Supporting Information

Chen et al. 10.1073/pnas.1515623112

SI Methods

Generation of APEX *Drosophila* Lines. Plasmids encoding APEX were obtained from Martell et al. (7). APEX is wild-type APX with three engineered mutations (K41D, W41F, E112K). The sequence encoding APEX, in-frame with the sequence encoding a signal peptide at the N terminus (Fig. 1), was cloned into the Gateway vector pENTR (Invitrogen) and subcloned into pTWE or pTWF vectors. Signal peptides used in this study are NLS (nuclear localization signal) (3): PKKKRKV; NES (nuclear export signal) (4): LALKLAGLDI; and mito (mitochondrial signal peptide) (5): N-terminal 29 aa of human COXVIII. Plasmids were injected into embryos to generate transgenic lines that carry the UAS-APEX construct. The UAS/Gal4 system (2) was used for overexpression studies using *Dmef2-Gal4* (27) and *ptc-Gal4* (28, 29) drivers. All crosses were maintained at 25 °C.

Immunostaining and Confocal Fluorescence Imaging. S2R+ cells or dissected fly tissues were fixed with 4% (wt/vol) paraformaldehyde in PBS (phosphate-buffered saline) buffer for 30–60 min on ice. After washing with PBT (0.3% TritonX-100 in PBS buffer) and blocking with PBTN [2% normal donkey serum in PBT], samples were incubated in primary antibody overnight at 4 °C. Samples were washed extensively and incubated with Alexa-conjugated streptavidin (Invitrogen; 1:500) or with Alexa-conjugated secondary antibodies (Molecular Probes; 1:500). Nuclei were visualized by DAPI staining (1 μg/mL). To visualize biotinylated proteins, 5% (wt/vol) dialyzed BSA in PBS was used for blocking and diluting Alexa-conjugated streptavidin. The following primary antibodies were used: mouse anti-ATP5α (MitoSciences; 1:500), mouse anti-Flag (Sigma; 1:250), rabbit anti-Myc (Cell Signaling; 1:500), and mouse anti-HA (Roche; 1:500). Images were captured with a Zeiss LSM 780 laser scanning confocal microscope.

Western Blotting. Fly tissues were lysed in RIPA buffer (50 mM Tris·Cl, pH 7.4, 150 mM NaCl, 0.5% Triton X-100, 0.1% SDS). Protein concentration was measured by Pierce 660-nm protein assay (Thermo Scientific). After homogenization, debris was removed by centrifuging once at 20,000 × *g* for 10 min. Western blotting was performed using standard protocols. The following primary antibodies were: mouse anti-ATP5α (MitoSciences; 1:1,000); rabbit anti-Flag (Sigma; 1:5,000); and mouse anti-GFP (Molecular Probes; 1:5,000). Membranes labeled with primary antibodies were incubated with anti-rabbit HRP-conjugated antibody (Amersham; 1:5,000) or anti-mouse (Amersham; 1:5,000). For imaging biotinylation, blocking was performed with 5% (wt/vol) BSA in TBST (50 mM Tris, 150 mM NaCl, 0.1% Tween 20) at 4 °C overnight, and streptavidin-conjugated HRP (Invitrogen; 1:50,000) was diluted in blocking buffer.

DAB Staining and EM. Larval muscle cells were dissected and fixed with 2% (wt/vol) glutaraldehyde (Electron Microscopy Sciences) in buffer (100 mM sodium cacodylate, 2 mM $CaCl_2$, pH 7.4) on ice for 60 min. After fixation, samples were rinsed for 2 min five times in chilled buffer and then treated for 5 min in buffer containing 20 mM glycine to quench unreacted glutaraldehyde, followed by five 2-min rinses in chilled buffer. SIGMAFAST DAB (3,3′-diaminobenzidine tetrahydrochloride) with metal enhancer tablets (Sigma-Aldrich) was dissolved in 5 mL buffer and added to the samples for 10–30 min. The reaction was stopped by removing the DAB solution and washing the samples for 2 min five times with chilled buffer. Tissues were postfixed with 1% osmium tetroxide/1.5% (wt/vol) potassium ferrocyanide ($KFeCN_6$) in chilled buffer for 30 min. Cells were washed three times in water, incubated in 1% aqueous uranyl acetate for 30 min followed by two washes in water, and subsequently dehydrated in grades of alcohol [50%, 70%, 95%, and twice 100% (vol/vol)] for 5 min each, incubated in 100% propylene oxide for 30 min, and infiltrated overnight in a 1:1 mixture of propyleneoxide and TAAB Epon (Marivac). The samples were subsequently embedded in TAAB Epon and polymerized at 60 °C for 48 h. Embedded samples were cut into 60-nm sections using a Reichert Ultracut S microtome. Sections were stained with lead citrate and images were taken using a JEOL 1200-EX transmission electron microscope operating at 80 kV with an AMT 2k CCD camera.

APEX Labeling of Fly Tissues. Fly tissues were dissected in PBS and incubated with 500 μM biotin-phenol in PBS for 30 min at room temperature. After incubation, the substrate-containing solution was removed. To activate APEX enzyme for protein labeling, 1 mM H_2O_2 in PBS was added to the samples for 1 min. To stop the labeling reaction, the samples were washed three times with PBS with radical quenchers and peroxidase inhibitor (1 mM sodium ascorbate, 2 mM Trolox, 5 mM sodium azide). Samples were fixed for immunostaining as previously described or lysed for Western blots and further proteomic analysis.

