HHS Public Access

Author manuscript

Nat Biotechnol. Author manuscript; available in PMC 2019 February 20.

Published in final edited form as:

Nat Biotechnol. 2018 October ; 36(9): 880–887. doi:10.1038/nbt.4201.

Efficient proximity labeling in living cells and organisms with TurboID

Tess C. Branon[1,2,3,4], **Justin A. Bosch**[5], **Ariana D. Sanchez**[4], **Namrata D. Udeshi**[6], **Tanya Svinkina**[6], **Steven A. Carr**[6], **Jessica L. Feldman**[4], **Norbert Perrimon**[5,7], and **Alice Y. Ting**[*], [1,2,3,4,7]

[1]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

[2]Departments of Genetics, Stanford University, Stanford, California, USA

[3]Departments of Chemistry, Stanford University, Stanford, California, USA

[4]Department of Biology, Stanford University, Stanford, California, USA

[5]Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA

[6]Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA

[7]Howard Hughes Medical Institute, Boston, Massachusetts, USA

[8]Chan Zuckerberg Biohub, San Francisco, California, USA

Abstract

Protein interaction networks and protein compartmentalization underlie all signaling and regulatory processes in cells. Enzyme-catalyzed proximity labeling (PL) has emerged as a new approach to study the spatial and interaction characteristics of proteins in living cells. However, current PL methods require over 18 hour labeling times or utilize chemicals with limited cell permeability or high toxicity. We used yeast display-based directed evolution to engineer two promiscuous mutants of biotin ligase, TurboID and miniTurbo, which catalyze PL with much greater efficiency than BioID or BioID2, and enable 10-minute PL in cells with non-toxic and easily deliverable biotin. Furthermore, TurboID extends biotin-based PL to flies and worms.

Enzyme-catalyzed proximity labeling (PL) is an alternative to immunoprecipitation and biochemical fractionation for proteomic analysis of macromolecular complexes, organelles,

[*]Corresponding author, ayting@stanford.edu.

Author contributions

T.C.B. and A.Y.T. designed the research and analyzed all the data except those noted. T.C.B. performed all experiments except those noted. T.C.B., A.Y.T., N.D.U. and S.A.C. designed the proteomics experiments. T.C.B. prepared the proteomic samples. N.D.U. and T.S. processed the proteomic samples and performed mass spectrometry. J.A.B. performed *D. melanogaster* experiments. J.A.B. and N.P. analyzed *D. melanogaster* data. T.C.B., A.Y.T., A.D.S., and J.L.F. designed the *C. elegans* experiments. A.D.S. performed *C. elegans* experiments. A.D.S. and J.L.F. analyzed *C. elegans* data.

Competing financial interests

A.Y.T. and T.C.B. have filed a patent application covering some aspects of this work.

and protein interaction networks[1]. In PL, a promiscuous labeling enzyme is targeted by genetic fusion to a specific protein or subcellular compartment. Addition of a small molecule substrate, such as biotin, initiates covalent tagging of endogenous proteins within a few nanometers of the promiscuous enzyme (Figure 1a). Subsequently, the biotinylated proteins are harvested using streptavidin-coated beads and identified by mass spectrometry (MS).

Two enzymes are commonly used for PL: APEX2, an engineered soybean ascorbate peroxidase[2,3]; and BioID, a promiscuous mutant of *E. coli* biotin ligase[4,5]. The advantages of APEX2 are its speed – proximal proteins can be tagged in 1 minute or less[6,7] – and versatility, as APEX2 also captures endogenous RNAs[8] and generates contrast for electron microscopy[9]. However, APEX labeling requires the use of H_2O_2, which is toxic to living samples. By contrast, BioID labeling is simple and non-toxic: only biotin needs to be supplied to initiate tagging. This feature has resulted in >100 applications of BioID since its introduction 5 years ago, in cultured mammalian cells[5,10,11], plant protoplasts[12], parasites[13–21], slime mold[22,23], mouse[24], and yeast[25].

The major disadvantage of BioID, however, is its slow kinetics, which necessitates labeling with biotin for 18-24 hours (and sometimes much longer[24]) to produce sufficient biotinylated material for proteomic analysis. This precludes the use of BioID for studying dynamic processes that occur on the timescale of minutes or even a few hours. Furthermore, the low catalytic activity makes BioID difficult or impossible to apply in some contexts - such as in worms, flies, or the ER lumen of cultured mammalian cells. Recently, new promiscuous biotin ligase variants, BioID2[26] and BASU[27], have been reported, but the former still requires >16 hours of labeling[26,28–30], while BASU enriched a proteome of only two proteins[27]. Further characterization (*see below*) shows that the activities of BioID, BioID2, and BASU are all comparable.

A new PL enzyme that combines the simplicity and non-toxicity of BioID with the catalytic efficiency of APEX2 would greatly enhance PL applications. To achieve this, we undertook the directed evolution of *E. coli* biotin ligase (BirA) to generate new promiscuous variants. To begin, we compared BioID (BirA-R118G) to seven other mutations at the R118 position. We found that R118S is ~2-fold more active than R118G under identical conditions (Supplementary Figure 1, Supplementary 2 for all full blot images, Supplementary Note 1), and hence we selected this mutant rather than BioID as our starting template for evolution.

As in previous work[3,31], we combined yeast surface display of our protein library with Fluorescence Activated Cell Sorting (FACS) to perform the evolution. We used error prone PCR to mutagenize BirA-R118S, generating a ~10^7 library of mutants, each with an average of ~2 amino acid mutations relative to template. This library was then displayed on the yeast surface as a fusion to the Aga2p mating protein (Figure 1b). Biotin and ATP were added to the yeast pool to initiate promiscuous biotinylation, then streptavidin-fluorophore stained biotinylation sites on the surface of each yeast cell. FACS was used to enrich cells displaying a high degree of self-biotinylation over cells displaying low or moderate self-biotinylation (Figure 1b). We gradually reduced the biotinylation time window from 18 hours to 10 minutes over 29 rounds of selection, in order to progressively increase selection stringency (see Supplementary Note 1 and Supplementary Figure 3).

We encountered some technical hurdles during the evolution. First, the activity of our starting template (R118S) and input library were too low to be detected on the yeast surface. Thus, we used Tyramide Signal Amplification (TSA[32]) on the yeast surface to boost the biotin signal (Figure 1c, Supplementary Figure 3b) until the activity of the pool was sufficiently high as to no longer require it. Second, to avoid enriching mutants that strongly tagged their own lysine residues but failed to biotinylate neighboring proteins on the same cell, we treated yeast with the reducing agent TCEP in some rounds of selection, to cleave off the ligase after the biotinylation reaction (Supplementary Figure 3c). Finally, we introduced negative selections to deplete mutants that exhibited strong biotinylation activity prior to exogenous biotin addition, indicating that they can utilize the low levels of biotin naturally present in yeast media (Supplementary Figure 3f).

Our engineering efforts yielded two promiscuous ligases: 35 kD TurboID, with 15 mutations relative to wild-type BirA; and 28 kD miniTurbo, with N-terminal domain deleted and 13 mutations relative to wild-type BirA (Figure 1d, Supplementary Table 1, Supplementary Note 2). Figure 1e and Supplementary Figure 4 show the activity of these ligases on the yeast surface in a side-by-side comparison to BioID, BirA-R118S, and various intermediate clones from our evolution (G1-G3Δ).

To test TurboID and miniTurbo in mammalian cells, we expressed them in the cytosol of HEK 293T cells. Labeling was initiated with the addition of 50 or 500 μM (C_f) exogenous biotin and terminated by cooling cells to 4°C and washing away excess biotin (Supplementary Figure 5). Streptavidin blot analysis of whole cell lysates shows that TurboID and miniTurbo biotinylate endogenous proteins much more rapidly than BioID, giving ~3-6-fold difference in signal at early time points, and ~15-23-fold difference in signal at later time points (Figures 1f, g, Figures 2a, b, Supplementary Figure 4b, Supplementary Figure 6).

The newer promiscuous ligases BioID2[26] and BASU[27] are also included in the comparison, and after normalization to account for differences in ligase expression levels, give activities comparable to that of BioID (Figure 2a, b). Notably, TurboID gives as nearly much biotinylated product in 10 minutes as BioID/BioID2/BASU give in 18 hours (Figure 2a, b). Overall, miniTurbo is 1.5-2-fold less active than TurboID, but exhibits less labeling prior to addition of exogenous biotin; this feature makes miniTurbo potentially superior for precise temporal control of the labeling window. Supplementary Figures 6c-e show the same experiment but use 50 μM biotin instead of 500 μM biotin for labeling; the resulting trends are the same.

To compare ligases by a different modality, we also fixed ligase-expressing HEK 293T cells after biotinylation, stained with neutravidin-fluorophore, and performed confocal microscopy. Supplementary Figure 7 shows that TurboID and miniTurbo give clearly detectable biotinylation in most transfected cells after 10 minutes of biotin labeling. By contrast, BioID, BioID2, and BASU-catalyzed biotinylation are undetectable even at 1 hour, and only dimly detectable at 6 hours in a small fraction of transfected cells.

Different organelles have distinct pH, redox environments, and endogenous nucleophile concentrations, which may influence PL activity. We therefore compared TurboID, miniTurbo, and BioID in the nucleus, mitochondrial matrix, ER lumen, and ER membrane of HEK 293T cells (Figure 2c). We found that the absolute activities of each ligase, as well as the relative activities between ligases, varied across compartments (see Supplementary Note 3). However, TurboID signal was clearly detectable after 10 minutes in each compartment, and even stronger than BioID 18 hour labeling in the mitochondrial matrix and ER lumen. TurboID was superior to miniTurbo in each of these four organelles. Given our observations, we recommend that users test both TurboID and miniTurbo for PL applications, given the context-dependent variations in their activities.

We next evaluated TurboID and miniTurbo in full-scale proteomic experiments. Does 10 minute labeling with these ligases produce proteomic datasets of similar quality to BioID labeling for 18 hours, in terms of specificity, coverage, and labeling radius (Supplementary Note 4 and Supplementary Figure 8)? We selected three mammalian organelles for the analysis: the mitochondrial matrix, nucleus, and ER membrane (ERM) facing cytosol (Figure 2d-h, Supplementary Figures 8-11). Because the ERM is continuous with the cytosol, it is valuable for assessing labeling radius: a good PL enzyme should strongly enrich ERM-localized proteins over immediately adjacent cytosolic proteins.

The HEK293T samples we prepared for proteomic analysis are depicted in Figure 2d and Supplementary Figure 10a. TurboID and miniTurbo labeling were each performed for 10 minutes, whereas BioID labeling was 18 hours. Cells were lysed and biotinylated proteins enriched with streptavidin beads. After on-bead digestion of proteins to peptides, we chemically labeled the peptides with isotopically-distinct TMT (tandem mass tag) labels. This enabled us to quantify relative abundance of each protein across samples. After LC-MS/MS analysis of pooled peptides, we filtered the data via ROC analysis (Supplementary Figure 9c, Supplementary Figure 10f-g), using true positive and false positive protein lists for each organelle (Supplementary Table 2, Supplementary Table 3, and Supplementary Table 4)[2,33], to obtain BioID/TurboID/miniTurbo-derived proteomes for the ERM (Supplementary Table 5), nucleus (Supplementary Table 6), and mitochondrial matrix (Supplementary Table 7).

Figures 2e-h show that TurboID- and miniTurbo-derived 10-minute proteomes have similar size and specificity to BioID-derived 18-hour proteomes in all three compartments. In particular, we note that TurboID is just as effective as BioID in enriching secretory proteins over cytosolic proteins when localized to the ERM (Figures 2e-g), suggesting comparable labeling radius despite much faster labeling kinetics. Depth of coverage is comparable in the mitochondrial matrix and ERM for the three ligases, but slightly lower for TurboID and miniTurbo in the nucleus (Supplementary Figure 9f, Supplementary Figure 10j).

Given the extremely high activity of TurboID, we wondered whether increasing the labeling time would produce a bigger and better proteome. For the ERM, we found that 1 hour labeling with TurboID did increase proteome size by 46% compared to 10 minute labeling, but at the expense of specificity (Figure 2e). With increased labeling time, proximal

nucleophiles may become saturated with biotin, enabling TurboID-generated biotin-AMP to travel farther and biotinylate distal, non-specific proteins.

Despite the widespread application of BioID, there have been only two *in vivo* demonstrations to date[24,34], which we suspect is related to BioID's low catalytic activity. We wondered whether TurboID and miniTurbo's increased activity might enable biotin-based PL in new settings. We first tested these ligases in bacteria (*E. coli*) and yeast (*S. cerevisiae*). Figures 3a, b show that, like in mammalian cells, TurboID and miniTurbo are considerably more active than BioID. In particular, we and others[25] observe that BioID activity is nearly undetectable in yeast, perhaps in part because yeast is cultured at 30°C whereas BioID functions optimally at 37°C[26]. We carried out our directed evolution in yeast at 30°C, so TurboID and miniTurbo are selected for high activity at 30°C.

BioID has not previously been reported in flies (*D. melanogaster*) or worms (*C. elegans*), despite their appeal as highly genetically tractable model organisms. To test biotin-based PL in flies, we expressed BioID, TurboID, or miniTurbo selectively in the larval wing disc, which gives rise to the adult wing, and raised animals on biotin-containing food for 5 days from early embryo stages (Figures 3c). Staining of dissected wing discs with streptavidin-fluorophore shows that TurboID- and miniTurbo-catalyzed biotinylation are 22-fold and 10-fold higher, respectively, than BioID-catalyzed biotinylation (Figures 3d, e). Consistent with our observations in HEK 293T cells, TurboID also gives some low biotinylation signal in flies fed regular, non-biotin supplemented, food.

We also generated flies expressing BioID, TurboID, or miniTurbo in all tissues (*Act-Gal4* driver, Figure 3g), in muscle (*Mef2-Gal4* driver, Supplementary Figure 12a, b), and in all tissues at non-permissive temperature (*tub-Gal4/tub-Gal80^{ts}* driver, Supplementary Figure 12c, d). Animals were either raised on biotin-containing food from early embryo stages to adulthood (13 days) (Figure 3f, g), or raised on regular food to adulthood (13 days) and then switched to biotin-supplemented food for 4 or 16 hours (Supplementary Figure 12). Streptavidin blotting of whole fly lysate showed extensive biotinylation in TurboID and miniTurbo flies, as early as 4 hours post-biotin addition (Supplementary Figure 12), but no signal was detectable in BioID flies, even after 13 days of biotin exposure (Figure 3g). The absence of detectable BioID signal here, compared to the wing experiment (Figure 3d), may result from endogenous biotinylated proteins drowning out specific signal in the streptavidin blot.

To test for possible toxicity of TurboID, miniTurbo, and BioID expression in flies, we performed morphological and survival assays. We observed no evidence of toxicity when any of the ligases were expressed tissue-specifically. However, we did find a decrease in fly viability and size when TurboID was ubiquitously and constitutively expressed and exogenous biotin was withheld (Supplementary Figure 13, Supplementary Note 5, Supplementary Figure 14). We hypothesize that under these conditions,TurboID consumes all the biotin, effectively biotin-starving cells.

We also tested BioID, TurboID, and miniTurbo in *C. elegans*. We expressed the ligases early in the intestinal lineage (~150 min after the first cleavage) and assessed biotinylation activity

~4 hours later, at the embryonic bean stage (stage 1), ~5.5 hours later at the embryonic comma stage (stage 2), or 3 days later in the adult worm (Figure 3h). Figures 3i, j and Supplementary Figure 15 show that TurboID and miniTurbo were much more active than BioID by both imaging and streptavidin blotting at all observed developmental stages. We also observed that TurboID expression was much stronger than miniTurbo expression in adult worms, resulting in much stronger labeling (Figure 3i, Supplementary Figure 15g). Supplementary Figure 15g shows that TurboID and miniTurbo labeling yield can be further increased by raising worms at higher temperatures (25°C vs 20°C).

While we observe background labeling by TurboID in adult biotin-depleted worms (Figure 3i), similar to our observations in flies and mammalian cell culture, we found that miniTurbo, but not TurboID, gave some background labeling in biotin-depleted worm embryos at stage 2 (Figure 3k, Supplementary Figure 15a). We also assessed viability and developmental timing, and did not observe decreased survival in worms expressing any of the three ligases in intestinal cells; however, evident developmental delay was observed in worms expressing TurboID (Supplementary Figure 16, Supplementary Note 5).

In summary, we have performed yeast display-based directed evolution, incorporating TSA signal amplification, reductive removal of ligases, and negative selections, to generate two new ligases for PL applications: TurboID and miniTurbo. TurboID is the most active, and should be used when the priority is to maximize biotinylation yield and sensitivity/recovery. However, in many contexts, we observe a small degree of labeling before exogenous biotin is supplied, indicating that TurboID can utilize the low levels of biotin present in cells/organisms grown in typical biotin-containing media/food. Hence, if the priority is to precisely define the labeling time window, miniTurbo may preferable to TurboID. Though 1.5-2-fold less active than TurboID, miniTurbo gives much less background in the biotin-omitted condition, and it is also 20% smaller (28 versus 35 kD), which may decrease the likelihood of interference with fusion protein trafficking and function. Yet another factor to consider when choosing a ligase for PL is ligase stability. Our results indicate that miniTurbo is less stable than TurboID (likely due to removal of its N-terminal domain), resulting in lower expression levels in the adult worm intestine and adult fly, for example. miniTurbo also exhibits biotin-dependent stability, similar to BioID (see anti-V5 Western blots in Figure 2a, for example).

Up to now, *in vivo* applications of PL have required very long labeling times[24,34] or extensive genetic or manual manipulation[35,36,37] to deliver chemical substrates to relevant cells. TurboID and miniTurbo offer facile substrate delivery and rapid labeling *in vivo*. In addition to increased catalytic efficiency, we believe that the temperature-activity profiles of TurboID and miniTurbo help to explain their superior performance to BioID *in vivo*. Whereas BioID is derived from E. coli (37°C), TurboID and miniTurbo were evolved in yeast (30°C). Flies grow at 25°C, while worms are typically grown at 20°C.

