramachandra M. Bhaskara.

Peer review information *Nature Communications* thanks Chen Ding, Fuchu He, and the other, anonymous, reviewers for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025