Enrichment of Biotinylated Proteins with Streptavidin Beads. Each cell lysate (250 μg) of fly muscles (at 500 ng/μL protein concentration) was mixed with 500 μL of streptavidin-coated magnetic bead slurry (Pierce) that was prewashed twice with RIPA buffer. The mixtures were incubated at room temperature for 1 h with rotation. After incubation, the beads were washed twice with 1 mL RIPA buffer, once with 1 mL of 2 M urea in 10 mM Tris·HCl (pH 8.0), and twice with 1 mL RIPA buffer. On-bead digestion was subsequently performed by incubating the beads in trypsin solution (80 μL 1 mM DTT, 5 μg/mL trypsin in 2 M urea in 50 mM Tris, pH 8) overnight to retrieve peptides of biotinylated proteins from the beads for the following analysis. The supernatant was removed and reduced (4 mM DTT), alkylated (10 mM iodoacetamide) for 30 min each, and digested with 0.5 μg of trypsin overnight (37 °C).

iTRAQ Labeling of Peptides. The resulting digested peptides were desalted and reconstituted in 30 μL iTRAQ reconstitution buffer. Four-plex iTRAQ labeling was conducted per the manufacturer's instructions (SCIEX). Briefly, iTRAQ labels were reconstituted with ethanol to a final volume of 145 μL, followed by individual labeling at room temperature for 1 h by adding 140 μL iTRAQ reagent to the samples. Labels were used as follows: 114 for wild-type control, 115 for Gal4 control, 116 for mito-APEX replicate A, and 117 for mito-APEX replicate B. Label incorporation was evaluated on an Orbitrap before quenching with 100 mM (final) Tris for 10 min at room temperature.

Fractionation and Mass Spectrometry. Labeled peptides were separated by StageTip SCX fractionation into three fractions using a protocol adapted from Rappsilber et al. (30). Briefly, StageTips were prepared containing two C18 material discs (3M) above three SCX discs and were conditioned with MeOH, washed with 80% acetonitrile, 0.5% acetic acid followed by 500 mM NH_4AcO, 0.1% NH_4OH, 20% acetonitrile and equilibrated with 0.5% acetic acid. Peptides were loaded, washed in 0.5% acetic acid, transeluted to the SCX discs with 80% acetonitrile, 0.5% acetic acid, and stepwise eluted for collection using three elution buffers (50 mM NH_4AcO, 20% acetonitrile, pH 5.5, 8.5, and 11,

respectively). All StageTip steps were done in 100-μL volumes at 3,000 × *g* for 2 min. Following fractionation, samples were desalted and reconstituted in 3% acetonitrile, 0.1% formaldehyde. Fractionated peptides were analyzed by a data-dependent method where the top 12 most abundant precursors were selected for MS/MS prior to being placed on an exclusion list, a Thermo Scientific Q Exactive coupled to a Proxeon UHPLC using the same MS parameters as Rhee et al. (1). Briefly, peptides were separated over a 180-min gradient using a heated PicoFrit (New Objective) column (50C) packed with 20 cm of 1.9-μm C18 material (Dr. Maisch). Data were searched with Spectrum Mill (Agilent) using the UniProt *Drosophila* database. A fixed modification of carbamidomethylation of cysteine and variable modifications of N-terminal protein acetylation, oxidation of methionine, and 4-plex iTRAQ labels were searched. The enzyme specificity was set to trypsin, allowing cleavages N-terminal to proline, and a maximum of two missed cleavages was used for searching. The maximum precursor-ion charge state was set to 6. The precursor mass tolerance and MS/MS tolerance were set to 20 ppm. The peptide and protein false discovery rates were set to 0.01, and the minimum peptide length was set to 6. The raw mass spectrometry data and the sequence database used for searches may be downloaded from MassIVE (massive.ucsd.edu/) using the identifier MSV000079107. Download this dataset directly from ftp://MSV000079107:a@massive.ucsd.edu.

Determination of the Cutoff Point for Matrix Proteome Analysis and Assembly of Positive and Negative Control Lists. Protein-level information was obtained from Spectrum Mill. Only proteins identified by >1 unique peptide with quantified ratios were retained for further analysis. Identified peptides with UniProt accession numbers were mapped to FlyBase gene identifiers using an ID mapping tool at UniProt that correspond to 2,222 unique annotated *Drosophila* genes (FlyBase release 5.54). Protein iTRAQ ratios were normalized to the median value of each channel. The $\log_2$ ratio of two replicates of mito-APEX samples to two control samples were calculated and normalized to the median value of each channel, respectively. All of the genes identified by iTRAQ along with their annotation are listed in Dataset S1.

To determine the cutoff of the iTRAQ ratio for the dataset, proteins were cross-referenced to positive and negative control lists. The positive control list was assembled from both human and *Drosophila* data. The list of human mitochondrial genes was assembled from three resources: (*i*) genes based on the corresponding subcellular location annotation from UniProt and/or cellular compartmental annotation from Gene Ontology annotation; (*ii*) genes from a specialized database, MitoCarta (31); and (*iii*) genes published by Rhee et al. (1). Genes identified by at least two resources were selected and then mapped to orthologous genes in *Drosophila* using DIOPT (21) with the least-stringent filter. The list of *Drosophila* mitochondrial genes was assembled from two resources: (*i*) genes based on the corresponding subcellular location annotation from UniProt and/or cellular compartmental annotation from Gene Ontology annotation; and (*ii*) genes from MitoDrome (19). Next, mitochondrial lists from both human and fly data were integrated and compared. There are 1,290 genes on the final positive control list, of which 552 genes are present in both species and 100 genes are present in *Drosophila* only, whereas 638 genes were mapped from human data/annotation. Similarly, the negative control list is assembled from a false positive list of 2,410 nonmitochondrial proteins compiled by Vamsi Mootha's laboratory (31) and *Drosophila* proteins annotated with nucleus, endosome, endoplasmic reticulum (ER), Golgi, lysosome, or peroxisome subcellular annotations. There are 1,223 genes on the final negative control list, of which 423 genes have annotations in both species, 517 genes have annotations from *Drosophila* only, and 283 genes were mapped from human annotation.