Our toxicity analyses in flies, worms, and mammalian cell culture (Supplementary Figures 13-14, 16) do suggest some necessary precautions when using TurboID and miniTurbo *in vivo*. First, if TurboID is expressed ubiquitously, it can sequester endogenous biotin and cause toxicity; the solution is to supplement animals with exogenous biotin. Second, users

should empirically optimize the *in vivo* labeling time window, and use the shortest labeling time that produces sufficient biotinylated material for analysis. Longer-than-necessary labeling can cause toxicity via chronic biotinylation of endogenous proteomes, and degrade spatial specificity due to saturation of proximal labeling sites (as shown in Figure 2e).

Methods

Cloning

See Supplementary Table 8 for a list of genetic constructs used in this study, with detailed description of construct designs, linker orientations, epitope tags, and signal sequence identities. All ligase variants were derived from *E. coli* biotin protein ligase, have the residue A146 deleted to suppress dimerization[48], and are codon optimized for expression in mammalian cells. For cloning, PCR fragments were amplified using Q5 polymerase (New England BioLabs (NEB)). The vectors were double-digested using standard enzymatic restriction digest and ligated to gel purified PCR products by T4 DNA ligation or Gibson assembly. Ligated plasmid products were introduced by heat shock transformation into competent XL1-Blue bacteria. Ligase mutants were either generated using QuikChange mutagenesis (Stratagene) or isolated from individual yeast clones and transferred to mammalian expression vectors using standard cloning techniques.

Yeast cell culture

For yeast-display (Figure 1c, e and Supplementary Figure 3, 4a), *S. cerevisiae* strain EBY100 was cultured according to previously published protocols[40]. Cells were propagated at 30°C in synthetic dextrose plus casein amino acid (SDCAA, “regular”) medium supplemented with tryptophan (20 mg/L). Yeast cells were transformed with the yeast-display plasmid pCTCON2[40] using the Frozen E-Z Yeast Transformation II kit (Zymoprep) according to manufacturer protocols. Transformed cells containing the *TRP1* gene were selected on SDCAA plates and propagated in SDCAA medium at 30°C. Protein expression was induced by inoculating saturated yeast culture into 10% SD/GCAA (SDCAA medium with 90% of dextrose replaced with galactose), or into “biotin-depleted” medium[41] (1.7 g/L YNB-Biotin (Sunrise Science Products), 5 g/L ammonium sulfate, 2 g/L dextrose, 18 g/L galactose, complete amino acids, 0.125 ng/mL d-biotin), at a 1:1000 dilution and incubating at 30°C for 18 – 24 hr.

For comparison of ligase variants in the yeast cytosol (Figure 3a), *S. cerevisiae* strain BY4741 cells were propagated at 30°C in supplemented minimal medium (SMM; 6.7 g/L Difco nitrogen base without amino acids, 20 g/L dextrose, 0.54 g/L CSM –Ade –His –Leu –Lys –Trp –Ura (Sunrise Science Products), 20 mg/L adenine, 20 mg/L uracil, 20 mg/L histidine, 30 mg/L lysine) supplemented with leucine (20 mg/L). Yeast cells were transformed with pRS415 plasmids using the Frozen E-Z Yeast Transformation II kit (Zymoprep) according to manufacturer protocols. Transformed cells containing the *LEU2* gene were selected on SMM plates (SMM with 20 g/L agar) and propagated in SMM at 30°C. Protein expression was induced by inoculating saturated yeast culture into 10% D/G SMM (SMM medium with 90% of dextrose replaced with galactose) supplemented with 50 μM biotin at a 1:100 dilution and incubating at 30°C for approximately 12 hr.

Generation of ligase libraries for yeast display

Libraries of ligase mutants were generated by error-prone PCR according to published protocols[42]. In brief, 150 ng of the template ligase in vector pCTCON2[40] were amplified for 10 – 20 rounds with 0.4 μM forward and reverse primers (F: 5′-CTAGTGGTGGAGGAGGCTCTGGTGGAGGCGGTAGCGGAGGCGGAGGGTCGGCTAGC-3′, R: 5′-TATCAGATCTCGAGCTATTACAAGTCCTCTTCAGAAATAAGCTTTTGTTCGGATCC-3′), 2 mM $MgCl_2$, 5 units of *Taq* polymerase (NEB), and 2 – 20 μM each of the mutagenic nucleotide analogs 8-oxo-2′-deoxyguanosine-5′-triphosphate (8-oxo-dGTP) and 2′-deoxy-P-nucleoside-5′-triphosphate (dPTP). The PCR products were then gel purified and reamplified for another 30 cycles under normal PCR conditions (F: 5′ CAAGGTCTGCAGGCTAGTGGTGGAGGAGGCTCTGGTG-3′, R: 5′ - CTACACTGTTGTTATCAGATCTCGAGCTATTACAAGTC-3′). The inserts were then electroporated into electrocompetent *S. cerevisiae* EBY100[40] with the BamHI-NheI linearized pCTCON2 vector (10 μg insert/1 μg vector) backbone. The electroporated cultures were rescued in 100 mL of SDCAA medium supplemented with 50 units/mL penicillin and 50 μg/mL streptomycin for 2 days at 30 °C. Refer to “Directed evolution of TurboID and miniTurbo” section below for further details on library generation for each generation of evolution.

Yeast display selections

For each round of evolution (Supplementary Figure 3) At least a ten-fold excess of yeast cells relative to the original library size (approximated via transformation efficiency) was propagated and labeled each round to ensure oversampling. Library protein expression was induced by inoculating saturated yeast culture into 10% SD/GCAA or biotin-depleted medium at a minimum of 1:100 dilution and incubating at 30°C for 18 – 24 hr. For samples biotin labeled for “18 hr,” yeast were induced in 10% SD/GCAA or biotin-depleted medium supplemented with 50 μM biotin, 1 mM ATP, and 5 mM $MgCl_2$ at 30°C for 24 hr. For samples labeled for shorter time points, yeast were induced in 10% SD/GCAA or biotin-depleted medium for 18 hr at 30°C prior to supplementation with 50 μM biotin, 1 mM ATP, and 5 mM $MgCl_2$ for the remaining labeling time indicated. For samples in which biotin labeling was omitted, yeast were induced in 10% SD/GCAA or biotin-depleted medium at 30°C for 18-24 hr. After labeling, approximately 5 million cells were pelleted at 5000*g* for 30 s at 4°C and washed five times with 1 mL PBS (phosphate buffered saline) + 0.5% bovine serum albumin (BSA; 1 mg/mL) (PBS-B).

For removal of ligase proteins via TCEP reduction (Supplementary Figure 3c), yeast were incubated in 500 μL PBS-B + 2 mM TCEP at 30°C for 90 minutes, then washed four times with 1 mL PBS-B. For tyramide signal amplification (TSA, Supplementary Figure 3b), yeast cells were incubated in 50 μL PBS-B + 1:100 streptavidin-horseradish peroxidase (HRP) for 1 hr at 4°C, then washed three times with 1 mL PBS-B. HRP labeling was performed by incubating yeast in 750 μL PBS-B with 50 μM biotin-phenol and 1 mM H_2O_2 for 1 min at room temperature. The reaction was quenched by adding 750 μL PBS-B + 20 mM sodium ascorbate and 10 mM Trolox followed by rapid mixing via inversion. Cells were then

washed two times with 1 mL PBS-B + 10 mM sodium ascorbate and 5 mM Trolox, and once with 1 mL PBS-B.

Yeast cells were then incubated in 50 μL PBS-B + 1:400 chicken anti-myc and 1:50 rabbit anti-biotin (when detecting biotinylated proteins with anti-biotin antibody) for 1 hr at 4°C, then washed three times with 1 mL PBS-B. Yeast cells were then incubated in 50 μL PBS-B + 1:200 Alexa Fluor 647 goat anti-chicken IgG and 1:50 phycoerythrin (PE) goat anti-rabbit IgG (when detecting biotinylated proteins with anti-biotin antibody) or streptavidin-PE (when detecting biotinylated proteins with streptavidin) for 1 hr at 4°C, then washed three times with 1 mL PBS-B for FACS analysis.

For two-dimensional FACS sorting, samples were resuspended in PBS-B at a maximal concentration of 100 million cells/mL and sorted on a BD FACS Aria II cell sorter (BD Biosciences) with the appropriate lasers and emission filters (561 nm and 530/30 for AF488, 640 nm and 575/26 for PE). To analyze and sort single yeast cells, cells were plotted by a forward-scatter area (FSC-A) and side-scatter area (SSC-A) and a gate was drawn around cells clustered between 10^4 – 10^5 FSC-A, 10^3 – 10^5 SSC-A to give population P1 (Supplementary Figure 3i). Cells from population P1 were then plotted by side-scatter width (SSC-W) and side-scatter height (SSC-h) and a gate was drawn around cells clustered between 10 – 100 SSC-W and 10^3 – 10^5 SSC-H to give population P2 (Supplementary Figure 3i). Cells from population P2 were then plotted by forward-scatter width (FSC-W) and forward-scatter height (FSC-H) and a gate was drawn around cells clustered between 10 – 100 FSC-W and 10^3 – 10^5 FSC-H to give population P3 (Supplementary Figure 3i). Population P3 often represented >90% of the total population analyzed.

From population P3, gates were drawn to collect yeast with the highest activity/expression ratio, i.e., positive for AF647 signal that also had high PE signal (Supplementary Figure 3i). For TCEP treated samples, gates were drawn to collect yeast with high PE signal and no AF647 signal above background (Supplementary Figure 3i). For negative selections (Supplementary Figure 3f), gates were drawn to collect yeast with AF647 signal and no PE signal above background (Supplementary Figure 3i). After sorting, yeast were collected in SDCAA medium containing 1% penicillin-streptomycin and incubated at 30°C for 24 h. 1 mL of the growing culture was removed for DNA extraction using the Zymoprep yeast Plasmid Miniprep II (Zymo Research) kit according to manufacturer protocols (using 6 μL zymolyase, vigorously vortex after lysis), and at least ten-fold excess of the number of cells retained during sorting were propagated in SDCAA + 1% pen-strep to ensure oversampling (yeast cells were passaged in this manner at least two times prior to the next round of selection). To analyze yeast populations and clones by FACS (Figure 1c, e; Supplementary Figures 3, 4a), yeast samples were prepared on a small scale (1 mL cultures) as described above, and analyzed on a BD Accuri flow cytometer (BD Biosciences). BD FACSDIVA software was used to analyze all data from FACS sorting and analysis. Refer to “Directed evolution of TurboID and miniTurbo” section below for further details on selections for each round of each generation of evolution. Refer to Supplementary Table 9 for a list of antibodies used in this study.

Directed evolution of TurboID and miniTurbo

Summaries of all yeast display directed evolution and resulting mutants are shown in Figure 1e; Supplementary Figures 3, 4, and Supplementary Table 1.

For the first round of evolution (Supplementary Figure 3b), three libraries were generated using BirA-R118S (Supplementary Table 8) as the starting template. The three libraries were generated using error prone PCR as described above, using the following conditions to result in varying levels of mutagenesis:

Library 1: 2 μM 8-oxo-dGTP, 2 μM dPTP, 10 PCR cycles

Library 2: 2 μM 8-oxo-dGTP, 2 μM dPTP, 20 PCR cycles

Library 3: 20 μM 8-oxo-dGTP, 20 μM dPTP, 10 PCR cycles

The library sizes, as calculated by transformation efficiency, were 1.4 x 10^7 for Library 1, 1.7 x 10^7 for Library 2, and 8 × 10^6 for Library 3. FACS analysis of the three libraries showed robust expression and wide range of activities for Library 1 and Library 2, however Library 3 showed poor expression and no activity. Sequencing of 24 clones in Library 1 revealed an average of 1.5 amino acid changes per ligase gene. Sequencing of 24 clones in Library 2 revealed an average of 2.4 amino acid changes per ligase gene.

Library 1 and Library 2 were combined and used as the initial population for the first round of selections. This combined library was induced as described above, supplemented with 50 μM biotin, 1 mM ATP, and 5 mM $MgCl_2$, for 24 hr. From this culture, approximately 5 x 10^8 cells were prepared for sorting (assuming 1 OD_{600} ≈ 3 x 10^7 cells[43]) as described above with TSA treatment (Supplementary Figure 3b). 6.24 x 10^7 cells were sorted by FACS. A square gate that collected cells positive for both anti-myc and streptavidin (conjugated to fluorophores, see Supplementary table 9) was drawn, and approximately 2.5 x 10^6 cells were collected (4%) to give population E1-R1.

Population E1-R1 was passaged twice, and analyzed by FACS side-by-side with the original combined library and BirA-R118S to ensure the sort was successful (resulting population still had expression and had higher or equivalent activity). Sequencing of 24 clones from E1-R1 revealed an average of 1.5 mutations per ligase gene. Population E1-R1 was induced, supplemented with 50 μM biotin, 1 mM ATP, and 5 mM $MgCl_2$, for 24 hr. From this culture, approximately 10-fold excess (i.e. >2.5 x 10^7) cells were prepared for sorting with TSA treatment. A square gate that collected cells positive for both anti-myc and streptavidin was drawn, and approximately 3.8% of cells were collected to give population E1-R2.

Population E1-R2 was passaged twice, and analyzed by FACS side-by-side with previous rounds and BirA-R118S. Sequencing of 24 clones from E1-R2 revealed an average of 1.5 mutations per ligase gene. Population E1-R2 was induced for ~18 hr then supplemented with 50 μM biotin, 1 mM ATP, and 5 mM $MgCl_2$ for 6 hours. From this culture, approximately 10-fold excess cells were prepared for sorting with TSA treatment. A square gate that collected cells positive for both anti-myc and streptavidin was drawn, and approximately 0.7% of cells were collected to give population E1-R3.

Population E1-R3 was passaged twice, and analyzed by FACS side-by-side with previous rounds and BirA-R118S. Population E1-R3 was induced for ~18 hr then supplemented with 50 μM biotin, 1 mM ATP, and 5 mM $MgCl_2$ for 6 hours. From this culture, approximately 10-fold excess cells were prepared for sorting with TSA treatment. A square gate that collected cells positive for both anti-myc and streptavidin was drawn, and approximately 2.4% of cells were collected to give population E1-R4.

Population E1-R4 was passaged twice, and analyzed by FACS side-by-side with previous rounds and BirA-R118S. Population E1-R4 was induced for ~18 hr then supplemented with 50 μM biotin, 1 mM ATP, and 5 mM $MgCl_2$ for 3 hours. From this culture, approximately 10-fold excess cells were prepared for sorting. A square gate that collected cells positive for both anti-myc and streptavidin was drawn, and approximately 2.6% of cells were collected to give population E1-R5.

Population E1-R5 was passaged twice, and analyzed by FACS side-by-side with previous rounds and BirA-R118S. Population E1-R5 was induced for ~18 hr then supplemented with 50 μM biotin, 1 mM ATP, and 5 mM $MgCl_2$ for 1 hour. From this culture, approximately 10-fold excess cells were prepared for sorting. A square gate that collected cells positive for both anti-myc and streptavidin was drawn, and approximately 0.9% of cells were collected to give population E1-R6.

Population E1-R6 was passaged twice, and analyzed by FACS side-by-side with previous rounds and BirA-R118S. Sequencing of E1-R6 revealed several mutants with the mutation E313K. Several mutants with and without this mutation were assayed as single clones on the yeast surface, and the most promising mutants, including two with the E313K mutation, were assayed in the mammalian cell cytosol. While neither of the E313K mutants showed significant difference in activity to R118S over 24 hours, they both showed very strong self-labeling at shorter time points, e.g. 1 hr. The crystal structure of BirA[44] shows that this residue points directly into the active site, where a lysine mutation could easily react with the phosphate group of biotin-5′-AMP. We removed this mutation from the two promising clones bearing it and assayed again in the mammalian cell cytosol. One of the mutants, denoted in this study as G1 (Supplementary Table 1), displayed significantly higher promiscuous activity than R118S after 24 hours of labeling. Another mutant from the mammalian cell screen, denoted in this study as R6-1 (Supplementary Table 1), also displayed significantly higher promiscuous activity than R118S after 24 hours of labeling. Both of these mutants, with 4 mutations each, had each of their mutations removed individually and in different combinations. Analysis of the resulting mutants in mammalian cells showed that each mutation was contributing to increased activity relative to R118S observed for R6-1 and G1.

For the second round of evolution (Supplementary Figure 3c), six libraries were generated. Three libraries were made using R6-1 (Supplementary Table 1, 8) as the starting template, and the three libraries were made using G1 (Supplementary table 1, 8) as the starting template, both using error prone PCR with the following conditions:

Library 1: R6-1, 2 μM 8-oxo-dGTP, 2 μM dPTP, 10 PCR cycles

Library 2: R6-1, 2 μM 8-oxo-dGTP, 2 μM dPTP, 20 PCR cycles

Library 3: R6-1, 20 μM 8-oxo-dGTP, 20 μM dPTP, 10 PCR cycles

Library 4: G1, 2 μM 8-oxo-dGTP, 2 μM dPTP, 10 PCR cycles

Library 5: G1, 2 μM 8-oxo-dGTP, 2 μM dPTP, 20 PCR cycles

Library 6: G1, 20 μM 8-oxo-dGTP, 20 μM dPTP, 10 PCR cycles

The library sizes, as calculated by transformation efficiency, were 3.8×10^7 for Library 1, 1.9×10^7 for Library 2, 1.6×10^7 for Library 3, 8×10^7 for Library 4, 3.9×10^7 for Library 5, and 3.9×10^7 for Library 6. FACS analysis of the three libraries showed robust expression and wide range of activities for Libraries 1, 2, 4, and 5, however Libraries 3 and 6 showed poor expression and no activity.

