Supplementary Information for: Defining mitochondrial protein functions through deep multi-omic profiling

Jarred W. Rensvold, Evgenia Shishkova, Yuriy Sverchkov, Ian J. Miller, Arda Cetinkaya, Angela Pyle, Mateusz Manicki, Dain R. Brademan, Yasemin Alanay, Julian Raiman, Adam Jochem, Paul D. Hutchins, Sean R. Peters, Vanessa Linke, Katherine A. Overmyer, Austin Z. Salome, Alexander S. Hebert, Catherine E. Vincent, Nicholas W. Kwiecien, Matthew J. Rush, Michael S. Westphall, Mark Craven, Nurten A. Akarsu, Robert W. Taylor, Joshua J. Coon, and David J. Pagliarini

Contents:

Methods

Cell Culture

HAP1 wild-type (WT) and knockout (KO) cells (Horizon Discovery, see Supplementary Table 7 for product information) were cultured in IMDM (Thermo, 12440053) with 10% FBS (Sigma, F2442) and 1× penicillin-streptomycin (Thermo, 15140122) at 37°C and 5% CO_2. C2C12 (Sigma, 91031101-1VL), HeLa (ATCC, CCL-2), 293 cells[26], normal adult human primary dermal fibroblasts (HDFa) (ATCC, PCS-201-012), and primary patient fibroblasts with the *RAB5IF* c.75G>A mutation were cultured in DMEM (Thermo, 11965092) with 10% FBS (Sigma, F2442) and 1× penicillin-streptomycin (Thermo, 15140122) at 37 °C and 5% CO_2. The Genome Editing and Animal Models Core at the University of Wisconsin–Madison Biotechnology Center generated the HAP1 knock-in cells containing the *PYURF* c.289_290dup patient mutation, and The Genome Engineering and iPSC Center (GEiC) at Washington University in St. Louis generated the HAP1 knock-in cells containing the *SLC30A9* c.1047_1049delGCA patient mutation and HAP1 knock-in cells containing the *RAB5IF* c.75G>A patient mutation. HAP1, C2C12, and HeLa cells were authenticated by the commercial source. 293 cells were not authenticated. HAP1, C2C12, and HeLa cells were negative for mycoplasma contamination as tested by the commercial source. 293 cells were not tested.

Cell Growth (Multiomics). The HAP1 KO cell lines were grown in 14 separate batches. Each batch was comprised of the same 15 KO lines except the final batch which had eight. The KO lines included in each batch were randomly selected as they were generated and received from Horizon Discovery. A separate batch of seven different HAP1 WT lines was collected for data normalization. 3-4 replicates of each batch were collected consecutively over the same number of weeks. For each replicate, a vial of each cell line was thawed and passaged at a ratio of 1:8 to 1:10 the following day. Two days later, the cells were collected by dissociation with trypsin (Thermo, 25300062), counted using a Millipore Muse Cell Analyzer and Muse Count and Viability Assay Kit (Sigma, MCH600103), and 1.5×10^6 cells were seeded on a 10-cm plate (3 plates per cell line). After 1 day, spent media was removed and replaced with fresh media; the following day, samples were collected for proteomic, lipidomic, and metabolomic analysis (1 plate per analysis). Sigma FBS (F2442, lot 16K293) was used for all cells grown for multiomic analysis.

Cell Growth (Density Measurements). Cells were grown in batches and replicates exactly as described above but were seeded at 2×10^5 cells per well on 6-well plates. Cells were collected by dissociation with trypsin (Thermo, 25300062), and counted using a Millipore Muse Cell Analyzer and Muse Count and Viability Assay Kit (Sigma, MCH600103) after 2 days, with the media refreshed (removed and replaced) after 1 day of growth. A single HAP1 WT cell line (C631, Horizon Discovery manufacture date of 160318) was grown with each batch and used to calculate fold-changes for the KO cell lines in that batch. p-values were calculated using a two-sided Student's unequal variances t-test (Welch's t-test). Mean $\log_2$-fold-change was computed as the quotient of the average cell density across technical replicates of the KO cell line and the average

cell density across technical replicates of the WT cell line, which were then $\log_2$ transformed; to compute p-values we used a two-sided Student's unequal variances t-test (Welch's t-test) comparing the set of cell densities in the technical replicates of the KO cell line and the set of cell densities in the technical replicates of the WT cell line.

siRNA Knockdown. HeLa and 293 cells were transfected with 20 nM siRNA (Dharmacon, see Supplementary Table 5 for product information) using 0.2% DharmaFECT 1 (Dharmacon, T-2001-02). C2C12 cells were transfected with 25 nM siRNA using 0.25% DharmaFECT 1. Cells were passaged or collected for analysis after 2 days. For siRNA knockdown lasting a total of 5 days, the cells were passaged after the initial 2-day knockdown, transfected again with siRNA the following day, and collected for analysis after an additional 2 days.

Cycloheximide Chase. 293 cells were seeded a 1.2×10^5 cells per well on 6-well plates and treated with siRNA the following day as described above. After 32 hours, cells were treated with 5 μM cycloheximide (Sigma, 239765-1ML), and samples were collected every 8 hours for 40 hours.