The false positive rate (FPR) is calculated as a function of iTRAQ ratio using the equation (1)

$$FPR = \frac{\mathrm{P}(\text{iTRAQ ratio}|\text{false positive})}{\mathrm{P}(\text{iTRAQ ratio}|\text{mito positive})}.$$

The denominator is the conditional probability of finding a known mitochondrial protein in this range, which is calculated as the percentage of proteins on the positive control list in this range over all proteins identified on the positive control list. The numerator is the conditional probability of finding a false positive protein in a particular iTRAQ ratio range. The result calculated using this equation represents the percentage of false positive proteins in this iTRAQ ratio range over the total false positive proteins identified. We plotted FPR over iTRAQ ratio range (Fig. S4) and selected the iTRAQ ratio cutoff based on an FPR of 0.1, which means that a protein is 10 times more likely to be a true mitochondrial protein than a false positive. We set the cutoff points for the four datasets of iTRAQ ratios (116/114, 116/115, 117/114, and 117/115). The genes that were above the cutoff in all of the four datasets were selected as our final matrix proteome (Dataset S1).

COMPLEAT Analysis. The bioinformatics tool COMPLEAT (12) (www.flyrnai.org/compleat/) was used to identify complexes enriched among the genes identified by APEX. The analysis was done using the average $\log_2$ ratio of each gene with complex size ranging from 5 to 100 and identified 69 complexes with *P* value <0.01 as well as complex IQM (interquartile mean) score >0.585 (Dataset S3). Representative complexes were manually selected for display (Fig. S4).

Cell Culture and Transfection. S2R+ cells were transfected with overexpression constructs generated by Guruharsha et al. (13) using Effectene (Invitrogen). After 24 h, cells were incubated with 500 μM $CuSO_4$ for 48 h to induce expression. Cells were fixed and immunostainings were performed as previously described.

Assembly of the MitoMax Database. To build MitoMax, a comprehensive database for *Drosophila* mitochondrial genes with subcompartmental annotation, genes identified from isolation-based studies and/or APEX labeling were combined and integrated with genes from annotation (see assembly of positive control list) as well as *Drosophila* genes annotated at MitoMiner (18) and MitoDrome (19). We annotated the genes encoding proteins localized or exposed to the mitochondrial matrix based on APEX data and Gene Ontology. Score 4 is assigned to genes identified by multiple experiments. Score 3 is assigned to genes identified by one experiment but which are also annotated by TargetP prediction. Genes with score 4 or 3 are considered to be high confidence mitochondrial genes, which can be used as a gold standard reference set for fly mitochondrial genes. In addition, score 2 is assigned to genes from annotation only, whereas score 1 is assigned to genes identified by one experiment without any other evidence.