Libraries 1, 2, 4, and 5 were combined and used as the initial population for the first round of selections. This combined library was induced, supplemented with 50 μM biotin, 1 mM ATP, and 5 mM $MgCl_2$, for 24 hr. From this culture, approximately 10-fold excess cells were prepared for sorting with TSA treatment. A square gate that collected cells positive for both anti-myc and streptavidin was drawn, and approximately 8.4% of cells were collected to give population E2-R1.

Population E2-R1 was passaged twice, and analyzed by FACS side-by-side with the combined library template. Population E2-R1 was induced, supplemented with 50 μM biotin, 1 mM ATP, and 5 mM $MgCl_2$, for 24 hr. From this culture, approximately 10-fold excess cells were prepared for sorting with TCEP treatment (Supplementary Figure 3c) followed by TSA treatment. A square gate that collected cells positive for streptavidin but negative for anti-myc was drawn, and approximately 1.2% of cells were collected to give population E2-R2.

Population E2-R2 was passaged twice, and analyzed by FACS side-by-side with the combined library template and previous rounds. Population E2-R2 induced, supplemented with 50 μM biotin, 1 mM ATP, and 5 mM $MgCl_2$, for 24 hr. From this culture, approximately 10-fold excess cells were prepared for sorting with TSA treatment. A square gate that collected cells positive for both anti-myc and streptavidin was drawn, and approximately 19% of cells were collected to give population E2-R3.

Population E2-R3 was passaged twice, and analyzed by FACS side-by-side with previous rounds. Population E2-R3 was induced, supplemented with 50 μM biotin, 1 mM ATP, and 5 mM $MgCl_2$, for 24 hr. From this culture, approximately 10-fold excess cells were prepared for sorting. A trapezoidal gate that collected cells positive for both anti-myc and streptavidin, but with high streptavidin/anti-myc ratios, was drawn, and approximately 1.4% of cells were collected to give population E2-R4. From here on, only trapezoidal gates as described here were used for double-positive selections.

Population E2-R4 was passaged twice, and analyzed by FACS side-by-side with previous rounds. Population E2-R4 was induced for ~18 hr, then supplemented with 50 μM biotin, 1 mM ATP, and 5 mM $MgCl_2$ for 1 hr. From this culture, approximately 10-fold excess cells

were prepared for sorting. A trapezoidal gate that collected cells positive for both anti-myc and streptavidin was drawn, and approximately 1.1% of cells were collected to give population E2-R5.

Population E2-R5 was passaged twice, and analyzed by FACS side-by-side with the combined library template and previous rounds. Population E2-R5 was induced for ~18 hr, then supplemented with 50 μM biotin, 1 mM ATP, and 5 mM $MgCl_2$ for 6 hr. From this culture, approximately 10-fold excess cells were prepared for sorting with TCEP treatment followed by TSA. A square gate that collected cells positive for streptavidin and negative for anti-myc was drawn, and approximately 1.5% of cells were collected to give population E2-R6.

Population E2-R6 was passaged twice, and analyzed by FACS side-by-side with previous rounds and the combined library template. Sequencing of E2-R6 revealed several mutations that appeared in multiple clones. Several of these mutants were assayed as single clones on the yeast surface, however it was found after re-sequencing that many of the most promising mutants had mutated stop codons. After mutating back the stop codons, the mutants were re-assayed on the yeast surface, and the mutants that remained promising were assayed in the mammalian cell cytosol. One of the mutants, denoted in this study as G2 (Supplementary Table 1), displayed significantly higher promiscuous activity than R118S, G1 (its template), or any other mutant tested after 1 hour of labeling. G1, with 2 additional mutations relative to G1, had each or both of its mutations removed. Analysis of the resulting mutants in mammalian cells showed that each mutation was contributing to activity boost observed for G2.

For the third round of evolution (Supplementary Figure 3d), three libraries were made using G2 as the starting template (Supplementary table 1, 8) using error prone PCR with the following conditions:

Library 1: 2 μM 8-oxo-dGTP, 2 μM dPTP, 10 PCR cycles

Library 2: 2 μM 8-oxo-dGTP, 2 μM dPTP, 20 PCR cycles

Library 3: 10 μM 8-oxo-dGTP, 20 μM dPTP, 10 PCR cycles

The library sizes, as calculated by transformation efficiency, were 3.5×10^8 for Library 1, 3.6×10^7 for Library 2, and 6.8×10^6 for Library 3. FACS analysis of the three libraries showed robust expression and wide range of activities for Library 1 and Library 2, however Library 3 showed weak expression and no activity.

Libraries 1 and 2 were combined and used as the initial population for the first round of selections. This combined library was induced for ~18 hr, then supplemented with 50 μM biotin, 1 mM ATP, and 5 mM $MgCl_2$ for 1 hr. From this culture, approximately 10-fold excess cells were prepared for sorting. A trapezoidal gate that collected cells positive for both anti-myc and streptavidin was drawn, and less than 0.1% of cells were collected to give population E3-R1.

Population E3-R1 was passaged twice, and analyzed by FACS side-by-side with G2 and the combined library template. Population E3-R1 was induced for ~18 hr, then supplemented

with 50 μM biotin, 1 mM ATP, and 5 mM $MgCl_2$ for 1 hr. From this culture, approximately 10-fold excess cells were prepared for sorting. A trapezoidal gate that collected cells positive for both anti-myc and streptavidin was drawn, and 0.15% of cells were collected to give population E3-R2.

Population E3-R2 was passaged twice, and analyzed by FACS side-by-side with G2, the combined library template, and previous rounds. Population E3-R2 was induced for ~18 hr, then supplemented with 50 μM biotin, 1 mM ATP, and 5 mM $MgCl_2$ for 10 min. From this culture, approximately 10-fold excess cells were prepared for sorting. A trapezoidal gate that collected cells positive for both anti-myc and streptavidin was drawn, and less than 0.1% of cells were collected to give population E3-R3.

At E3-R3, it was noted that the population had strong streptavidin signal in the absence of exogenous biotin addition. Sequencing of population E3-R3 revealed that the majority of clones had a large insertion at the 5′ of the ligase gene. Removal of this insertion restored biotin dependence, but also resulted in decreased activity (5-fold less than E3-R3). The library was “cleaned” by removing this insertion via PCR with primers that restored the wild-type N-terminal sequence, and subjected to one additional round of double-positive selection with 10 minute labeling and 0.1% cells collected. The resulting population was E3-R4.

Population E3-R4 was passaged twice, and analyzed by FACS side-by-side with previous rounds. Sequencing of E3-R4 revealed several mutations that appeared in multiple clones. Several of these mutants were assayed as single clones on the yeast surface, the most promising mutants were assayed in the mammalian cell cytosol. Two mutants had significantly higher activity than the template G2 or any other mutants. The mutations from these mutants were combined in various combinations, resulting in the highest activity mutant, denoted in this study as G3 (Supplementary Table 1).

G3 was the highest activity mutant found to date, but it also appeared to have streptavidin signal without the addition of exogenous biotin. This was observed in yeast, where this signal proved to be biotin-dependent (Supplementary Figure 3e), and also in the mammalian cytosol (Figure 1f, g, Supplementary Figure 4b). From this point, we continued with two evolutions as follows:

In one path, we truncated the N-terminal domain (aa 1-63) of G3 to give G3Δ (Supplementary Table 1). Consistent with literature[45,46], this truncation resulted in reduced streptavidin signal when exogenous biotin was omitted (Supplementary Figure 3e). Using G3Δ as the starting template (Supplementary table 1, 8) for another round of evolution (Supplementary Figure 3g), we generated three libraries using error prone PCR with the following conditions:

Library 1: 2 μM 8-oxo-dGTP, 2 μM dPTP, 10 PCR cycles

Library 2: 2 μM 8-oxo-dGTP, 2 μM dPTP, 20 PCR cycles

Library 3: 4 μM 8-oxo-dGTP, 2 μM dPTP, 20 PCR cycles

The library sizes, as calculated by transformation efficiency, were 4.9 x 10^8 for Library 1, 4.6 × 10^8 for Library 2, and 3.7 x 10^8 for Library 3. FACS analysis of the three libraries showed robust expression and wide range of activities for all libraries, therefore all were combined and used for the first round of selections.

This combined library was induced in biotin-depleted media, supplemented with 50 μM biotin, 1 mM ATP, and 5 mM $MgCl_2$, for 18 hr. From this culture, approximately 10-fold excess cells were prepared for sorting with streptavidin. A trapezoidal gate that collected cells positive for both anti-myc and streptavidin was drawn, and 0.1% of cells were collected to give population E4-R1.

Population E4-R1 was passaged twice, and analyzed by FACS side-by-side with G3Δ and the combined library template. Population E4-R1 was induced for ~18 hr in biotin-depleted media, then supplemented with 50 μM biotin, 1 mM ATP, and 5 mM $MgCl_2$ for 3.5 hr. From this culture, approximately 10-fold excess cells were prepared for sorting with anti-biotin antibody. A trapezoidal gate that collected cells positive for both anti-myc and anti-biotin was drawn, and 1% of cells were collected to give population E4-R2.

Population E4-R2 was passaged twice, and analyzed by FACS side-by-side with G3Δ, the combined library template, and previous rounds. Population E4-R2 was induced for ~18 hr in biotin-depleted media, then supplemented with 50 μM biotin, 1 mM ATP, and 5 mM $MgCl_2$ for 1 hr. From this culture, approximately 10-fold excess cells were prepared for sorting with streptavidin. A trapezoidal gate that collected cells positive for both anti-myc and streptavidin was drawn, and 0.2% of cells were collected to give population E4-R3.

Population E4-R3 was passaged twice, and analyzed by FACS side-by-side with G3Δ, the combined library template, and previous rounds. Population E4-R3 was induced for ~18 hr in biotin-depleted media, then supplemented with 50 μM biotin, 1 mM ATP, and 5 mM $MgCl_2$ for 1 hr. From this culture, approximately 10-fold excess cells were prepared for sorting with anti-biotin antibody. A trapezoidal gate that collected cells positive for both anti-myc and anti-biotin was drawn, and 0.1% of cells were collected to give population E4-R4.

Population E4-R4 was passaged twice, and analyzed by FACS side-by-side with G3Δ, the combined library template, and previous rounds. Population E4-R4 was induced for ~18 hr in biotin-depleted media, labeling was omitted for negative selection (Supplementary Figure 3f). From this culture, approximately 10-fold excess cells were prepared for sorting with streptavidin. A square gate that collected cells positive for anti-myc and negative for streptavidin was drawn, and 50% of cells were collected to give population E4-R5.

Population E4-R5 was passaged twice, and analyzed by FACS side-by-side with G3Δ, the combined library template, and previous rounds. Two selections were performed on E4-R5. In the first selection, population E4-R5 was induced for ~18 hr in biotin-depleted media, labeling was omitted for negative selection. From this culture, approximately 10-fold excess cells were prepared for sorting with anti-biotin antibody. A square gate that collected cells positive for both anti-myc and anti-biotin was drawn, and 45% of cells were collected to give population E4-R6.1.

In the second selection, population E4-R5 was induced for ~18 hr in biotin-depleted media, then supplemented with 50 μM biotin, 1 mM ATP, and 5 mM $MgCl_2$ for 20 min. From this culture, approximately 10-fold excess cells were prepared for sorting with streptavidin. A trapezoidal gate that collected cells positive for both anti-myc and streptavidin was drawn, and 0.1% of cells were collected to give population E4-R6.2. One more round of selections was performed on E4-R6.1, which was induced for ~18 hr in biotin-depleted media, then supplemented with 50 μM biotin, 1 mM ATP, and 5 mM $MgCl_2$ for 1 hr. From this culture, approximately 10-fold excess cells were prepared for sorting with streptavidin. A trapezoidal gate that collected cells positive for both anti-myc and streptavidin was drawn, and 0.2% of cells were collected to give population E4-R7.

Population E4-R7 was passaged twice, and analyzed by FACS side-by-side with previous rounds. Sequencing of E4-R7 revealed several mutations that appeared in multiple clones. Several of these mutations were assayed as single mutations and in various combinations in the mammalian cytosol. One mutation, K194I, was found to significantly increase activity while not increasing signal exogenous when biotin is omitted. Introducing K194I into G3Δ resulted in miniTurbo (Supplementary Table 1).

In a second path, we continued with evolving G3 (Supplementary Figure 3h). Two libraries were made using G3 as the starting template (Supplementary table 1, 8) using error prone PCR with the following conditions:

Library 1: 2 μM 8-oxo-dGTP, 2 μM dPTP, 10 PCR cycles

Library 2: 2 μM 8-oxo-dGTP, 2 μM dPTP, 20 PCR cycles

The library sizes, as calculated by transformation efficiency, were 2×10^7 for Library 1 and 1.1×10^7 for Library 2. FACS analysis of the libraries showed robust expression and wide range of activities for Library 1 and Library 2.

Libraries 1 and 2 were combined and used as the initial population for the first round of selections. This combined library was induced for ~18 hr in biotin-depleted media, then supplemented with 50 μM biotin, 1 mM ATP, and 5 mM $MgCl_2$ for 10 min. From this culture, approximately 10-fold excess cells were prepared for sorting with anti-biotin antibody (Supplementary table 9) in place of streptavidin. A trapezoidal gate that collected cells positive for both anti-myc and anti-biotin was drawn, and 0.1% of cells were collected to give population E5-R1.

Population E5-R1 was passaged twice, and analyzed by FACS side-by-side with G3 and the combined library template. Population E5-R1 was induced for ~18 hr in biotin-depleted media, then supplemented with 50 μM biotin, 1 mM ATP, and 5 mM $MgCl_2$ for 10 min. From this culture, approximately 10-fold excess cells were prepared for sorting with anti-biotin antibody. A trapezoidal gate that collected cells positive for both anti-myc and anti-biotin was drawn, and 0.1% of cells were collected to give population E5-R2.

Population E5-R2 was passaged twice, and analyzed by FACS side-by-side with G3, the combined library template, and previous rounds. Population E5-R2 was induced for ~18 hr in biotin-depleted media, then supplemented with 50 μM biotin, 1 mM ATP, and 5 mM

$MgCl_2$ for 10 min. From this culture, approximately 10-fold excess cells were prepared for sorting with anti-biotin antibody. A trapezoidal gate that collected cells positive for both anti-myc and anti-biotin was drawn, and 1.7% of cells were collected to give population E5-R3.

Population E5-R3 was passaged twice, and analyzed by FACS side-by-side with G3, the combined library template, and previous rounds. Population E5-R3 was induced for ~18 hr in regular media, labeling was omitted for negative selection. From this culture, approximately 10-fold excess cells were prepared for sorting with anti-biotin antibody. A square gate that collected cells positive for anti-myc and negative for anti-biotin was drawn, and 34% of cells were collected to give population E5-R4.

Population E5-R4 was passaged twice. FACS analysis side-by-side with G3, the combined library template, and previous rounds showed that the negative selection that resulted E5-R4 reduced overall activity of the population. Population E5-R4 was induced for ~18 hr in biotin depleted media, then supplemented with 50 μM biotin, 1 mM ATP, and 5 mM $MgCl_2$ for 10 min. From this culture, approximately 10-fold excess cells were prepared for sorting with streptavidin. A trapezoidal gate that collected cells positive for both anti-myc and streptavidin was drawn, and 0.8% of cells were collected to give population E5-R5.

Population E5-R5 was passaged twice, and analyzed by FACS side-by-side with G3, the combined library template, and previous rounds. Population E5-R5 was induced for ~18 hr in regular media, labeling was omitted for negative selection. From this culture, approximately 10-fold excess cells were prepared for sorting with anti-biotin antibody. A square gate that collected cells positive for anti-myc and negative for anti-biotin was drawn, and 11.6% of cells were collected to give population E5-R6.

Population E5-R6 was passaged twice, and analyzed by FACS side-by-side with previous rounds. Sequencing of E5-R6 revealed several mutations that appeared in multiple clones. Several of these mutations were assayed as single mutations and in various combinations in the mammalian cytosol. None of the mutations gave dramatic increases in activity, but one mutation M241T, appeared to impart benefits to activity. Screening of mutations present in E4-R6.2 in the mammalian cell cytosol revealed one mutation, S263P, which boosted activity, but also increased signal when biotin was omitted. This mutation, along with K194I from E4-R7 and M241T from E5-R6, were introduced into G3 to give TurboID (Supplementary Table 1). We also tested M241T in miniTurbo, however it was not added because it increased background signal when biotin was omitted.

Mammalian cell culture, transfection, and stable cell line generation

HEK 293T cells from ATCC (passage number <25) were cultured as a monolayer in growth media (either MEM (Cellgro) or a 1:1 DMEM:MEM mixture (Cellgro) supplemented with 10% (w/v) fetal bovine serum (VWR), 50 units/mL penicillin, and 50 μg/mL streptomycin at 37°C under 5% CO_2. Mycoplasma testing was not performed before experiments. For confocal imaging experiments, cells were grown on 7 x 7 mm glass coverslips in 48-well plates with 250 μL growth medium. To improve adherence of HEK 293T cells, glass coverslips were pretreated with 50 μg/mL fibronectin (Millipore) in MEM for at least 20 min

at 37°C before cell plating. For Western blotting, cells were grown on polystyrene 6-well plates (Greiner) with 2.5 mL growth medium. For transient expression (Figure 1f, 2c and Supplementary Figures 1, 4b, 5, 10b, d), cells were typically transfected at approximately 60% confluency using 3.2 μL/mL Lipofectamine2000 (Life Technologies) and 800 ng/mL plasmid in serum-free media (250 μL total volume for 48-wells, 2.5 mL total volume for 6-wells) for 3-4 hr, after which time Lipofectamine-containing media was replaced with fresh serum-containing media.

In attempts to achieve similar expression levels of ligase in the experiment presented in Figure 2a, b, Supplementary Figure 6a, b, and Supplementary Figure 7, cells were transfected at approximately 60% confluency using 1.6 μL/mL Lipofectamine2000 (Life Technologies) in serum-free media with the following amounts of each plasmid (250 μL total volume for 48-wells, 2.5 mL total volume for 6-wells): 160 ng/mL V5-BioID-NES, 80 ng/mL V5-TurboID-NES, 200 ng/mL V5-miniTurbo-NES, 30 ng/mL V5-BioID2-NES, and 1000 ng/mL V5-BASU-NES (Supplementary Table 8). After 3-4 hr, the Lipofectamine-containing media was replaced with fresh serum-containing media.