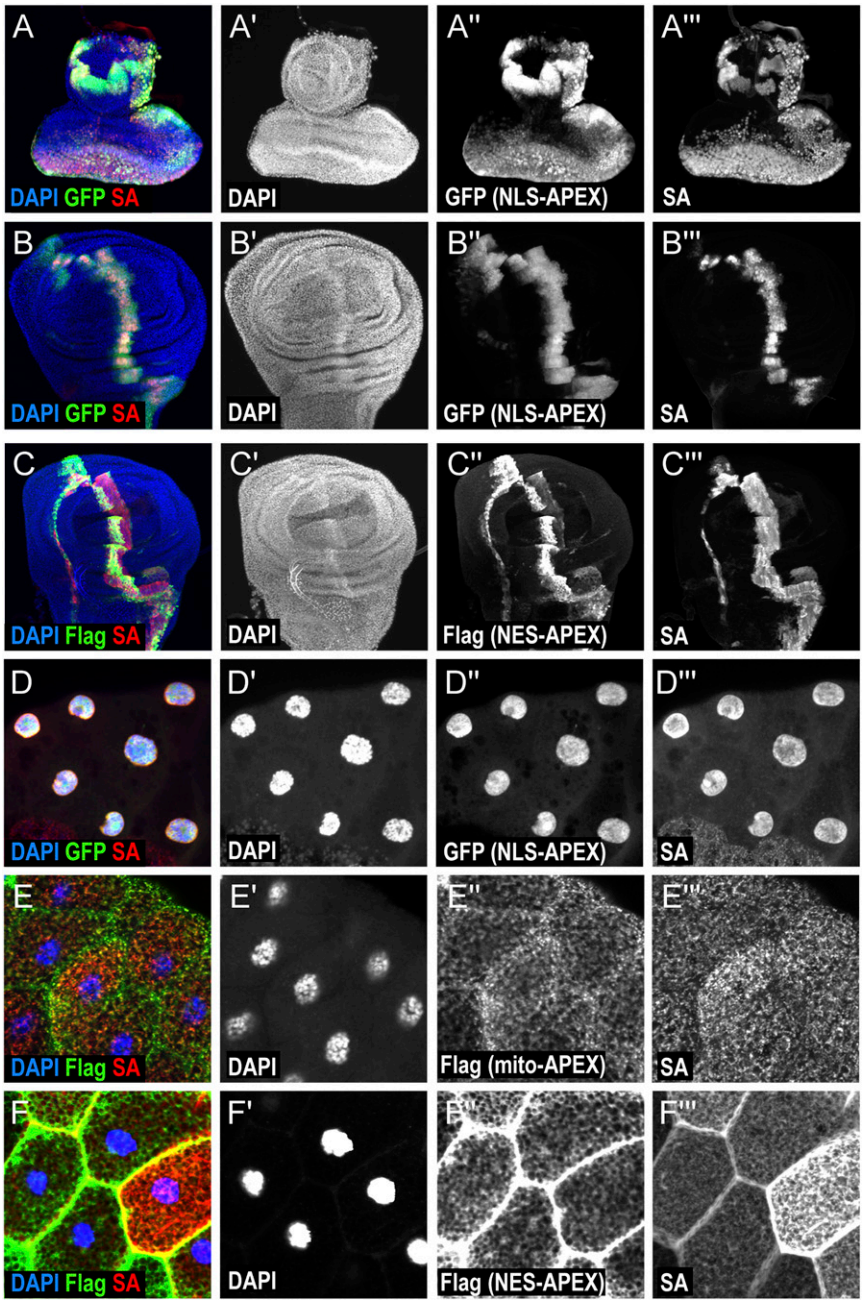


Fig. S1. Characterization of APEX labeling in *Drosophila* tissues. Expression and labeling activities of APEX constructs targeted to different subcellular localizations. APEX is expressed in larval tissues using *ptc-Gal4*. (*A*) Immunostainings of eye imaginal disc overexpressing NLS-APEX-GFP. (*B* and *C*) Immunostainings of wing imaginal disc overexpressing NLS-APEX-GFP (*B*) and NES-APEX-Flag (*C*). (*D–F*) Immunostainings of salivary glands overexpressing NLS-APEX-GFP (*D*), mito-APEX-Flag (*E*), and NES-APEX-Flag (*F*). Nuclei are identified by DAPI staining (blue in *A–F*; gray in *A′–F′*). APEX is detected by GFP expression or Flag staining (green in *A–F*; gray in *A′′–F′′*). Biotinylated proteins are visualized by streptavidin (SA)-coupled fluorescent staining (red in *A–F*; gray in *A′′′–F′′′*).

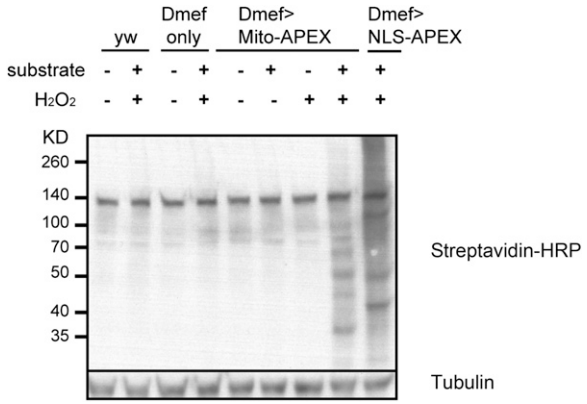


Fig. S2. Western blotting analysis of biotinylation in fly tissues. Genotypes and chemical supplements are as indicated. The provided substrate is biotin-phenol. Endogenously biotinylated proteins are shown in the negative controls: wild-type (yw) cells or *Dmef-gal4*–only cells without incubating with substrate and H_2O_2. Providing both substrate and H_2O_2 to either mito-APEX– or NLS-APEX–expressing cells gives rise to specific and strong biotinylation.

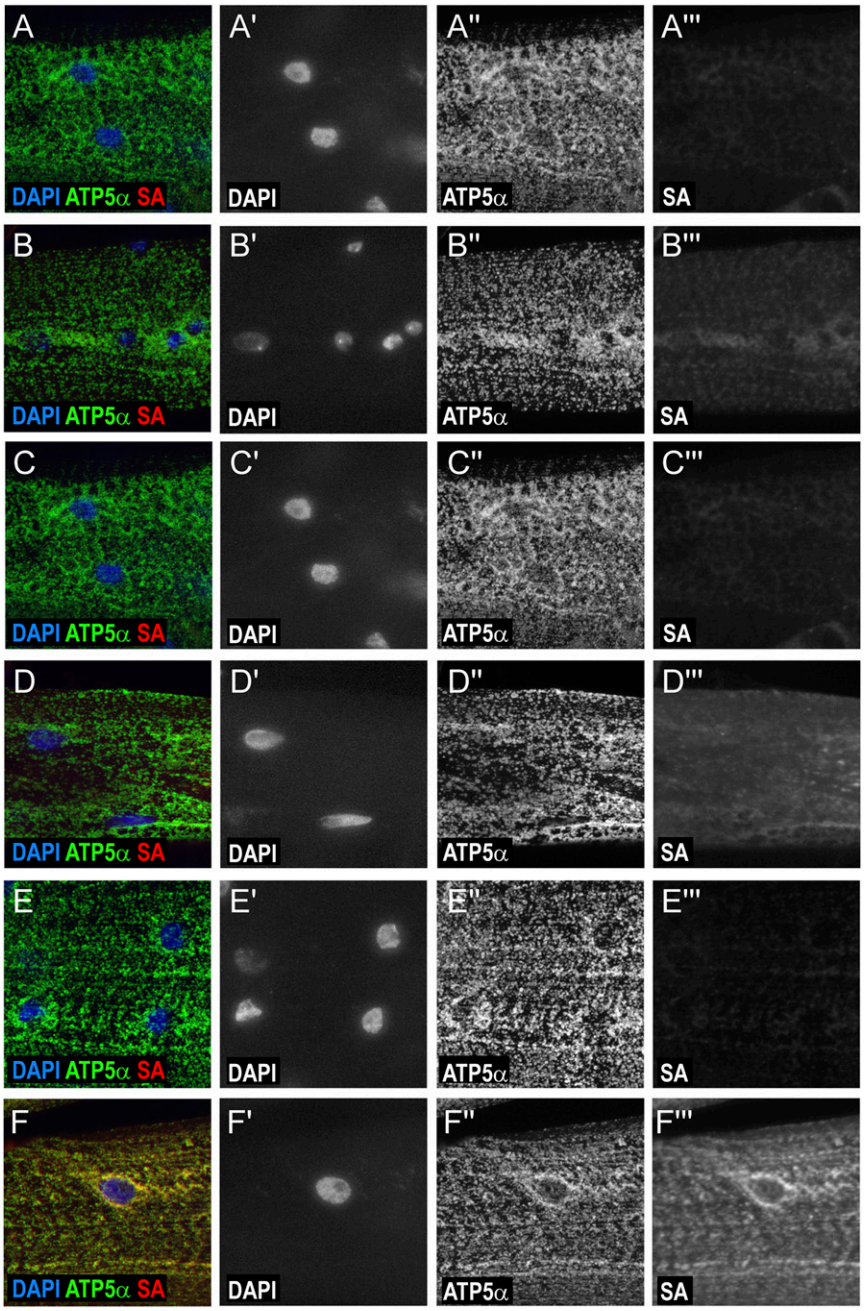
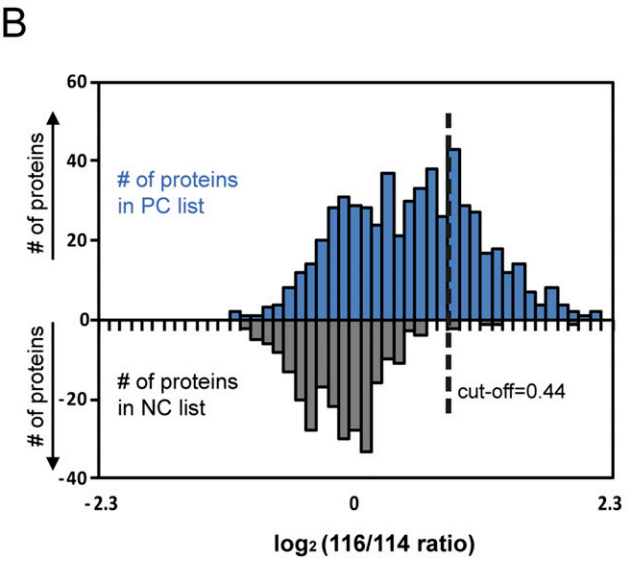
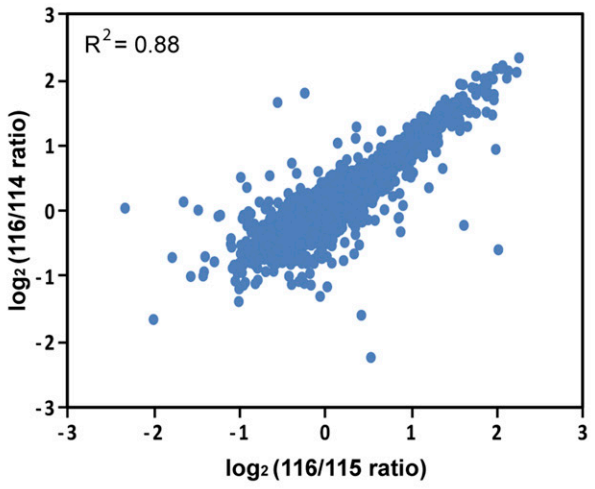


Fig. S3. Analysis of the background of biotinylation in fly larval muscles. (*A* and *B*) Immunostainings of control wild-type (yw) larval muscle cells. (*C–F*) Immunostainings of larval muscle cells overexpressing mito-APEX using *Dmef-Gal4*. After dissection, larval muscles were unlabeled (*A* and *C*) or treated with H_2O_2 only (*D*), biotin-phenol only (*E*), or both substrate and H_2O_2 (*B* and *F*). Wild-type cells have weak background labeling (*B*). Incubating cells with substrate alone (*D*) increased background, whereas supplementing H_2O_2 alone to cells had negligible effect on biotinylation (*E*). Providing both substrate and H_2O_2 to mito-APEX–expressing cells gives rise to specific and strong biotinylation (*D*). Nuclei are detected by DAPI staining (blue in *A–F*; gray in *A′–F′*). Mitochondrion is visualized by ATP5α staining (green in *A–F*; gray in *A″–F″*). Biotinylated proteins are visualized by SA-coupled fluorescent staining (red in *A–F*; gray in *A‴–F‴*).