In attempts to achieve similar expression levels of ligase in the experiment presented in Supplementary Figure 6c-e, cells were transfected at approximately 60% confluency using 1.6 μL/mL Lipofectamine2000 (Life Technologies) in serum-free media with the following amounts of each plasmid (250 μL total volume for 48-wells, 2.5 mL total volume for 6-wells): 320 ng/mL V5-BioID-NES, 160 ng/mL V5-TurboID-NES, 400 ng/mL V5-miniTurbo-NES, 60 ng/mL V5-BioID2-NES, and 1000 ng/mL V5-BASU-NES (Supplementary Table 8). After 3-4 hr, the Lipofectamine-containing media was replaced with fresh serum-containing media.

BioID expressing-cells were typically labeled by supplementing media with 50 or 500 μM biotin for 18 hr, approximately 18 hr after transfection; for shorter time-points, labeling was initiated approximately 30-36 hr after transfection. TurboID and miniTurbo expressing-cells were typically labeled by supplementing 50 or 500 μM biotin for 10 min, approximately 36 hr after transfection; for longer time-points, labeling was initiated between 18-35 hr after transfection. Labeling was stopped by placing cells on ice and washing five times with PBS (Supplementary Figure 5).

For preparation of lentiviruses, HEK 293T cells in T25 flasks (BioBasic) were transfected at ~60-70% confluency with the lentiviral vector pLX304 containing the gene of interest (2500 ng; Supplementary Table 8), and the lentiviral packaging plasmids pVSVG (250 ng; Supplementary Table 8) and Δ8.9 (2250 ng; Supplementary Table 8) with 30 μL Lipofectamine2000 in serum-free media for 3 hr, after which time the Lipofectamine-containing media was replaced with fresh serum-containing media. Approximately 60 hr after transfection, the cell medium containing the lentivirus was harvested and filtered through a 0.45-μm filter. To generate stable cell lines, HEK cells were then infected at ~50% confluency, followed by selection with 8 μg/mL blasticidin in growth medium for at least 7 days before further analysis (Figure 2c, d and Supplementary Figure 8; 9a, b; 10c, e, 14a, b). Cells stably expressing BioID were labeled by supplementing media with 50 or 500 μM biotin for 18 hr. Cells stably expressing BioID were typically by supplementing media with

50 or 500 μM biotin for 10 min. Labeling was stopped by placing cells on ice and washing five times with PBS (Supplementary Figure 5).

Synthesis of homemade neutravidin-AlexaFluor647 conjugate

A reaction mixture was assembled in a 1.5 mL Eppendorf tube with the following components (added in this order): 200 μL of 5 mg/mL Neutravidin (Life Technologies) in PBS, 20 μL of 1 M sodium bicarbonate in water, and 10 μL of 10 mg/mL AlexaFluor647-NHS Ester (Life Technologies) in anhydrous DMSO. The tube was incubated at room temperature with rotation in the dark for 3 h. The neutravidin-AlexaFluor647 conjugate was purified from unreacted dye using a NAP-5 size-exclusion column (GE Healthcare Life Sciences) according to the manufacturer's instructions. The conjugate was typically eluted from the column in 500 μL cold PBS. Absorbance values, determined using a Nanodrop 2000c UV-vis spectrophotometer (Thermo Scientific), were typically as follows: A280 = ~0.284 and A647 = ~1.625. The conjugate was stable at 4 °C in the dark for at least 4 months and was flash frozen and stored at −80 °C for longer term storage. For mammalian cell labeling experiments, the conjugate was diluted 1,000-fold in PBS containing 1% BSA.

Gels and Western blots

For gels and Western blots experiments in Figure 1f, 2a, c and Supplementary Figure 1, 4b, 5, 6, 8b-d, 9a, 10b-c, HEK 293T cells expressing the indicated constructs were plated, transfected, and labeled with biotin as described above, and subsequently scraped and pelleted by centrifugation at 1500 rpm for 3 min. The pellet was lysed by resuspending in RIPA lysis buffer (50 mM Tris pH 8, 150 mM NaCl, 0.1% SDS, 0.5% sodium deoxycholate, 1% Triton X-100, 1X protease inhibitor cocktail (Sigma-Aldrich), and 1 mM PMSF) by gentle pipetting and incubating for 5 min at 4°C. Lysates were clarified by centrifugation at 10000 rpm for 10 min at 4°C. Protein concentration in clarified lysate was estimated with Pierce BCA Protein Assay Kit (ThermoFisher) prior to separation on a 9% SDS-PAGE gel. Silver-stained gels (Supplementary Figure 9a, 10b-c) were generated using Pierce Silver Stain Kit (ThermoFisher).

For the Western blot experiment in Figure 3a, BY4741 yeast expressing the indicated constructs (Supplementary table 8) were induced as described above and supplemented with 50 μM biotin for the duration of induction. After approximately 12 hr, the saturated induced culture was diluted 1:30 in fresh induction media supplemented with 50 μM biotin and allowed to grow for approximately 6 hr more until reaching OD_{600} ~1. Three milliliters of this culture was pelleted (normalized across samples so that the same approximate amount of cells are collected for each sample), and lysed on ice in 50 μL 1.85 M NaOH + 300 mM β-mercaptoethanol for 10 min on ice. The protein in the lysate was then precipitated by adding 50 μL 50% (w/v) TCA and incubating on ice for 15 min. The protein was pelleted at 12000*g* for 5 min, then dissolved in 120 μL urea/SDS buffer (0.48 g/mL urea, 50 mg/mL SDS, 29.2 mg/mL EDTA, 15.4 mg/mL DTT, 1 mg/mL bromophenol blue, 12 mg/mL Tris base, 0.2 mL/mL 1M Tris pH6.8). Proteins were boiled for 10 min prior to separation on a 9% SDS-PAGE gel.

For the Western blot experiment in Figure 3b, BL21 bacteria expressing the indicated constructs (Supplementary table 8) were induced overnight (18 hr) at 37°C in Lysogeny Broth (LB) supplemented with 100 μg/mL ampicillin, 100 μg/mL IPTG, and with or without 50 μM biotin. Grown to approximately OD_{600} = 0.6, 100 μL of each culture was pelleted (normalized across samples so that the same approximate amount of cells are collected for each sample) and resuspended in 15 μL 6X protein loading buffer (0.33 M Tris-HCl pH .8, 34% glycerol, 94 mg/mL SDS, 88 mg/mL DTT, 113 μg/mL bromophenol blue). The protein was boiled for 5 min, diluted to 1X, and then separated on a 9% SDS-PAGE gel.

For all Western blots in Figure 1f; 2a, c; 3a, b and Supplementary Figure 1; 4b; 5; 6; 8b-d; 9a; 10b-c, proteins separated on SDS-PAGE gels were transferred to nitrocellulose membrane, and then stained by Ponceau S (5 min in 0.1% (w/v) Ponceau S in 5% acetic acid/water). The blots were then blocked in 5% (w/v) milk (LabScientific) in TBS-T (Tris-buffered saline, 0.1% Tween 20) for at least 30 min at room temperature, or as long as overnight at 4°C. Blots were then stained with primary antibodies (Supplementary table 9) in 3% BSA (w/v) in TBS-T for 1 - 16 hr at 4°C, washed four times with TBS-T for 5 min each, then stained with secondary antibodies or 0.3 μg/mL streptavidin-HRP (Supplementary table 9) in 3% BSA (w/v) in TBS-T for 1 at 4°C. The blots were washed four times with TBS-T for 5 min each prior to developing with Clarity Western ECL Blotting Substrates (BioRad) and imaging on a UVP BioSpectrum Imaging System. Quantitation of Western blots was performed using ImageJ on raw images under non-saturating conditions.

Confocal fluorescence imaging of cultured cells

For fluorescence imaging experiments in Supplementary Figure 8e, 9b and 10d, e, HEK 293T cells expressing the indicated constructs were plated, transfected, and labeled with biotin as described above, and subsequently fixed with 4% (v/v) paraformaldehyde in PBS at 4°C for 45 min. Cells were then washed three times with PBS and permeabilized with cold methanol at −20°C for 5 min. Cells were then washed three times with PBS, and then incubated with primary antibody (Supplementary table 9) in PBS supplemented with 3% (w/v) BSA for 1 hr at 4°C. After washing three times with PBS, cells were then incubated with DAPI/secondary antibody, and neutravidin-Alexa Fluor647 (Supplementary table 9) in PBS supplemented with 3% (w/v) BSA for 1 hr at 4°C. Cells were then washed three times with PBS and imaged by confocal fluorescence microscopy.

Confocal imaging was performed using a Zeiss AxioObserver.Z1 microscope, outfitted with a Yokogawa spinning disk confocal head, a Cascade II:512 camera, a Quad-band notch dichroic mirror (405/488/568/647), 405 (diode), 491 (DPSS), 561 (DPSS), and 640 (diode) nm lasers (all 50 mW). DAPI (405 laser excitation, 445/40 emission), Alexa Fluor568 (561 laser excitation, 617/73 emission), and Alexa Fluor647 (640 laser excitation, 700/75 emission), and differential intereference contrast (DIC) images were acquired through a 63X oil-immersive objective; Acquisition times ranged from 50 to 100 ms. All images were collected and processed using SlideBook 6.0 software (Intelligent Imaging Innovations). The data in Supplementary Figure 8e, 9b, and 10d, e are representative of at least 10 fields of view.

Sample preparation for proteomics

HEK 293T cells were grown in T150 flasks per proteomic sample as described above. Nuclear samples were transfected with 30 μg DNA using 150 μL Lipofectamine 2000 for 4 hr. BioID samples were labeled using 50 μM biotin for 18 hr, TurboID and miniTurbo samples were labeled using 50 μM biotin for 18 hr. ER membrane and mitochondrial matrix samples were generated using stable cell lines. Imaging of samples cultured and labeled in the same manner as the larger scale proteomic samples were prepared for quality controls (Supplementary Figure 9b and 10d, e). Cell pellets were collected and lysed in approximately 1.5 mL RIPA lysis buffer as described above, and clarified by centrifugation at 10,000 rpm for 10 min at 4°C. 2.5% of this lysate was separated and used for quality control analysis of expression and labeling by Western blotting as described above (Supplementary Figure 9a and 10b, c), and for estimating protein concentration in clarified lysate using Pierce BCA Protein Assay Kit (ThermoFisher).

This preparation was also employed for samples in the proximity labeling experiment shown in Supplementary Figure 8, where ER membrane and outer mitochondrial membrane stable cell lines were used to generate samples.

Streptavidin bead enrichment of biotinylated material

For enrichment of biotinylated material, 350 μL streptavidin-coated magnetic beads (Pierce) were washed twice with RIPA buffer, then incubated with clarified lysates containing approximately 3 mg protein for each sample with rotation for 1 hr at room temperature, after which 5% of beads were removed for quality control analysis of enrichment (Supplementary Figure 9a and 10b, c), and then the remaining beads were moved to 4°C and incubated overnight. The beads were subsequently washed twice with 1 mL of RIPA lysis buffer, once with 1 mL of 1 M KCl, once with 1 mL of 0.1 M Na_2CO_3, once with 1 mL of 2 M urea in 10 mM Tris-HCl (pH 8.0), and twice with 1 mL RIPA lysis buffer. For quality control analysis, biotinylated proteins were eluted by boiling the beads in 75 μL of 3X protein loading buffer supplemented with 20 mM DTT and 2 mM biotin, run on SDS-PAGE gel, and stained using Pierce Silver Stain Kit.

This enrichment protocol was also employed for samples in the proximity labeling experiment shown in Supplementary Figure 8, but instead protein was eluted from the total amount of beads and separated on SDS-PAGE gel for Western blotting as described above with antibodies against the endogenous proteins indicated in Supplementary Figure 8 (Supplementary Table 9).

On-bead trypsin digestion of biotinylated peptides

To prepare proteomic samples for mass spectrometry analysis, proteins bound to streptavidin beads (~300 μL of slurry) were washed twice with 200 μL of 50 mM Tris HCl buffer (pH 7.5) followed by two washes with 2 M urea/50 mM Tris (pH 7.5) buffer. The final volume of 2 M urea/50 mM Tris buffer (pH 7.5) was removed and beads were incubated with 80 μL of 2 M urea/50 mM Tris containing 1 mM DTT and 0.4 μg trypsin for 1 h at 25 C with shaking. After 1 h, the supernatant was removed and transferred to a fresh tube. The streptavidin beads were washed twice with 60 μL of 2 M urea/50 mM Tris buffer (pH 7.5) and the

washes were combined with the on-bead digest supernatant. The eluate was reduced with 4 mM DTT for 30 min at 25°C with shaking. The samples were alkylated with 10 mM iodoacetamide for 45 min in the dark at 25°C with shaking. An additional 0.5 μg of trypsin was added to the sample and the digestion was completed overnight at 25°C with shaking. After overnight digestion, samples was acidified (pH < 3) by adding formic acid (FA) such that the sample contained ~1% FA. Samples were desalted on C18 StageTips and evaporated to dryness in a vacuum concentrator, exactly as previously described[47].

TMT labeling and fractionation of peptides

Desalted peptides were labeled with TMT (6-plex or 11-plex) reagents. Peptides were reconstituted in 100 μL of 50 mM HEPES. Each 0.8 mg vial of TMT reagent was reconstituted in 41 μL of anhydrous acetonitrile and added to the corresponding peptide sample for 1 h at room temperature. Labeling of samples with TMT reagents was completed with the design indicated in Figure 2d and Supplementary Figure 10a. TMT labeling reactions were quenched with 8 μL of 5% hydroxylamine at room temperature for 15 minutes with shaking, evaporated to dryness in a vacuum concentrator, and desalted on C18 StageTips. For each TMT 6-plex cassette and the TMT 11-plex cassette, 50% of the sample was fractionated by basic pH reversed phase using StageTips while the other 50% of each sample was reserved for LC-MS analysis by a single-shot, long gradient. One StageTip was prepared per sample using 2 plugs of Styrene Divinylbenzene (SDB) (3M) material. The StageTips were conditioned two times with 50 μL of 100% methanol, followed by 50 μL of 50%MeCN/0.1% FA, and two times with 75 μL of 0.1% FA. Sample, resuspended in 100 μL of 0.1% FA, was loaded onto the stage tips and washed with 100 μL of 0.1% FA. Following this, sample was washed with 60 μL of 20mM NH_4HCO_2/2% MeCN, this wash was saved and added to fraction 1. Next, sample was eluted from StageTip using the following concentrations of MeCN in 20 mM NH_4HCO_2: 10%, 15%, 20%, 25%, 30%, 40%, and 50%. For a total of 6 fractions, 10 and 40% (fractions 2 and 7) elutions were combined, as well as 15 and 50% elutions (fractions 3 and 8). The six fractions were dried by vacuum centrifugation.

Liquid chromatography and mass spectrometry

Desalted peptides were resuspended in 9 μL of 3% MeCN/0.1% FA and analyzed by online nanoflow liquid chromatography tandem mass spectrometry (LC-MS/MS) using an Orbirtrap Fusion Lumos Tribrid MS (ThermoFisher Scientific) coupled on-line to a Proxeon Easy-nLC 1200 (ThermoFisher Scientific). Four microliters of each sample was loaded onto a microcapillary column (360 μm outer diameter × 75 μm inner diameter) containing an integrated electrospray emitter tip (10 μm), packed to approximately 24 cm with ReproSil-Pur C18-AQ 1.9 μm beads (Dr. Maisch GmbH) and heated to 50 °C. The HPLC solvent A was 3% MeCN, 0.1% FA, and the solvent B was 90% MeCN, 0.1% FA. The SDB fractions were measured using a 110 min MS method, which used the following gradient profile: (min:%B) 0:2; 1:6; 85:30; 94:60; 95:90; 100:90; 101:50; 110:50 (the last two steps at 500 nL/min flow rate). Non-fractionated samples were analyzed using a 260 min MS method with the following gradient profile: (min:%B) 0:2; 1:6; 235:30; 244:60; 245:90; 250:90; 251:50; 260:50 (the last two steps at 500 nL/min flow rate).

The Orbitrap Fusion Lumos Tribrid was operated in the data-dependent mode acquiring HCD MS/MS scans (resolution = 15,000 for TMT6-plex, or resolution = 50,000 for TMT11-plex) after each MS1 scan (resolution = 60,000) on the most abundant ions within a 2 s cycle time using an MS1 target of 3×10^6 and an MS2 target of 5×10^4. The maximum ion time utilized for MS/MS scans was 50 ms for TMT6-plex experiments and 105 ms for the TMT 11-plex experiment; the HCD normalized collision energy was set to 34 for TMT6 and 38 for TMT11; the dynamic exclusion time was set to 45 s, and the peptide match and isotope exclusion functions were enabled. Charge exclusion was enabled for charge states that were unassigned, 1 and >6.

Proteomic data analysis

Collected data were analyzed using Spectrum Mill software package v6.1pre-release (Agilent Technologies). Nearby MS scans with the similar precursor m/z were merged if they were within ± 60 s retention time and ±1.4 m/z tolerance. MS/MS spectra were excluded from searching if they failed the quality filter by not having a sequence tag length 0 or did not have a precursor MH+ in the range of 750 – 4000. All extracted spectra were searched against a UniProt[48] database containing human reference proteome sequences. Search parameters included: parent and fragment mass tolerance of 20 ppm, 30% minimum matched peak intensity, trypsin allow P enzyme specificity with up to four missed cleavages, and calculate reversed database scores enabled. Fixed modifications were carbamidomethylation at cysteine. TMT labeling was required at lysine, but peptide N termini were allowed to be either labeled or unlabeled. Allowed variable modifications were protein N-terminal acetylation and oxidized methionine. Individual spectra were automatically assigned a confidence score using the Spectrum Mill autovalidation module. Score at the peptide mode was based on target-decoy false discovery rate (FDR) of 1%. Protein polishing autovalidation was then applied using an auto thresholding strategy. Relative abundances of proteins were determined using TMT reporter ion intensity ratios from each MS/MS spectrum and the mean ratio is calculated from all MS/MS spectra contributing to a protein subgroup. Proteins identified by 2 or more distinct peptides were considered for the dataset.