A

# of detected peptides	18600
# of proteins detected with >1 iTRAQ ratio	2241
# of genes detected with >1 unique peptide	2222
# of genes enriched in any combination	632
# of genes enriched in all 4 combinations	389





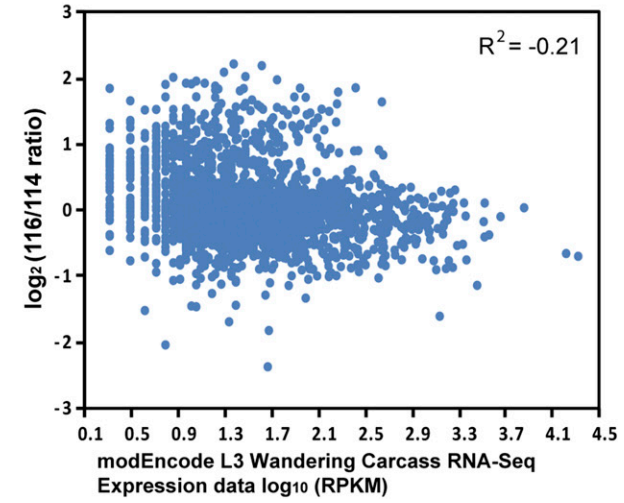


Fig. S4. Analysis of data from quantitative mass spectrometry. (*A*) Summary of information obtained from quantitative mass spectrometry. (*B*) Distribution of genes on the positive and negative control lists over iTRAQ ratios $\log_2$(mito-APEX replicate A/wild-type control). (*Top*) Genes on the positive control (PC) list. (*Bottom*) Genes on the negative control (NC) list. (*C*) High correlation between two different controls: iTRAQ ratios of proteins from mito-APEX replicate A (116) versus wild-type control (114) are plotted against iTRAQ ratios of proteins from mito-APEX replicate A (116) versus *Dmef-Gal* control (114). R^2 was calculated using Pearson correlation based on all detected proteins. (*D*) We retrieved transcript abundance data of L3 larval carcass from the RNA sequencing data generated by the modENCODE Consortium (32). The iTRAQ ratios were compared with $\log_{10}$ of RPKM (reads per kilobase per million mapped reads) values, and R^2 was calculated using Pearson correlation.

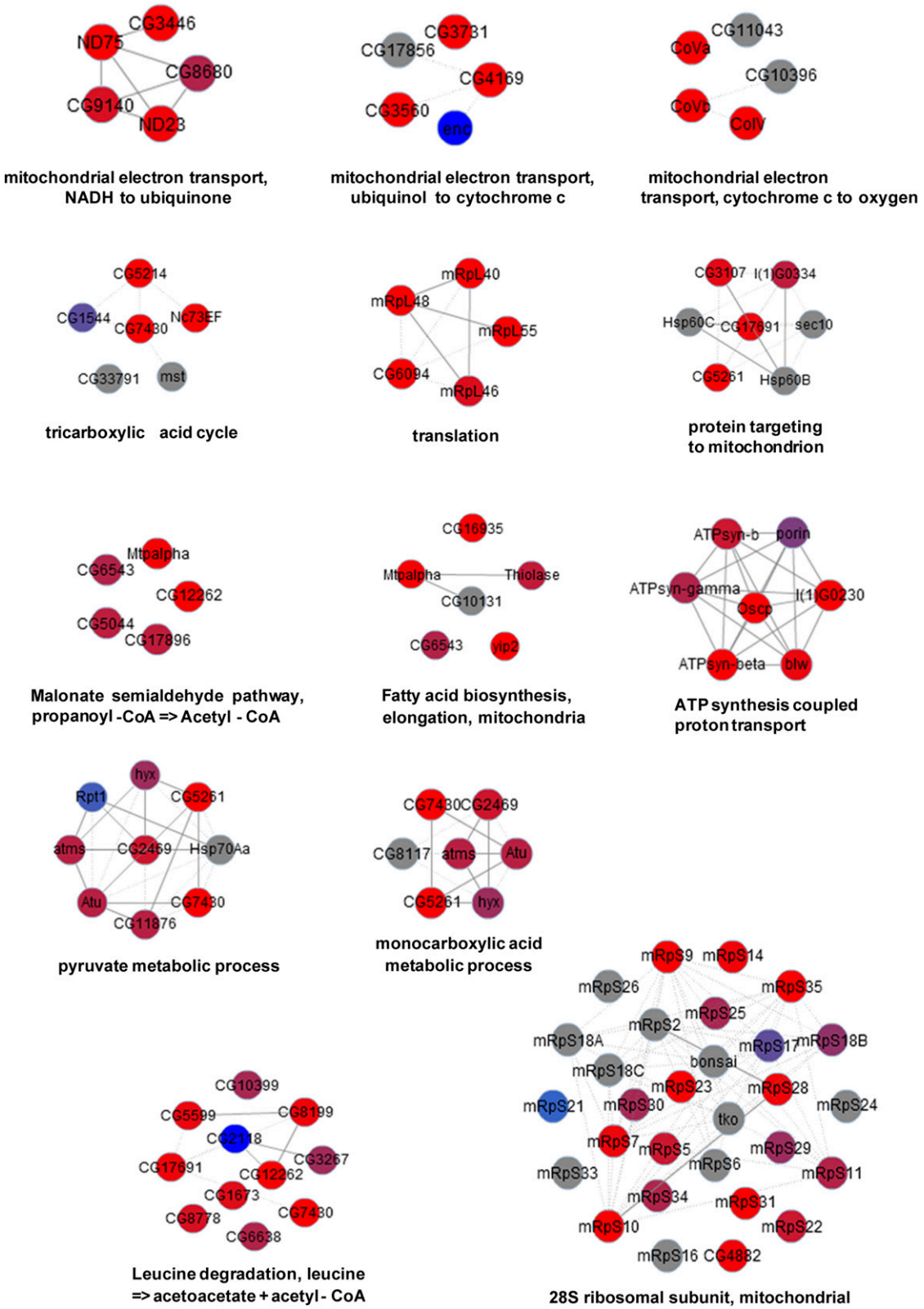


Fig. S5. COMPLEAT analysis of identified mitochondrial matrix proteins. Representative protein complexes enriched among the proteins identified by APEX. Each node represents one component of the complex, and node color reflects the average $\log_2$ ratio of each protein. Red, positive $\log_2$ ratio; blue, negative $\log_2$ ratio; gray, proteins not detected. Solid lines represent binary protein–protein interactions identified in *Drosophila*, whereas dotted lines represent binary protein–protein interactions identified in other species.

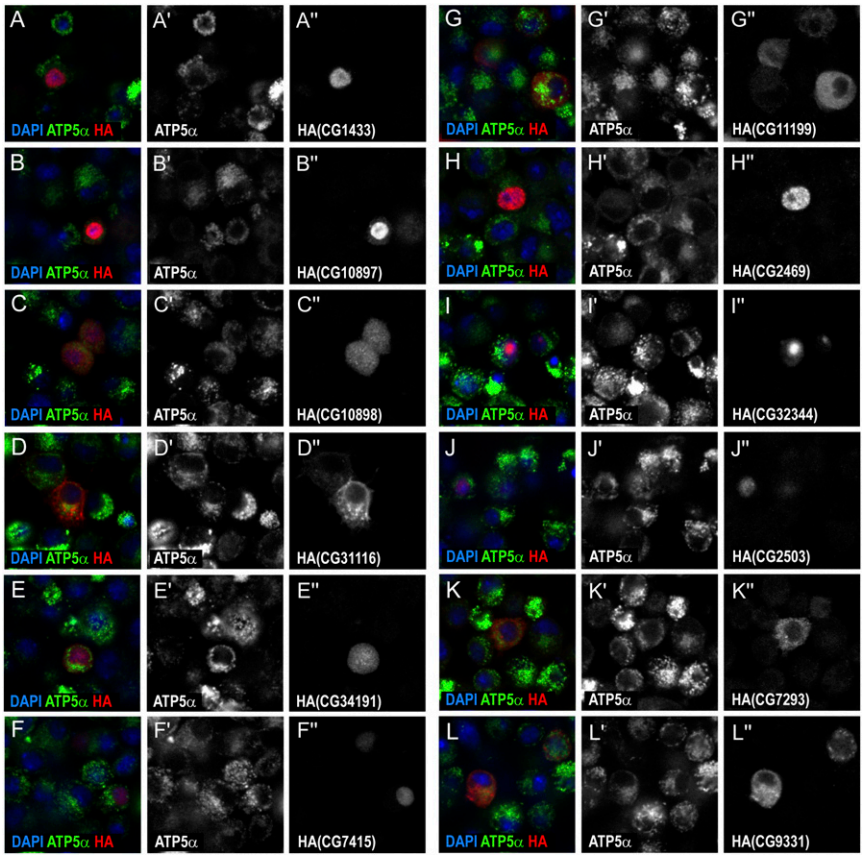


Fig. S6. Characterization of identified proteins in S2R+ cells. S2R+ cells were transfected with constructs carrying different proteins, as indicated, that are fused with HA at the C terminus. Localization of proteins of interest was detected by HA staining (red in *A*–*L*; gray in *A*′′–*L*′′). Mitochondrion is visualized by ATP5α staining (green in *A*–*L*; gray in *A*′–*L*′). Nuclei are detected by DAPI staining (blue in *A*–*L*).

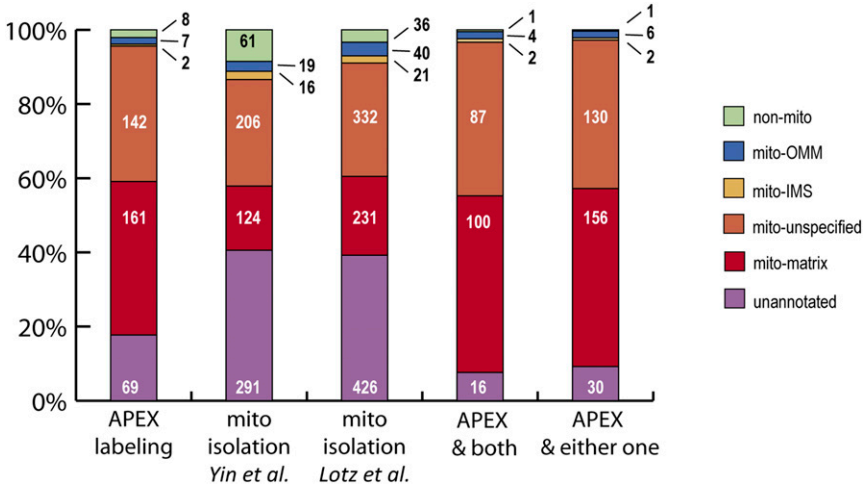


Fig. S7. Analysis of *Drosophila* mitochondrial proteins identified from different studies. Fractions of fly mitochondrial proteomes from different studies with subcompartmental annotation from human. Annotation of protein localization is as indicated. A higher percentage of genes identified by the APEX labeling method than by isolation methods [Yin et al. (16); Lotz et al. (17)] have mitochondrial matrix annotation (41.4% compared with 17.8% and 21.3% from isolation approaches). Genes identified by both APEX labeling and isolation-based approaches encode many proteins annotated as mitochondrial matrix proteins (48%). IMS, intermembrane space; OMM, outer membrane.