ER membrane proteomic data analysis

Complete mass spectrometry data for the ER membrane (ERM) proteomic experiment are shown in Supplementary Table 5 Tab 6. To select cutoffs for proteins biotinylated by the indicated ligase over non-specific bead binders, we classified the detected proteins into three groups:

1. ERM proteins (Supplementary Table 2 Tab 1; true positive list of 90 well-established ERM proteins[33]).
2. soluble matrix proteins (Supplementary Table 2 Tab 2; false positive list of 173 soluble mitochondrial matrix proteins[2]).
3. all other proteins.

We then normalized the TMT ratios in order to account for differences in total protein quantity between samples within the TMT 11-plex experiment. To do this, the Log_2(TMT

ratios) corresponding to ERM-ligase/untransfected (Log_2(127N/126C), Log_2(128N/126C), Log_2(129N/126C), (Log_2(128C/126C), Log_2(130N/126C), Log_2(131N/126C), Log_2(131C/126C)) were normalized to the median for class (2) proteins, which was set to 0 (i.e. TMT ratios set to 1). To calculate optimal cut-offs, we then calculated the true positive rate (TPR) and false positive rate (FPR) we would obtain if we retained only proteins above that TMT ratio. We defined TPR as the fraction of class (1) proteins above the TMT ratio in question, and FPR as the fraction of class (2) above the TMT ratio in question. We selected TMT ratios that maximize the difference between TPR and FPR as our cutoffs (Supplementary Figure 9c).

To select cutoffs for proteins biotinylated by the indicated ERM-ligase over proteins biotinylated by the corresponding cytosol-targeted ligase, we classified the detected proteins into three groups:

1. ERM proteins (Supplementary Table 2 Tab 1; true positive list of 90 well-established ERM proteins[33]).
2. non-secretory proteins, (Supplementary Table 2 Tab 3; false positive list of 7421 human proteins that are not predicted to be secretory by Phobius[49] or are not annotated with the following Gene Ontology[38,39] terms: GO:0005783, GO:0005789, GO:0007029, GO:0030867, GO:0048237, GO:0061163, GO:0016320, GO:0030868, GO:0006983, GO:0000139, GO:0051645, GO:0031985, GO:0005796, GO:0005795, GO:0005794, GO:0007030, GO:0090168, GO:0005886, GO:0007009, GO:1903561, GO:0070062, GO:0005576, GO:0031012, GO:0005615, GO:0005769, GO:0035646, GO:0005765, GO:0090341, GO:0090340, GO:0005635, GO:0007084, GO:0007077, GO:0006998, GO:0051081, GO:0005641, GO:0031965, GO:0005637, GO:0071765, GO:0048471, GO:1905719, GO:0031982, GO:0006906, GO:0048278, GO:0032587, GO:0016021, GO:0005887, GO:0005768, GO:0071816, GO:0031526, GO:0005913, GO:0072546, GO:1990440, GO:0030968, GO:1902236, GO:1990441, GO:0034976, GO:0005788, GO:0005790, GO:1902237, GO:0070059, GO:0005786, GO:0005793, GO:0044322, GO:0098554, GO:0005791, GO:1902010, GO:0043001, GO:0005802, GO:0006888, GO:0006890, GO:0005801, GO:0012510, GO:0006892, GO:0042147, GO:0034499, GO:0032588, GO:0006895, GO:0030140, GO:0051684, GO:0000042, GO:0032580, GO:0030173, GO:0006891, GO:0030198, GO:0031668, GO:0010715, GO:0035426, GO:1903053, GO:1903551, GO:0005578, GO:1903055, GO:0001560, GO:0022617, GO:0006887, GO:0012505.
3. all other proteins.

We then normalized the TMT ratios in order to account for differences in total protein quantity between samples within the TMT 11-plex experiment. To do this, the Log_2(TMT ratios) corresponding to ERM-ligase/ligase-NES (Log_2(127N/127C), Log_2(128N/127C), Log_2(129N/129C), (Log_2(128C/129C), Log_2(130N/130C), Log_2(131N/130C), Log_2(131C/129C)) were normalized to the median for class (2) proteins, which was set to 0 (i.e. TMT ratios set to 1). To calculate optimal cut-offs, we then calculated the true positive rate (TPR) and false positive rate (FPR) we would obtain if we retained only proteins above that TMT ratio. We defined TPR as the fraction of class (1) proteins above the TMT ratio in question,

and FPR as the fraction of class (2) above the TMT ratio in question. We selected TMT ratios that maximize the difference between TPR and FPR as our cutoffs (Supplementary Figure 9c).

After applying both cutoffs to each experimental replicate, we then intersected both filtered replicates to produce the final proteomes (Supplementary Table 5 Tabs 1-3). Overlap of proteins between proteomes obtained with BioID, TurboID 10 minute labeling, and TurboID 1 hour labeling are shown in Supplementary Figure 9g, Supplementary Table 5 Tab 4.

To assess the specificity of our proteomes, we determined the secretory specificity of the respective proteomes (Figure 2e). To calculate specificity, we report the percentage of proteins present in Supplementary Table 2 Tab 4, a list of 11,838 human proteins with secretory annotation according to Phobius[49], the Human Protein Atlas[50] (protein localized to endoplasmic reticulum, Golgi apparatus, plasma membrane, vesicles, nuclear membrane, cell junctions; or predicted membrane proteins and predicted secreted proteins), the Plasma Proteome Database[51], literature (reference cited in table), or are annotated with the following Gene Ontology[38,39] terms: GO:0005783, GO:0005789, GO:0007029, GO:0030867, GO:0048237, GO:0061163, GO:0016320, GO:0030868, GO:0006983, GO:0000139, GO:0051645, GO:0031985, GO:0005796, GO:0005795, GO:0005794, GO:0007030, GO:0090168, GO:0005886, GO:0007009, GO:1903561, GO:0070062, GO:0005576, GO:0031012, GO:0005615, GO:0005769, GO:0035646, GO:0005765, GO:0090341, GO:0090340, GO:0005635, GO:0007084, GO:0007077, GO:0006998, GO:0051081, GO:0005641, GO:0031965, GO:0005637, GO:0071765, GO:0048471, GO:1905719, GO:0031982, GO:0006906, GO:0048278, GO:0032587, GO:0016021, GO:0005887, GO:0005768, GO:0071816, GO:0031526, GO:0005913, GO:0072546, GO:1990440, GO:0030968, GO:1902236, GO:1990441, GO:0034976, GO:0005788, GO:0005790, GO:1902237, GO:0070059, GO:0005786, GO:0005793, GO:0044322, GO:0098554, GO:0005791, GO:1902010, GO:0043001, GO:0005802, GO:0006888, GO:0006890, GO:0005801, GO:0012510, GO:0006892, GO:0042147, GO:0034499, GO:0032588, GO:0006895, GO:0030140, GO:0051684, GO:0000042, GO:0032580, GO:0030173, GO:0006891, GO:0030198, GO:0031668, GO:0010715, GO:0035426, GO:1903053, GO:1903551, GO:0005578, GO:1903055, GO:0001560, GO:0022617, GO:0006887, GO:0012505.

The specificity of the “entire human proteome” reported in Figure 2e was calculated as the percentage of human proteins that are not present in category (2) non-secretory proteins, i.e. proteins that are predicted to be secretory by Phobius[49], or are annotated with the following Gene Ontology[38,39] terms: GO:0005783, GO:0005789, GO:0007029, GO:0030867, GO:0048237, GO:0061163, GO:0016320, GO:0030868, GO:0006983, GO:0000139, GO:0051645, GO:0031985, GO:0005796, GO:0005795, GO:0005794, GO:0007030, GO:0090168, GO:0005886, GO:0007009, GO:1903561, GO:0070062, GO:0005576, GO:0031012, GO:0005615, GO:0005769, GO:0035646, GO:0005765, GO:0090341, GO:0090340, GO:0005635, GO:0007084, GO:0007077, GO:0006998, GO:0051081, GO:0005641, GO:0031965, GO:0005637, GO:0071765, GO:0048471, GO:1905719, GO:0031982, GO:0006906, GO:0048278, GO:0032587, GO:0016021, GO:0005887, GO:0005768, GO:0071816, GO:0031526, GO:0005913, GO:0072546, GO:1990440, GO:

0030968, GO:1902236, GO:1990441, GO:0034976, GO:0005788, GO:0005790, GO:1902237, GO:0070059, GO:0005786, GO:0005793, GO:0044322, GO:0098554, GO:0005791, GO:1902010, GO:0043001, GO:0005802, GO:0006888, GO:0006890, GO:0005801, GO:0012510, GO:0006892, GO:0042147, GO:0034499, GO:0032588, GO:0006895, GO:0030140, GO:0051684, GO:0000042, GO:0032580, GO:0030173, GO:0006891, GO:0030198, GO:0031668, GO:0010715, GO:0035426, GO:1903053, GO:1903551, GO:0005578, GO:1903055, GO:0001560, GO:0022617, GO:0006887, GO:0012505.

To calculate subsecretory specificity, we took a subset of proteins with the following Gene Ontology[38,39] terms: GO:0005783 for endoplasmic reticulum, GO:0005794 for Golgi apparatus, and GO:0005886 for plasma membrane and classified them according to this priority: endoplasmic reticulum>Golgi apparatus>plasma membrane (Supplementary Table 2 Tab 5). We then took the subset of proteins in the ERM proteomes with these GO terms and plotted their percentages in Figure 2f. To calculate ER specificity, the subset of proteins with GOCC[38,39] annotation for endoplasmic reticulum (GO:0005783) were subdivided into those with membrane annotation, soluble cytosolic annotation, or soluble luminal annotation according to GOCC[38,39], UniProt[48], TMHMM[52], or literature (Supplementary Table 2 Tab 6); these percentages are reported in Figure 2g.

To assess the recall of our proteomes for ERM proteins, we determined the coverage of our proteomes for lists of true positive ERM (Supplementary Figure 9f, Supplementary Table 2 Tab 1). In the scatter plot analyses shown in Supplementary Figure 9e, true positive ERM proteins (Supplementary Table 2 Tab 1) are shown in green, cytosolic proteins (Supplementary Table 2 Tab 7; human proteins with Gene Ontology[38,39] term GO:0005829 that lack annotated or predicted transmembrane domains according to UniProt[48] or TMHMM[33,52]) are shown in red, all other proteins are shown in black.

Mitochondrial matrix and nuclear proteomic data analysis

Complete mass spectrometry data for both the nucleus and mitochondrial matrix are shown in Supplementary Table 6 Tab 6 and Supplementary Table 7 Tab 6 respectively. Each of the two replicates for each proteomics experiment (mitochondrial matrix and nucleus) were analyzed separately. To select cutoffs for proteins biotinylated by the indicated ligase over non-specific bead binders, we classified the detected proteins into three groups:

(1) nuclear annotated proteins (Supplementary Table 3 Tab 1; true positive list of 6710 human proteins annotated with the following Gene Ontology[38,39] terms: GO:0016604, GO:0031965, GO:0016607, GO:0005730, GO:0001650, GO:0005654, GO:0005634).

(1) mitochondrial annotated proteins (Supplementary Table 4 Tab 1; true positive list of 1555 human proteins present in MitoCarta2.0[53] or annotated with the following Gene Ontology[38,39] term: GO:0005739, but excluding any proteins also present in category 2 (Supplementary Table 4 Tab 2).

(2) proteins with non-nuclear annotation (Supplementary Table 3 Tab 2; false positive list of 6815 human proteins annotated with the following Gene

Ontology[38,39] terms: GO:0015629, GO:0016235, GO:0030054, GO:0005813, GO:0045171, GO:0000932, GO:0005829, GO:0005783, GO:0005768, GO:0005929, GO:0005794, GO:0045111, GO:0005811, GO:0005764, GO:0005815, GO:0015630, GO:0030496, GO:0070938, GO:0005739, GO:0072686, GO:0005777, GO:0005886, GO:0043231; and are not annotated with the following Gene Ontology[38,39] terms: GO:0016604, GO:0031965, GO:0016607, GO:0005730, GO:0001650, GO:0005654, GO:0005634, “nucleus localization”, “nuclear envelope”, “nuclear matrix”, “nuclear chromatin”, “nuclear pore”, “nuclear inner membrane”, “nuclear chromosome”, “nuclear heterochromatin”, “nuclear euchromatin”, “nuclear inclusion body”).

(2) proteins with non-mitochondrial annotation (Supplementary Table 4 Tab 2; previously curated false positive list of 2410 human proteins that are not annotated to be mitochondrial[2,54]).

(3) all other proteins.

We then normalized the TMT ratios in order to account for differences in total protein quantity between samples within the TMT 6-plex experiments. To do this, the Log_2(TMT ratios) corresponding to ligase experimentals/negative control (Log_2(126/127), Log_2(128/129), Log_2(130/131), Log_2(129/127) for replicate 1, and (Log_2(130/131), Log_2(129/126), Log_2(127/126) Log_2(128/126) for replicate 2) were normalized to the median for class (2) proteins, which was set to 0 (i.e. TMT ratios set to 1). To calculate optimal cut-offs, we then calculated the true positive rate (TPR) and false positive rate (FPR) we would obtain if we retained only proteins above that TMT ratio. We defined TPR as the fraction of class (1) proteins above the TMT ratio in question, and FPR as the fraction of class (2) above the TMT ratio in question. We selected TMT ratios that maximize the difference between TPR and FPR as our cutoffs (Supplementary Figure 10f, g).

After applying cutoffs to each replicate, we then intersected both to produce the final proteomes (Supplementary Table 6 Tabs 1-3 for nuclear proteomes, Supplementary Table 7 Tabs 1-3 for mitochondrial matrix proteomes). Overlap of proteins between proteomes obtained with BioID, TurboID, and miniTurbo for both the nucleus and mitochondrial matrix are shown in Supplementary Figure 10k, Supplementary Table 6 Tab 4 for nucleus, and Supplementary Table 7 Tab 4 for mitochondrial matrix. To assess the specificity of our proteomes, we determined the nuclear and mitochondrial specificity of the respective proteomes (Figure 2h). To calculate specificity, we report the percentage of proteins present in class (I) (Supplementary Table 3 Tab 1 for nuclear specificity, Supplementary Table 4 Tab 1 for mitochondrial specificity). To assess the recall of our proteomes for known proteins of the respective compartment being mapped, we determined the coverage of our proteomes for lists of well-established nuclear or mitochondrial proteins (Supplementary Figure 10j). To calculate coverage of our nuclear proteome, we constructed a list of 230 proteins using Cell Atlas data and hyperLOPIT data[50] that have annotated nuclear detection by a validated antibody to nuclear bodies, nuclear membrane, nuclear speckles, nucleoli, fibrillary centers, nucleoplasm, or nucleus; and also have hyperLOPIT location annotated to nucleus, or nucleus-chromatin; and also have expression in HEK cells (Supplementary Table 3 Tab 3).

To calculate coverage of our mitochondrial matrix proteome, we used a previously curated list of 173 well-established mitochondrial matrix proteins[2] (Supplementary Table 4 Tab 3).

Generation of *UAS-ligase* transgenic *Drosophila* lines

V5-BioID, *V5-TurboID*, and *V5-miniTurboID* coding sequence was PCR amplified from CMV-plasmids using the same F and R primers:

V5-ligase_F: ccgcggccgccccttcaccATGGGCAAGCCCATCCCC

V5-ligase_R gggtcggcgcgcccaccttCTATTAGTCCAGGGTCAGGCG

DNA fragments were cloned into pEntr plasmids (Invitrogen) using Gibson assembly (NEB). *pEntr_V5-ligase* entry plasmids were recombined into *pWalium10-roe*[55] using Gateway LR Clonase II Enzyme (Invitrogen). *pWalium10-roe* contains 10x *UAS* enhancer elements for Gal4-controlled expression, *attB* sequence, and a *white*$^{+}$ transgene. Transgenic flies were generated using PhiC31 integration by injecting *pWalium10-V5-ligase* plasmids into flies carrying an *attP* docking site on chromosome III (*attP2*)[56]. Final fly strains are referred to as *UAS-V5-BioID*, *UAS-V5-TurboID*, and *UAS-V5-miniTurboID*.

Drosophila culture and genetics

Experiments on flies were performed with wild type or transgenic strains of *Drosophila melanogaster*. The age and sex of animals involved in experiments are indicated in figure legends and methods below. The Harvard Medical School Standing Committee on Animals (through the Office of the Institutional Animal Care and Use Committee (IACUC)) deems flies as invertebrates with limited sentience and therefore not subject to formal review and approval by the committee.

Crosses were maintained on standard fly food at 25˚C. For temporal expression experiments using *tub-Gal4*, *tubGal80*^{ts}, animals were kept at 18˚C during all developmental stages until transferred to 29˚C to induce gene expression. Biotin food was prepared by microwaving standard fly food until liquid and adding 1 mM biotin dissolved in H_2O to a final concentration of 100 μM.

Unless otherwise noted, fly stocks were obtained from the Bloomington Drosophila Stock Center and are listed with the corresponding stock number: *ptc-Gal4* (2017), *Act5c-Gal4/CyO* (4414), *nub-Gal4* (25754), *w1118* (6326), *tub-Gal80*^{ts}; *tub-Gal4/TM6b* (Perrimon Lab), *UAS-Luciferase* (35788), *Desat-Gal4* (Oenocyte) (65405), *repo-Gal4* (Glia) (7415), *Mef2-Gal4* (Muscle) (27390), *Lpp-Gal4* (Fat body) (Perrimon Lab, see transgene in 67043 for information), *elav-Gal4* (Neurons) (8760), *Myo1a-Gal4* (Gut) (Perrimon Lab, see transgene in 67057 for information), *Hml-Gal4* (Hemocytes) (30140).