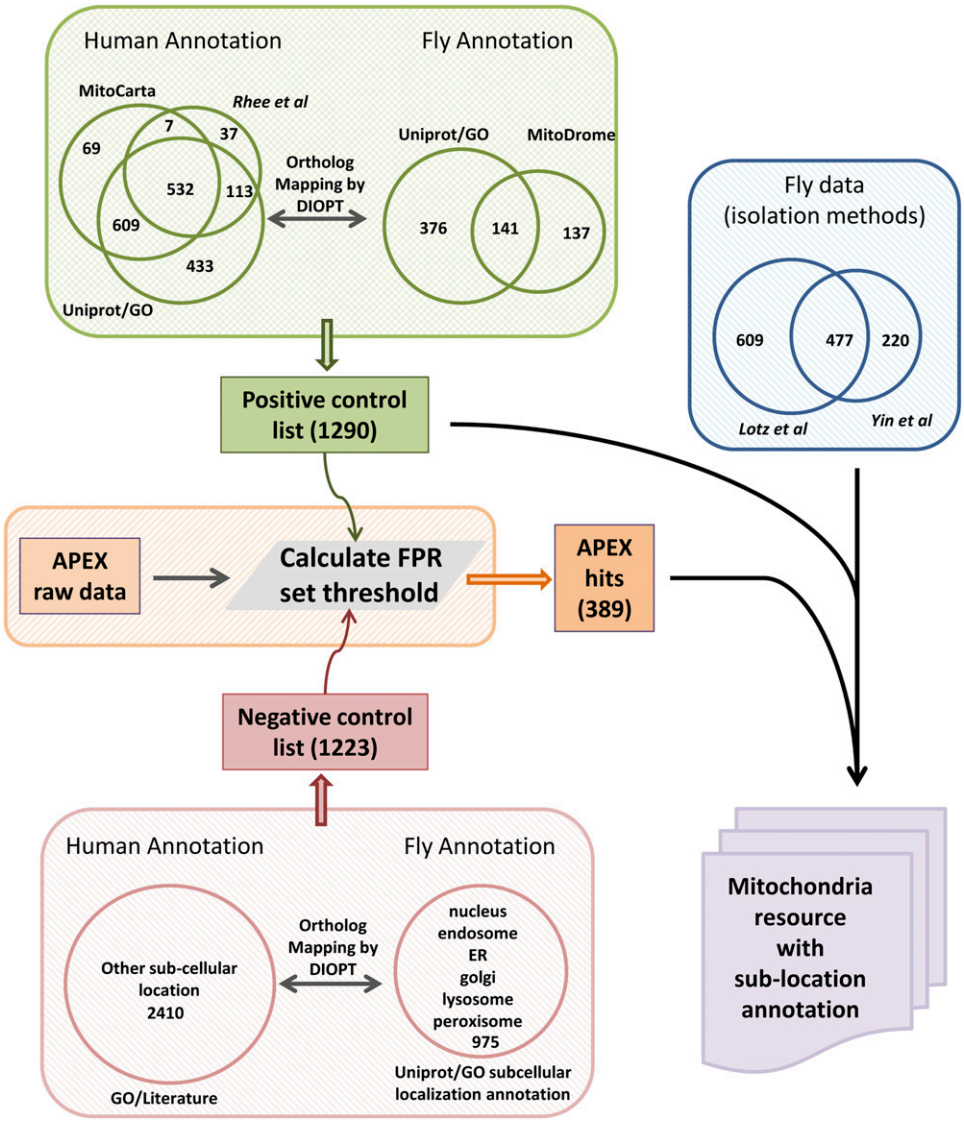


Fig. S8. Assembly of the MitoMax database. Data integration strategy used to generate MitoMax. The positive control list (1,290 genes) and negative control list (1,223 genes) were assembled from both human and *Drosophila* data to determine the cutoff of iTRAQ ratio for the APEX-labeled fly mitochondrial proteome. Genes, identified by either APEX labeling (389 genes) or isolation-based studies, are combined and integrated with genes from annotation (see assembly of positive control list) as well as *Drosophila* genes annotated at MitoMiner and MitoDrome to build MitoMax, a comprehensive database for *Drosophila* mitochondrial genes with subcompartmental annotation. We annotated the genes encoding proteins localized or exposed to the mitochondrial matrix based on Gene Ontology and APEX data from *Drosophila* and mammalian cell lines; 980 *Drosophila* genes are annotated with high confidence, and supporting evidence for mitochondrial localization of 2,126 proteins is reported at MitoMax.

Dataset S1. Mitochondrial matrix proteome identified using the APEX labeling method

Dataset S1

Dataset S2. Depth of coverage analysis of the five groups of mitochondrial proteins

Dataset S2

Dataset S3. COMPLEAT analysis of enriched complexes among mitochondrial proteins that are identified using the APEX approach

Dataset S3