Western blotting of *Drosophila* adults

For experiments Figure 3g and Supplementary Figure 12, adult flies were aged 3 days after eclosion from pupal cases (13 days old after egg deposition). For each condition, five females and five males were lysed in RIPA buffer (Thermo Fisher, 89900) on ice using a blue pestle in a microcentrifuge tube. Samples were centrifuged at 14,000 xg for 20 min at 4˚C. Supernatant was retained and transferred to a new centrifuge tube. Protein

concentration was calculated using a BCA kit (Pierce 23225) and RIPA buffer was added to samples to normalize to 4 μg/μL. Normalized protein samples were mixed with an equal volume of 4x SDS sample buffer and boiled for 5 minutes at 95˚C. 10 μg/sample was loaded onto a 4-20% Mini-PROTEAN TGX PAGE gel (Biorad 4561095), transferred to Immobilon-FL PVDF membrane (Millipore IPFL00010), incubated in PBS + 0.1% Tween (PBST) for 15 minutes, and blocked overnight in 3% BSA in PBST (PBST-BSA) at 4˚C. To detect biotinylated proteins, blots were incubated with 0.3 μg/mL streptavidin-HRP (Thermo Fisher S911) in PBST-BSA for 1 hour at room temperature. Blots were washed extensively with PBST and exposed using Pico Chemiluminescent Substrate (Thermo Fisher 34577). To detect expressed V5-tagged ligases, blots were incubated with 1:10,000 mouse anti-V5 (Invitrogen R960-25) with PBST-BSA overnight at 4˚C, washed with PBST, incubated with 1:5000 anti-mouse Alexa 800 (Thermo Fisher A32730), washed with PBST, and imaged on an Aerius Fluorescent imager (LI-COR 9250).

Immunohistochemistry of *Drosophila* wing discs

For Figure 3d, wandering 3rd instar larvae were bisected and inverted to expose the imaginal discs. Inverted carcasses were fixed for 20 min in 4% paraformaldehyde in 1x PBS. Fixed carcasses with attached wing discs were permeabilized with PBS + 0.1% Triton-X100 (PBST) for 20 min and blocked with PBST + 5% normal goat serum (PBST-NGS) for 1 hour. Blocked carcasses were incubated overnight at 4˚C in PBST-NGS with 1:500 mouse anti-V5 (Invitrogen R960-25) and 1:500 streptavidin-555 (Invitrogen S32355). Carcasses were washed 3x with PBST and incubated for 1 hour at room temperature in PBST-NGS with 1:500 anti-mouse Alexa 647 (Thermo Fisher A-21236) and 1:1000 DAPI (stock 1mg/ml). Samples were washed with three times with PBST, once with PBS, and equilibrated in 70% Glycerol/1x PBS. Wing discs were dissected away from the carcass and mounted onto glass slides with Vectashield mounting media (Vector Labs H-1000) and glass coverslip. Mounted samples were imaged on a Zeiss 780 confocal microscope.

Quantitation of fluorescence signal intensity from *Drosophila* wing discs in Figure 3e

Average signal intensity of fluorescence of streptavidin-555 in wing discs was measured using raw images obtained under identical confocal settings and under non-saturating exposure settings. Using ImageJ software, the polygon tool was used to select a rectangular region of the *ptc-Gal4* expressing domain in the wing pouch. The average signal intensity in this selected region was determined separately for the streptavidin-555 channel and the anti-V5 channel. The average signal intensity in control samples (very low background staining) was subtracted from signal intensity of experimental conditions (BioID, turboID, miniturboID). For each wing disc, the signal intensity of streptavidin-555 was normalized to the signal intensity of anti-V5 (streptavidin-555/anti-V5). Fold change was determined by normalizing streptavidin-555/anti-V5 values from TurboID and miniturboID to values from BioID. Measurements were taken from at least three wing discs for each condition.

Quantification of adult *Drosophila* wing size and survival after ligase expression during development in Supplementary Figure 13

UAS-V5-ligase transgenes were expressed during development by crossing with different Gal4-expressing lines and their effects on the adult assessed.

To determine if larval wing disc expression of ligases affects adult wing morphology, *nub-Gal4* was crossed with *UAS-V5- ligase* transgenes and the resulting progeny analyzed. *nub-Gal4* was crossed with wild-type flies (w^{1118}) as a negative control. Adult flies were aged 3 days after eclosion from pupal cases. Wings were removed from adults, placed in a drop of 50% Permount/50% Xylenes on a glass slide, and a coverslip added. Mounted wings were imaged using a light microscope with a 10x objective. Wing area was measured using the polygon selection tool in ImageJ. Wings quantified and imaged are from female flies.

To determine if developmental expression of ligases reduces survival to adulthood, we crossed *UAS- ligase* lines with different Gal4 lines that express in major tissue types (Muscle, Fat, Neurons, Glia, Gut, Oenocytes, Hemocytes) or ubiquitously (*Act5c-Gal4*). To quantify toxicity, we counted the number of surviving adult animals after undergoing ~10 days of development (from fertilized egg through pupal stages) expressing *UAS- ligase* under Gal4 control, and compared to the number of wild-type siblings. *UAS-Luciferase* was used as a negative control transgene, which is widely considered as non-toxic to cells.

As an example, the following crossing scheme was used for *Act5c-Gal4*:

P0 *Act5c-Gal4*/*CyO* x *UAS-V5- ligase (homozygous)*

Segregation of the *Act5c-Gal4* chromosome and *CyO* balancer chromosome results in two possible F1 progeny genotypes:

F1 (genotype 1) Act5c-Gal4/UAS-V5- ligase

F1 (genotype 2) CyO/UAS-V5- ligase

The *CyO* chromosome has a dominant *Cy* mutation that causes adult flies to have curly wings. Therefore genotype 1 flies have straight wings and express the ligase transgene, and genotype 2 have curly wings and do not express the transgene. The fraction of surviving flies expressing a given UAS-transgene is calculated as: # genotype 1/(# genotype 1 + # genotype 2)

For example, a survival fraction of 0.5 indicates that equal numbers of genotype 1 and genotype 2 were observed in the adult population, and that no reduction in survival from expressing a UAS-transgene during development occurred.

Similar crossing schemes were used for tissue-specific Gal4 lines. Gal4 lines that are normally maintained as a homozygous stock were first outcrossed to an appropriate balancer line to obtain Gal4/Balancer flies, which were then crossed with *UAS-Luciferase* or *UAS-TurboID*. Gal4 lines on chromosome II were used with a *CyO* balancer, and Gal4 lines on chromosome III were used with a *TM3, Sb* balancer.

Adult flies were aged >3 days after eclosion from pupal cases before being counted. Females and males of the same genotype were counted together.

For imaging whole adults, flies were frozen at -20˚C overnight and images of adult flies were obtained using a dissection microscope connected to a digital camera.

To determine if changes in the fraction of surviving flies were statistically significant, a two-sided Chi-square test was applied to the number of adult flies for genotype 1 and genotype 2, comparing *UAS-Luciferase* to a *UAS- ligase* transgene.

Mammalian cell viability assays

For each experiment presented in Supplementary Figure 14, five sterile, white, clear bottom 96-well plates were pre-coated with 100 μL 50 μg/mL human fibronectin in MEM for at least 20 minutes at 37°C under 5% CO_2. For transiently transfected samples, HEK 293T cells were plated in 6-wells and transfected at ~90% confluency using 0.8 μL Lipofectamine2000 (Life Technologies) and 200 ng plasmid in serum-free media (2.5 mL total volume) for 3-4 hr, after which time Lipofectamine-containing media was replaced with fresh serum-containing media. After ~2 hours, cells were trypsinized and seeded in triplicate into wells of each fibronectin-coated 96-well plate at 2000 cells/well in 50:50 serum-containing MEM:DMEM with or without 50 μM biotin. The stable cell lines were seeded into wells in the same manner. An additional triplicate of coated wells without cells served as background subtraction in each plate. One plate was immediately assayed after plating for cell viability by the CellTiter-Glo 2.0 Luminescent Viability assay (Promega). Subsequent plates were assayed at the indicated time points.

C. elegans strains and culture conditions

Experiments on *Caenorhabditis elegans* were performed with wild type (N2) or transgenic strains expressing extrachromosomal arrays. The age and sex of animals involved in experiments are indicated in figure legends and methods below. The Stanford's Administrative Panel on Laboratory Animal Care (APLAC) deems *C. elegans* used in this study as invertebrates and not subject to formal review and approval by the committee.

Unless otherwise noted, *C. elegans* strains were cultured and maintained at 20°C on *E. coli* OP50 bacteria as previously described[57]. To deplete the animals of excess biotin, worms were grown for 2 generations on biotin auxotrophic *E. coli* (MG1655bioB:kan)[58] and washed twice with 1X M9 solution. Biotin auxotrophic *E. coli* MG1655bioB:kan was kindly donated by Dr. John E. Cronan, University of Illinois. Embryos dissected from one day-old adults of the following genotypes were compared for this study: JLF289 (wowEx66[ges1p::3xHA:BioID::unc-54, myo-2p:mCherry::unc-54]), JLF290 (wowEx67[ges1p::3xHA:BioID::unc-54, myo-2p::mCherry::unc-54]), JLF291 (wowEx68[ges1p::3xHA::TurboID::unc-54, myo-2p::mCherry::unc-54]), JLF292 (wowEx69[ges1p::3xHA::TurboID::unc-54, myo-2p::mCherry::unc-54]), JLF293 (wowEx70[ges1p::3xHA::miniTurbo::unc-54, myo-2p:mCherry::unc-54]), JLF294 (wowEx71[ges1p::3xHA::miniTurbo::unc-54, myo-2p:mCherry::unc-54]).

Transgenic ligase strain construction for *C. elegans*

C. elegans codon-optimized ligase genes BioID and TurboID (containing the 3 worm introns present in GFP) and miniTurbo (containing 2 worm introns present in GFP) were synthesized (IDT) and inserted into pJF241 to produce plasmids pAS28, pAS31, and pAS32, respectively. Transgenic worms were generated by injecting 50ng/μL ligase gene and 2.5ng/μL of the co-injection marker myo-2p::mCherry into day 1 N2 hermaphrodites.

Western blotting of *C. elegans* adults

Ligase expression and biotinylation (Figure 3i, Supplementary Figure 15g) were assessed by Western blotting one day-old adult worm lysates. For each condition, 50 N2 hermaphrodites (wild-type) or worms expressing a ligase transgene were transferred to Eppendorf tubes containing 1mL of M9 and washed once. Excess M9 was removed until ~50μL of M9 remained and an additional 50μL of 4x sample buffer was added. Worms were boiled at 95°C for 10 min, vortexed 10 seconds, and centrifuged at 13,000 x g for 5 min at 4°C. Equal volume of lysate was loaded onto a 4-20% Mini-PROTEAN TGX PAGE gel (BioRad), transferred to a nitrocellulose membrane (0.4 μm, BioRad), and stained with Ponceau S solution. Blots were blocked in 5% milk PBST solution, probed with anti-HA (1:5000, rat monoclonal, Roche) and anti-tubulin (1:5000, rat monoclonal, Abcam) primary antibodies, and detected with secondary antibody (1:5000, goat anti-rat IRDye 680RD, Licor) and streptavidin-IRDye (1:5000, 800CW, Licor). Blots were imaged on LI-COR Odyssey CLx.

Immunohistochemistry and microscopy of *C. elegans*

To visualize ligases and biotinylation (Figure 3j), embryos were isolated from one day-old adults, fixed, and stained as previously described[59]. Briefly, embryos were attached to poly-lysine coated microscope slides with Teflon spacers. Slides were frozen on dry ice and embryos were permeabilized by freeze-crack and fixed in 100% MeOH for 5 minutes at -20° C. Embryos were washed in PBS then PBST, and subsequently incubated in anti-HA primary antibody (Abcam, 1:200) overnight at 4°C to visualize ligase expression. Embryos were washed in PBST and then incubated in CY3-anti-mouse secondary antibody (Jackson Immunoresearch Laboratories, 1:200) Streptavidin Alexa Fluor 488 (Invitrogen, 1:200), and DAPI (Sigma, 1:10000). Embryos were mounted in Vectashield (Vector Laboratories) and stored at 4°C. Samples were imaged using 405 nm, 488 nm, and 561 nm lasers, a Yokogawa X1 confocal spinning disk head, and a 60x PLAN APO oil objective (NA=1.4) on a Nikon Ti-E inverted microscope (Nikon Instruments) equipped with a 1.5x magnifying lens. Images were captured using NIS Elements software (Nikon) and an Andor Ixon Ultra back thinned EM-CCD camera, at a sampling rate of 0.5μm. All samples were imaged with the same camera and laser settings, with the exception of embryos expressing miniTurbo. To avoid pixel saturation, a 25% reduction in camera exposure was used to capture the streptavidin-AF488 signal in miniTurbo expressing embryos. Thus, the miniTurbo images and quantifications in Figure 3 for streptavidin-AF488 are an underrepresentation of the signal resulting from biotinylation. Images were processed and assembled in NIS Elements and Adobe InDesign. In Figure 3j, images shown are maximum intensity projections of two Z-slices with brightness adjusted for visual clarity.

Quantitation of fluorescence signal intensity in *C. elegans* intestine

Bean and comma stage embryos were chosen for analyses. For each embryo, one slice of the Z-stack was used for analysis. A custom Python script including the modules scikit-image, matplotlib, and NumPy combined with ImageJ was used to analyze *C. elegans* imaging data. A threshold for the HA:ligase signal was calculated by the Otsu method to create a mask to isolate the intestine region for each embryo. Background streptavidin-AF488 signal for each embryo was determined by drawing a square in the anterior portion of the embryo outside of

the intestine and calculating the average pixel intensity within that square. The average background pixel intensity of streptavidin-AF488 was then subtracted from the average streptavidin-AF488 pixel intensity within the isolated intestine region, and the resulting corrected average was plotted for each embryo (Figure 3k). To measure the ratio of streptavidin-AF488 to HA:ligase pixel intensities, each pixel value for streptavidin-AF488 within the isolated intestine region was corrected for background and then divided by its corresponding HA:ligase pixel value. Then the average of the ratio values for each embryo was plotted (Supplementary Figure 15). Statistical significance was determined using the Mann-Whitney U test (Fig 3k). Samples were blinded for statistical analysis.

Quantitation of *C. elegans* viability in Supplementary Figure 16

C. elegans strains expressing ligase variants were maintained at 20°C on either biotin+ (OP50) or biotin- (MG1655(bioB:kan)) *E. coli*. For BioID and TurboID in each bacteria condition, a one-day old adult worm expressing the ligase transgene from an extrachromosomal array and a sibling one-day old adult worm lacking the transgene (control) were placed on separate plates containing the appropriate bacteria. For each plate, the adult was removed after laying eggs for 4 hours and the remaining embryos on the plate were counted. Three days later the number of living worms were counted and viability was calculated by dividing the number of hatched worms by the number of eggs that were initially laid. Worms were kept on biotin+ or biotin- bacteria for two generations. Developmentally delayed worms were defined as worms that were larval stage or non-gravid at the time of counting.

Statistics

Figure 3 d, e, wing disc imaging results are representative of at least 10 discs present on the microscope slide, and at least 3 of which were imaged. Sample sizes (n) in e from left column to right are 5, 6, 3. Error bars were calculated using s.e.m. This experiment was performed twice with similar results. Supplementary Figure 13c, sample sizes (n) from left column to right are 17, 14, 17, 15, 19, 18, 19, 18. Error bars were calculated using s.e.m. This experiment was performed twice with similar results. Supplementary Figure 13d, for each food type, a two-sided Chi-square test was used to determine if the difference in proportions (measure of effect) of UAS-ligase transgenes to UAS-Luciferase was statistically significant. Sample size values (n) from left column to right: 512, 586, 466, 563, 286, 524, 513, 459. 95% confidence intervals (CI) for the difference in proportions of BioID, TurboID, and miniTurboID compared to Luciferase were -0.02945 to 0.09046, 0.2757 to 0.384, and -0.02958 to 0.09206 for normal food, and -0.04278 to 0.104, -0.05999 to 0.08728, and -0.07191 to 0.07838 for biotin food. p-values for all datapoints - Columns 1/2: 0.3113, Columns 1/3: <0.0001, Columns 1/4: 0.3027, Columns 5/6: 0.4109, Columns 5/7: 0.718, Columns 5/8: 0.9369. This experiment was performed twice with similar results. Supplementary Figure 13f, Data was analyzed as in Supplementary Figure 13d. Sample size values (n) from left column to right: 350, 367, 196, 339, 203, 284, 194, 287, 214, 232, 215, 305, 240, 346. 95% confidence intervals (CI) for each Gal4 line from left to right were -0.06874 to 0.0807, 0.04977 to 0.2228, -0.07546 to 0.1081, -0.03849 to 0.148, - 0.07737 to 0.1115, 0.1398 to 0.3063, -0.07858 to 0.08951. p-values for all datapoints - Oenocytes:

0.8719, Glia: 0.0017, Muscle: 0.7295, Fat body: 0.2333, Neurons: 0.7143, Gut: <0.0001, Hemocytes: 0.8917. This experiment was performed twice with similar results.

Figure 3i, experiment was repeated 5 times with similar results, with the exception that miniTurbo expression was detectable in only 2 of those 5 replicates. Embryo imaging results shown in Figure 3j and Supplementary Figure 15a are representative images of complete quantitative data shown in Figure 3k. Figure 3k samples sizes (n) from left column to right column are 26, 18, 11, 16, 25, 8, 19, 23, 14, 14, 23, 9. Statistical significance was assessed via Mann-Whitney U test (two-sided). Error bars were calculated using s.e.m. Supplementary Figure 15c-f, statistical significance was assessed via Mann-Whitney U test (two-sided). Error bars were calculated using s.e.m. Supplementary Figure 16a-c, sample sizes (n) are indicated within an in column titled 'Replicates'. Statistical significance was assessed via Mann-Whitney U test (two-sided). Error bars were calculated using s.e.m.

For further detail on the experimental design, reagents, statistics, reproducibility, software, and data collection methods used in this study, please refer to the Life Sciences Reporting Summary.

Code Availability

The Python script used for *C. elegans* imaging analysis is available from the corresponding author upon reasonable request.

Data Availability

Source data for Figure 2e and Supplementary Figures 8c-g are provided in the paper in Supplementary Table 5. Source data for Figure 2h and Supplementary Figures 9f-k are provided in the paper in Supplementary Tables 6 and 7. The original mass spectra may be downloaded from MassIVE (http://massive.ucsd.edu) using the identifier: MSV000082304. The data is directly accessible via ftp://massive.ucsd.edu/MSV000082304. Any additional data that support the findings of this study are available from the corresponding author upon reasonable request.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

FACS was performed at the Koch Institute Flow Cytometry Core (MIT) and Stanford Shared FACS Facility. S. Han (Stanford) synthesized neutravidin-AlexaFluor647. S. Ax (Stanford) cloned the cell surface TurboID and miniTurbo constructs. We are grateful to I. Droujinine (Harvard) for advice on biotin labeling in *D. melanogaster*. Biotin auxotrophic *E. coli* MG1655bioB:kan was kindly donated by J. Cronan (University of Illinois). This work was supported by NIH R01-CA186568 (to A. Y. T.), Howard Hughes Medical Institute Collaborative Innovation Award (to A. Y. T., S. C., and N.P.), and NIH New Innovator Award DP2GM119136 (to J. L. F.). T.C.B. was supported by Dow Graduate Research and Lester Wolfe Fellowships. J.A.B. was supported by a Damon Runyon Post-Doctoral Fellowship. A.D.S. was supported by NIH Training Grant 2T32GM007276.

References

1. Kim DI, Roux KJ. Filling the Void: Proximity-Based Labeling of Proteins in Living Cells. Trends in Cell Biology. 2016; 26:804–817. [PubMed: 27667171]
2. Rhee HW, et al. Proteomic Mapping of Mitochondria in Living Cells via Spatially Restricted Enzymatic Tagging. Science. 2013; 339:1328–1331. [PubMed: 23371551]
3. Lam SS, et al. Directed evolution of APEX2 for electron microscopy and proximity labeling. Nat Methods. 2014; 12:51–54. [PubMed: 25419960]
4. Choi-Rhee E, Schulman H, Cronan JE. Promiscuous protein biotinylation by *Escherichia coli* biotin protein ligase. Protein Sci. 2004; 13:3043–50. [PubMed: 15459338]
5. Roux KJ, Kim DI, Raida M, Burke B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. J Cell Biol. 2012; 196:801–810. [PubMed: 22412018]
6. Paek J, et al. Multidimensional Tracking of GPCR Signaling via Peroxidase-Catalyzed Proximity Labeling. Cell. 2017; 169:338–349.e11. [PubMed: 28388415]
7. Lobingier BT, et al. An Approach to Spatiotemporally Resolve Protein Interaction Networks in Living Cells. Cell. 2017; 169:350–360.e12. [PubMed: 28388416]
8. Kaewsapsak P, Shechner DM, Mallard W, Rinn JL, Ting AY. Live-cell mapping of organelle-associated RNAs via proximity biotinylation combined with protein-RNA crosslinking. eLife. 2017; 6
9. Martell JD, et al. Engineered ascorbate peroxidase as a genetically encoded reporter for electron microscopy. Nat Biotechnol. 2012; 30:1143–1148. [PubMed: 23086203]
10. Gupta GD, et al. A Dynamic Protein Interaction Landscape of the Human Centrosome-Cilium Interface. Cell. 2015; 163:1483–1499.
11. Kim DI, et al. Probing nuclear pore complex architecture with proximity-dependent biotinylation. Proc Natl Acad Sci. 2014; 111:E2453–E2461. [PubMed: 24927568]
12. Lin Q, et al. Screening of Proximal and Interacting Proteins in Rice Protoplasts by Proximity-Dependent Biotinylation. Front Plant Sci. 2017; 8
13. Morriswood B, et al. Novel bilobe components in *Trypanosoma brucei* identified using proximity-dependent biotinylation. Eukaryot Cell. 2013; 12:356–367. [PubMed: 23264645]
14. Chen AL, et al. Novel components of the toxoplasma inner membrane complex revealed by BioID. MBio. 2015; 6
15. Nadipuram SM, et al. *In vivo* biotinylation of the toxoplasma parasitophorous vacuole reveals novel dense granule proteins important for parasite growth and pathogenesis. MBio. 2016; 7
16. Chen AL, et al. Novel insights into the composition and function of the *Toxoplasma* IMC sutures. Cell Microbiol. 2017; 19
17. Long S, et al. Calmodulin-like proteins localized to the conoid regulate motility and cell invasion by *Toxoplasma gondii*. PLoS Pathog. 2017; 13
18. Zhou Q, Hu H, Li Z. An EF-hand-containing protein in *Trypanosoma brucei* regulates cytokinesis initiation by maintaining the stability of the cytokinesis initiation factor CIF1. J Biol Chem. 2016; 291:14395–14409. [PubMed: 27226595]
19. Dang HQ, et al. Proximity interactions among basal body components in *Trypanosoma brucei* identify novel regulators of basal body biogenesis and inheritance. MBio. 2017; 8
20. Kehrer J, Frischknecht F, Mair GR. Proteomic Analysis of the *Plasmodium berghei* Gametocyte Egressome and Vesicular bioID of Osmiophilic Body Proteins Identifies Merozoite TRAP-like Protein (MTRAP) as an Essential Factor for Parasite Transmission. Mol Cell Proteomics. 2016; 15:2852–2862. [PubMed: 27371728]
21. Gaji RY, et al. Phosphorylation of a Myosin Motor by TgCDPK3 Facilitates Rapid Initiation of Motility during *Toxoplasma gondii* egress. PLoS Pathog. 2015; 11
22. Batsios P, Ren X, Baumann O, Larochelle D, Gräf R. Src1 is a Protein of the Inner Nuclear Membrane Interacting with the *Dictyostelium* Lamin NE81. Cells. 2016; 5:13.
23. Meyer I, et al. CP39, CP75 and CP91 are major structural components of the *Dictyostelium* centrosome's core structure. Eur J Cell Biol. 2017; 96:119–130. [PubMed: 28104305]

24. Uezu A, et al. Identification of an elaborate complex mediating postsynaptic inhibition. Science. 2016; 353:1123–1129. [PubMed: 27609886]
25. Opitz N, et al. Capturing the Asc1p/*R* eceptor for *A* ctivated *C K* inase *1* (RACK1) Microenvironment at the Head Region of the 40S Ribosome with Quantitative BioID in Yeast. Mol Cell Proteomics. 2017; 16:2199–2218. [PubMed: 28982715]
26. Kim DI, et al. An improved smaller biotin ligase for BioID proximity labeling. Mol Biol Cell. 2016; 27:1188–1196. [PubMed: 26912792]
27. Ramanathan M, et al. RNA–protein interaction detection in living cells. Nat Methods. 2018; doi: 10.1038/nmeth.4601
28. Birendra KC, et al. VRK2A is an A-type lamin-dependent nuclear envelope kinase that phosphorylates BAF. Mol Biol Cell. 2017; mbc.E17-03-0138. doi: 10.1091/mbc.E17-03-0138
29. Redwine WB, et al. The human cytoplasmic dynein interactome reveals novel activators of motility. eLife. 2017; 6
30. Jung EM, et al. Arid1b haploinsufficiency disrupts cortical interneuron development and mouse behavior. Nat Neurosci. 2017; 20:1694–1707. [PubMed: 29184203]
31. Martell JD, et al. A split horseradish peroxidase for the detection of intercellular protein–protein interactions and sensitive visualization of synapses. Nat Biotechnol. 2016; 34:774–780. [PubMed: 27240195]
32. Bobrow MN, Harris TD, Shaughnessy KJ, Litt GJ. Catalyzed reporter deposition, a novel method of signal amplification application to immunoassays. J Immunol Methods. 1989; 125:279–285. [PubMed: 2558138]
33. Hung V, et al. Proteomic mapping of cytosol-facing outer mitochondrial and ER membranes in living human cells by proximity biotinylation. eLife. 2017; 6
34. Dingar D, et al. BioID identifies novel c-MYC interacting partners in cultured cells and xenograft tumors. J Proteomics. 2015; 118:95–111. [PubMed: 25452129]
35. Reinke AW, Balla KM, Bennett EJ, Troemel ER. Identification of microsporidia host-exposed proteins reveals a repertoire of rapidly evolving proteins. Nat Commun. 2017; 8:14023. [PubMed: 28067236]
36. Reinke AW, Mak R, Troemel ER, Bennett EJ. In vivo mapping of tissue- and subcellular-specific proteomes in *Caenorhabditis elegans*. Sci Adv. 2017; 3:e1602426. [PubMed: 28508060]
37. Chen CL, et al. Proteomic mapping in live Drosophila tissues using an engineered ascorbate peroxidase. Proc Natl Acad Sci U S A. 2015; 112:1–6.
38. Gene Ontology Consortium. Gene Ontology Consortium: going forward. Nucleic Acids Res. 2015; 43:D1049–56. [PubMed: 25428369]
39. Ashburner M, et al. Gene Ontology: tool for the unification of biology. Nat Genet. 2000; 25:25–29. [PubMed: 10802651]
40. Chao G, et al. Isolating and engineering human antibodies using yeast surface display. Nat Protoc. 2006; 1:755–768. [PubMed: 17406305]
41. Jan CH, Williams CC, Weissman JS. Response to Comment on ‘Principles of ER cotranslational translocation revealed by proximity-specific ribosome profiling’. Science (80-). 2015; 348:1217–1217.
42. Colby DW, et al. Engineering antibody affinity by yeast surface display. Methods in Enzymology. 2004; 388:348–358. [PubMed: 15289082]
43. Ausubel FM, et al. Current Protocols in Molecular Biology. Molecular Biology. 2003; 1
44. Wood ZA, Weaver LH, Brown PH, Beckett D, Matthews BW. Co-repressor induced order and biotin repressor dimerization: A case for divergent followed by convergent evolution. J Mol Biol. 2006; 357:509–523. [PubMed: 16438984]
45. Xu Y, Beckett D. Evidence for interdomain interaction in the Escherichia coli repressor of biotin biosynthesis from studies of an N-terminal domain deletion mutant. Biochemistry. 1996; 35:1783–1792. [PubMed: 8639659]
46. Eginton C, Cressman WJ, Bachas S, Wade H, Beckett D. Allosteric coupling via distant disorder-to-order transitions. J Mol Biol. 2014; 427:1695–1704.

47. Hung V, et al. Spatially resolved proteomic mapping in living cells with the engineered peroxidase APEX2. Nat Protoc. 2016; 11:456–475. [PubMed: 26866790]
48. Apweiler R, et al. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 2017; 45:D158–D169. [PubMed: 27899622]
49. Käll L, Krogh A, Sonnhammer ELL. A combined transmembrane topology and signal peptide prediction method. J Mol Biol. 2004; 338:1027–1036. [PubMed: 15111065]
50. Thul PJ, et al. A subcellular map of the human proteome. Science (80-). 2017; 356:eaal3321.
51. Muthusamy B, et al. Plasma proteome database as a resource for proteomics research. Proteomics. 2005; 5:3531–3536. [PubMed: 16041672]
52. Krogh A, Larsson B, Von Heijne G, Sonnhammer ELL. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. J Mol Biol. 2001; 305:567–580. [PubMed: 11152613]
53. Calvo SE, Clauser KR, Mootha VK. MitoCarta2.0: An updated inventory of mammalian mitochondrial proteins. Nucleic Acids Res. 2016; 44:D1251–D1257. [PubMed: 26450961]
54. Pagliarini DJ, et al. A Mitochondrial Protein Compendium Elucidates Complex I Disease Biology. Cell. 2008; 134:112–123. [PubMed: 18614015]
55. Perkins LA, et al. The transgenic RNAi project at Harvard medical school: Resources and validation. Genetics. 2015; 201:843–852. [PubMed: 26320097]
56. Markstein M, Pitsouli C, Villalta C, Celniker SE, Perrimon N. Exploiting position effects and the gypsy retrovirus insulator to engineer precisely expressed transgenes. Nat Genet. 2008; 40:476–483. [PubMed: 18311141]
57. Brenner S. The genetics of Caenorhabditis elegans. Genetics. 1974; 77:71–94. [PubMed: 4366476]
58. Sulston JE, Schierenberg E, White JG, Thomson JN. The embryonic cell lineage of the nematode Caenorhabditis elegans. Developmental Biology. 1983; 100:64–119. [PubMed: 6684600]
59. Leung B, Hermann GJ, Priess JR. Organogenesis of the Caenorhabditis elegans Intestine. Dev Biol. 1999; 216:114–134. [PubMed: 10588867]

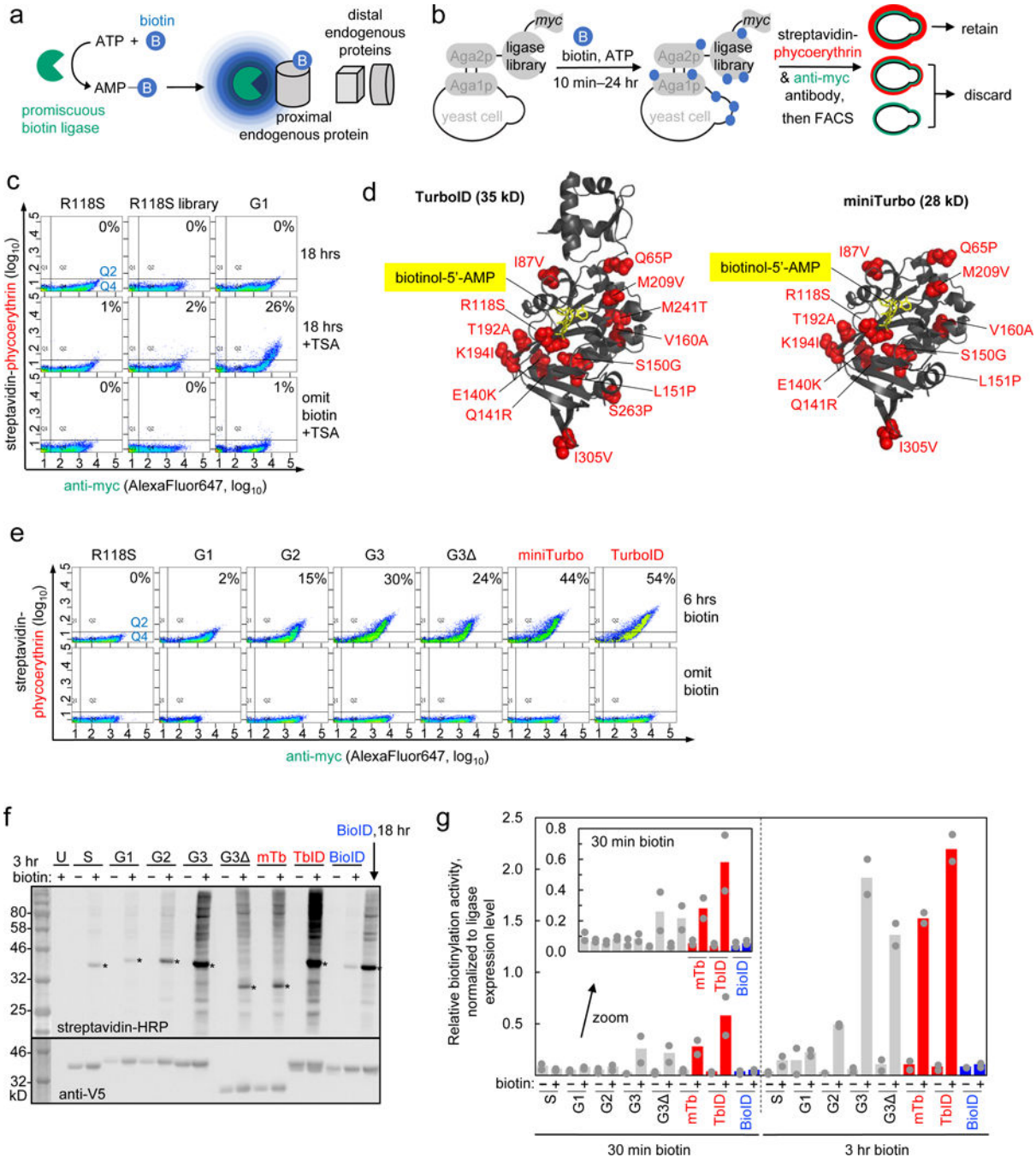

Figure 1. Directed evolution of TurboID

(**a**) Proximity-dependent biotinylation catalyzed by promiscuous biotin ligases. Ligases catalyze the formation of biotin-5′-AMP anhydride, which diffuses out of the active site to biotinylate proximal endogenous proteins on nucleophilic residues such as lysine. (**b**) Yeast display-based selection scheme. A >10^7 library of ligase variants is displayed on the yeast surface as a fusion to mating protein Aga2p. All ligases have a C-terminal myc epitope tag. Biotin and ATP are added to the yeast library for between 10 minutes and 24 hours. Ligase-catalyzed promiscuous biotinylation is detected by staining with streptavidin-phycoerythrin

and ligase expression is detected by staining with anti-myc antibody. Two-dimensional FACS sorting enables enrichment of cells displaying a high ratio of streptavidin to myc staining. (**c**) Tyramide signal amplification (TSA)[32] improves biotin detection sensitivity on the yeast surface. In the top row, the three indicated yeast samples (G1 is the winning ligase mutant from the first generation of evolution) were labeled with exogenous biotin for 18 hours then stained for FACS as in (b). The y-axis shows biotinylation extent, and the x-axis quantifies ligase expression level. In the second row, after 18 hours of biotin incubation, yeast were stained with streptavidin-HRP, reacted with biotin-phenol[2,32] to create additional biotinylation sites, then stained with streptavidin-phycoerythrin and anti-myc antibody before FACS. The third row omits biotin. Percentage of cells in upper right quadrant (Q2/(Q2+Q4)) shown in top right of each graph. This experiment was performed once, but each yeast sample has been analyzed under identical conditions at least twice in separate experiments with similar results. (**d**) *E. coli* biotin ligase structure (PDB: 2EWN) with sites mutated in TurboID (left) and miniTurbo (right) shown in red. The N-terminal domain (aa1-63) is also removed from miniTurbo. A non-hydrolyzable analog of biotin-5′-AMP, biotinol-5′-AMP, is shown in yellow stick. (**e**) FACS plots summarizing progress of directed evolution. G1-G3 are the winning clones from generations 1-3 of directed evolution. G3Δ has its N-terminal domain (aa1-63) deleted. Omit biotin samples were grown in biotin-deficient media (see **Methods**) for the entire induction period (~18-24 hr). This experiment was performed twice with similar results, except G3Δ omit biotin, which was performed once. (**f**) Comparison of ligase variants in the HEK cytosol showing that TurboID and miniTurbo are much more active than BioID, as well as the starting template and various intermediate clones from the evolution. Indicated ligases were expressed as NES (nuclear export signal) fusions in the HEK cytosol. 50 μM exogenous biotin was added for 3 hours, then whole cell lysates were analyzed by streptavidin blotting. Ligase expression detected by anti-V5 blotting. U, untransfected. S, BirA-R118S. Asterisks indicate ligase self-biotinylation. BioID labeling for 18 hours (50 μM biotin) shown for comparison in the last lane. This experiment was performed twice with similar results. (**g**) Quantitation of streptavidin blot data in (f) and from a 30 minute labeling experiment shown in Supplementary Figure 4b. Quantitation excludes self-biotinylation band. Sum intensity of each lane is divided by the sum intensity of the ligase expression band; ratios are normalized to that of BioID/18 hours, which is set to 1.0. Grey dots indicate quantitation of signal intensity from each replicate, colored bars indicate mean signal intensity calculated from the two replicates.

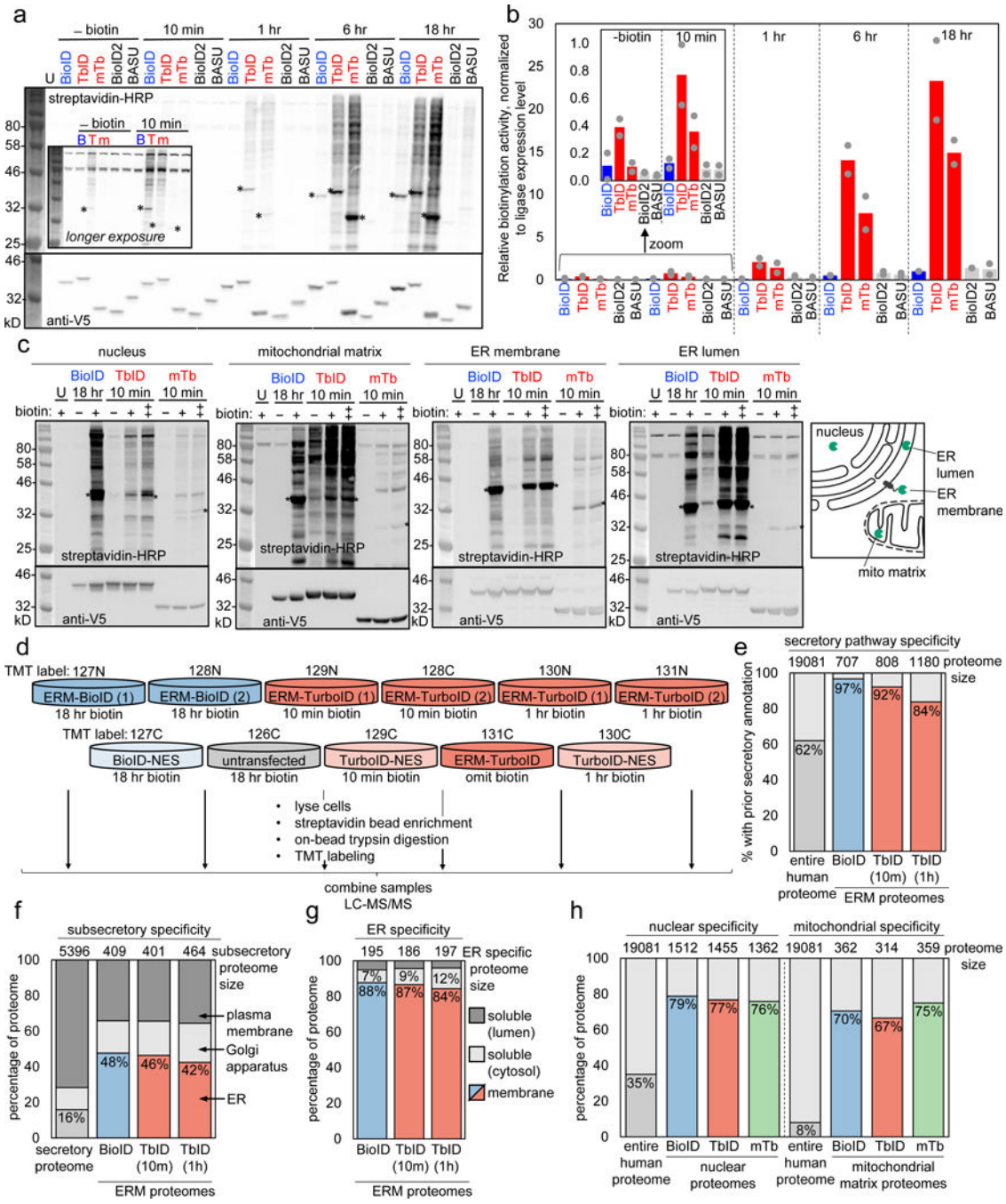

Figure 2. Characterization of TurboID and miniTurbo in mammalian cells

(**a**) Comparison of TurboID and miniTurbo to three other promiscuous ligases (BioID[5], BioID2[26], and BASU[27]) in the cytosol of HEK 293T cells. Here, 500 μM exogenous biotin was used for labeling, whereas 50 μM was used in Supplementary Figure 6c-e. Streptavidin-HRP blotting detects promiscuously biotinylated proteins, and anti-V5 blotting detects ligase expression. U, untransfected. Asterisks denote ligase self-biotinylation bands. This experiment was performed twice with similar results. (**b**) Quantitation of experiment in (a). For shorter timepoints (-biotin and 10 min), we used a longer-exposure image of the same

blot, shown in Supplementary Figure 6a; for longer timepoints (1, 6, 18 hr), we used a shorter-exposure image of the blot in (a), shown in Supplementary Figure 6b. Quantitation performed as in Figure 1g. Grey dots indicate quantitation of signal intensity from each replicate, colored bars indicate mean signal intensity calculated from the two replicates. (**c**) Comparison of promiscuous ligases in multiple HEK organelles. Each ligase was fused to a peptide targeting sequence (see Supplementary Table 8) directing them to the locations indicated in the scheme at right. BioID samples were treated with 50 μM biotin for 18 hours. TurboID and miniTurbo samples were labeled for 10 minutes with 50 (+) or 500 (++) μM biotin. U, untransfected. Asterisks denote ligase self-biotinylation. This experiment was performed five times for nuclear constructs, three for mitochondrial constructs, four times for ER membrane constructs, and twice for ER lumen constructs with similar results. (**d**) Mass spectrometry-based proteomic experiment comparing TurboID and BioID on the ER membrane (ERM), facing cytosol. Experimental design and labeling conditions. Ligase fusion constructs were stably expressed in HEK 239T. BioID samples were treated with 50 μM biotin for 18 hours, while TurboID samples were treated with 500 μM biotin for 10 minutes or 1 hr. After labeling, cells were lysed and biotinylated proteins were enriched with streptavidin beads, digested to peptides, and conjugated to TMT (tandem mass tag) labels. All 11 samples were then combined and analyzed by LC-MS/MS. This experiment was performed once with two replicates per condition. (**e**) Specificity analysis for proteomic datasets derived from experiment in (d). Size of each ERM proteome at top. Bars show percentage of each proteome with prior secretory pathway annotation, according to GOCC, Phobius, human protein atlas, human plasma proteome database, and literature (see **Methods** and Supplementary Table 2 Tab 4). (**f**) Same as (e), except for each ERM proteome, we analyze the subset with ER, Golgi, or plasma membrane annotation. Annotations from GOCC were assigned in the priority order: ER>Golgi>plasma membrane (see **Methods** and Supplementary Table 2 Tab 5). (**g**) Breakdown of ER proteins enriched by TurboID and BioID, by transmembrane or soluble. Soluble proteins were further divided into luminal or cytosol-facing. Annotations obtained from GOCC, UniProt, TMHMM, and literature (see **Methods** and Supplementary Table 2 Tab 6). (**h**) Characterization of nuclear and mitochondrial matrix proteomes obtained via BioID (18 hour), TurboID (10 min), and miniTurbo (10 min)-catalyzed labeling. Proteome sizes across top. Bars show fraction of each nuclear (left) or mitochondrial (right) proteome with prior nuclear or mitochondrial annotation, according to GOCC, MitoCarta, or literature (see **Methods** and Supplementary Table 3 Tab 1, Supplementary Table 4, Tab 1). Design of proteomic experiment shown in Supplementary Figure 10a, proteomic lists in Supplementary Tables 6-7; further analysis of proteome data in Supplementary Figure 10.

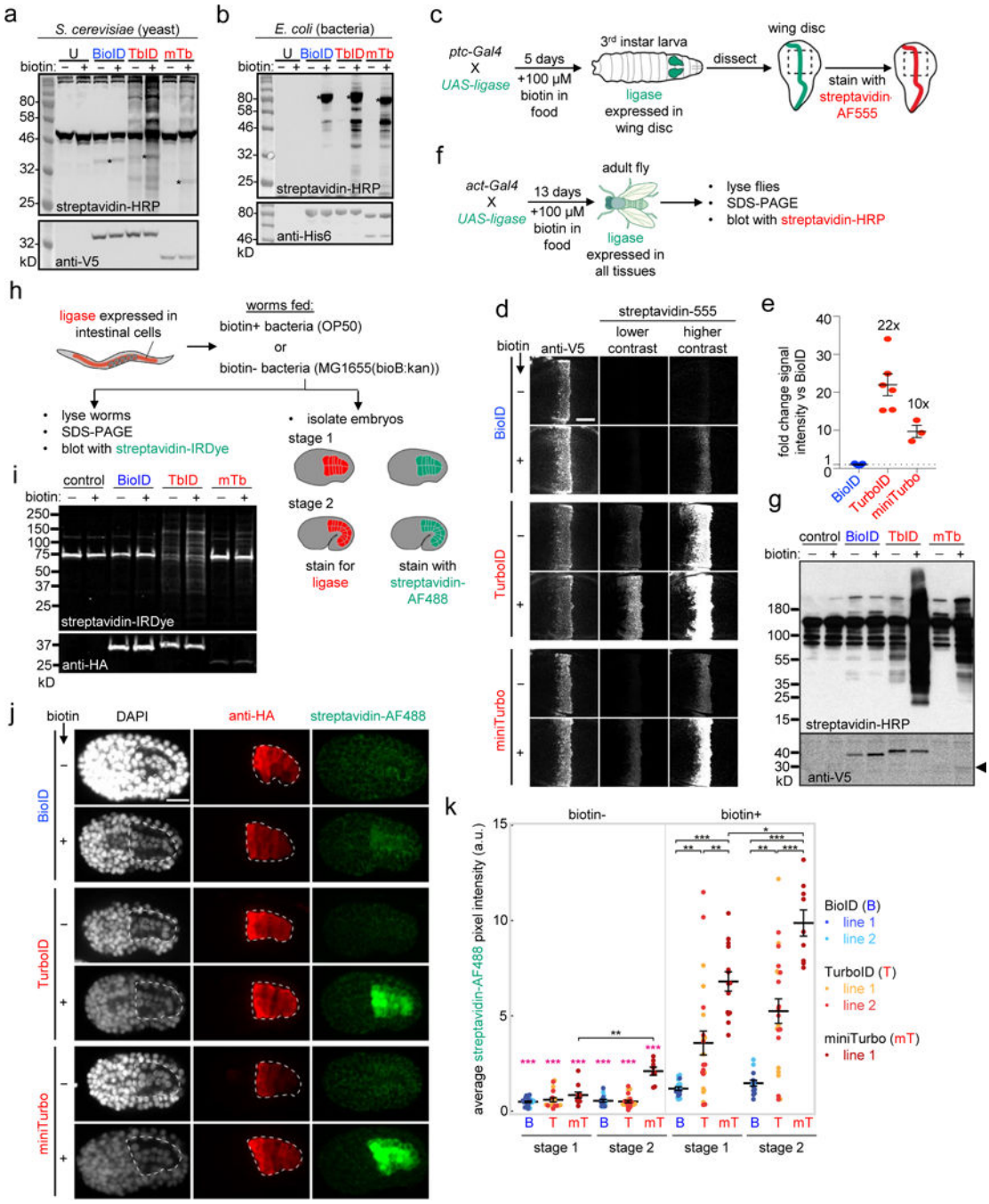

Figure 3. TurboID and miniTurbo in flies, worms, and other species

(**a**) Comparison of ligases in yeast. EBY100 *S. cerevisiae* expressing BioID, TurboID, or miniTurbo in the cytosol were treated with 50 μM biotin for 18 hours. Whole cell lysates were blotted with streptavidin-HRP to visualize biotinylated proteins, and anti-V5 antibody to visualize ligase expression. U, untransfected. Asterisks denote ligase self-biotinylation. Bands in untransfected lane are endogenous naturally-biotinylated proteins. This experiment was performed twice with similar results. (**b**) Comparison of ligases in *E. coli.* Ligases, fused at their N-terminal ends to His6-maltose binding protein, were expressed in the cytosol

of BL21 *E. coli* and 50 μM exogenous biotin was added for 18 hours. Whole cell lysates were analyzed as in (a). This experiment was performed twice with similar results. (**c**) – (**g**) Comparison of ligases in flies. (**c**) Scheme for tissue-specific expression of ligases in the wing disc of *D. melanogaster*. *ptc-Gal4* induces ligase expression in a strip of cells within the wing imaginal disc that borders the anterior/posterior compartments. (**d**) Imaging of larval wing discs after 5 days of growth on biotin-containing food. Biotinylated proteins are detected by staining with streptavidin-AlexaFluor555, and ligase expression is detected by anti-V5 staining. Panels show the pouch region of the wing disc, indicated by the dashed line in (c). Scale bar, 40 μm. Each experimental condition has at least three technical replicates; one representative image is shown. This experiment was independently repeated two times with similar results. (**e**) Quantitation of streptavidin-AlexaFluor555 signal intensities in (d). Error bars, s.e.m. Average fold-change shown as text above bars. Sample size values (n) from left column to right: 5, 6, 3. (**f**) Scheme for ubiquitous expression of ligases in flies, at all developmental timepoints, via the *act-Gal4* driver. (**g**) Western blotting of fly lysates prepared as in (f). Biotinylated proteins detected by blotting with streptavidin-HRP, ligase expression detected by anti-V5 blotting. In control sample, *act-Gal4* drives expression of *UAS-luciferase.* Bands in control lanes are endogenous naturally-biotinylated proteins. This experiment was performed twice with similar results. (**h**) – (**k**) Comparison of ligases in worms. (**h**) Scheme for tissue-specific expression of ligases in *C. elegans* intestine via *ges-1p* promoter. Transgenic strains are fed either biotin-producing *E. coli OP50* (biotin+), or biotin-auxotrophic *E. coli MG1655bioB:kan* (biotin-). Promoter *ges-1p* drives ligase expression approximately 150 minutes after the first cell cleavage. (**i**) Adult worms prepared as in (h) were shifted to 25°C for one generation, then lysed and analyzed by Western blotting. Control worms (N2) do not express ligase. Anti-HA antibody detects ligase expression. Streptavidin-IRDye detects biotinylated proteins. This experiment was performed five times (n = 5). In biotin+ conditions, BioID biotinylation activity was undetectable and TurboID gave robust biotinylation signal (n = 5/5). Despite high activity detected by immunofluorescence in embryos, we only detected some low level of biotinylation by miniTurbo in adults (n = 2/5), likely due to its low expression levels. (**j**) Representative images of bean stage worm embryos (stage 1) from (h). See Supplementary Figure 15a for representative images of comma stage worm embryos (stage 2). Embryos were fixed and stained with streptavidin-AF488 to detect biotinylated proteins, and anti-HA antibody to detect ligase expression. Intestine is outlined by a white dotted line. Scale bar, 10 μm. Quantitation of multiple replicates shown in (k). (**k**) Quantitation of streptavidin-AF488 signal acquired from IF staining of embryonic stages 1 and 2 shown in (j) and Supplementary Figure 15a. Mean streptavidin pixel intensities for each embryo assessed are plotted for BioID (B), TurboID (T), and miniTurbo (mT). Two independent transgenic lines for BioID and TurboID and one for miniTurbo were assessed. Number of embryos imaged (n) from left to right: 26, 18, 11, 16, 25, 8, 19, 23, 14, 14, 23, 9. Statistical significance via Mann-Whitney U test (two-sided). ***$p \leq 0.0001$, **$p \leq 0.001$, *$p \leq 0.01$. Pink asterisks indicate significance of pairwise comparisons between biotin- and corresponding biotin+ treated embryos. Mean (reported in Supplementary Figure 15b) is shown as a black horizontal line for each condition, and error bars indicate s.e.m. Note that the streptavidin-AF488 pixel intensities for miniTurbo are an underrepresentation of the signal as camera

exposure settings were lowered to avoid pixel saturation (see **Methods**). See Supplementary Figure 15 for more details.