Patient Fibroblast Rescue. Primary patient fibroblasts with the *RAB5IF* c.75G>A mutation were seeded at 1×10^5 cells per well of a 6-well plate in Fibroblast Basal Medium (ATCC, PCS-201-030) containing the Fibroblast Growth Kit-Low serum (ATCC, PCS-201-041) and 0.1× penicillin-streptomycin (Thermo, 15140122), and incubated at 37°C and 5% CO_2. The following day, the medium was replaced with the same but without penicillin-streptomycin, and the cells were then transfected with 2 μg of either pEGFP-N1-FLAG (Addgene, 60360) or pEGFP-N1 (Clontech, 6085-1) containing the 129 aa RAB5IF isoform using 4 μL X-tremeGENE HP (Sigma, 6366236001) per well. After 2 days of incubation, samples were collected for immunoblotting.

Quantitative PCR

Total RNA was purified from cultured cells using the RNeasy Mini Kit (QIAGEN, 74104). First-strand cDNA was synthesized from purified RNA (2 μg) using the SuperScript III First-Strand Synthesis System (Thermo, 18080051) with a 50:50 mixture of oligo(dT)_{20} and random hexamers. Real-time qPCR was performed using Power SYBR Green PCR Master Mix (Thermo, 4367659), primers at 300 nM (see Supplementary Table 5 for primer sequences), *RPLP0* (*PYURF* experiments) or *ACTB* (*RAB5IF* experiments) as the endogenous control, and the comparative C_T ($\Delta\Delta$ C_T) method on an Applied Biosystems QuantStudio 6 Flex Real-Time PCR System and QuantStudio Real-Time PCR Software (version 1.2).

Immunoblotting

Protein lysates from cultured cells were prepared by scraping in lysis buffer composed of 150 mM NaCl, 1% IGEPAL CA-630, 0.5% sodium deoxycholate, 0.1% SDS, 50 mM Tris (pH 8.0), 0.4 mM EDTA (pH 8.0), 10% glycerol, and protease inhibitors (Sigma, 11836170001) while on ice, and transferred to microcentrifuge tubes. The extracts were clarified by centrifugation (16,000 ×g,

15 minutes, 4°C), and the supernatants were transferred to new tubes. Then 15-30 µg of cleared whole-cell lysate, as determined by BCA assay (Thermo, 23225), was separated on a NuPAGE 4-12% Bis-Tris Gel (Thermo, NP0323BOX) with a protein standard (Thermo, LC5800), transferred to PVDF membrane (Sigma, IPFL00010), probed with primary antibodies (see Supplementary Table 5 for product information), and analyzed using a LI-COR Odyssey CLx Imaging System and LI-COR Image Studio Software (version 5.2.5) with secondary antibodies (LI-COR, 925-32210, 926-32211, 926- 68070, 926-68071).

Immunoprecipitation

293 cells were transfected with 10 µg of pRL-CMV-PreYF-PIG-YF (normal construct, variant a, and variant b)[32] using 30 µL TransIT-293 (Mirus, MIR 2704) per 10-cm plate (2 plates per construct). After two days, cells were removed from plates via scraping in 500 µL of cold DPBS (4°C) and transferred to microcentrifuge tubes, and collected by centrifugation (500 ×g, 3 min, 4°C). Two plates of cells were used for each condition and combined after scraping. Cells were washed twice with 1 mL cold DPBS, and the cell pellet was resuspended in 400 µL lysis buffer composed of 10 mM Tris (pH 7.5), 150 mM NaCl, 0.5 mM EDTA, 0.5% IGEPAL CA-630, and protease inhibitors (Sigma, 11836170001). The cells were incubated on ice for 30 min with pipetting every 10 min. The cell lysate was clarified by centrifugation (20,000 ×g, 10 min, 4°C), diluted in 600 µL of wash buffer composed of 10 mM Tris (pH 7.5), 150 mM NaCl, and 0.5 mM EDTA. 900 uL of the diluted lysate was added to 50 µL washed and equilibrated anti-FLAG M2 magnetic beads (Sigma, M8823-1ML), and mixed end-over-end for 4 hours at 4°C. A sample of the diluted lysate (input fraction) was saved for analysis. Following the incubation, the beads were washed three times in 500 µL wash buffer and then eluted in 100 µL 2× NuPAGE LDS Sample Buffer (Thermo, NP0007) for 10 min at 95°C.

Mitochondrial Preparation and Blue-Native PAGE

Mitochondrial preparation and blue-native PAGE were performed as previously described[49-51] with minor changes. HAP1 cells grown on 10-cm plates (10 plates per condition) were washed twice with 5 mL warm DPBS (37°C) (Thermo, 14190144) at room temperature (RT). Cells were removed from plates via dissociation in 1 mL of warm 0.05% trypsin-EDTA (37°C) (Thermo, 25300062), washed from plates with warm growth medium (37 °C), transferred and combined in a 15-mL conical tube, and pelleted by centrifugation (100 ×g, 5 min). Cells were then washed twice with cold DPBS (4°C), pelleted (100 ×g, 5 min, 4°C), and the supernatant was discarded. The cell pellet was resuspended in 2× packed cell volume (~250 µL) of cold hypotonic buffer (4°C) (10 mM KCl, 1.5 mM MgCl2, 10 mM Tris-HCl pH 7.5). The cells were pelleted (100 ×g, 5 min, 4°C) to check for swelling, the supernatant was discarded, the cells were resuspended in 1.5 mL hypotonic buffer, and incubated on ice for 10 min. Cells were homogenized with a 1-mL glass Dounce grinder and tight-fitting pestle (Wheaton, 357538) using 40 strokes, then 1 mL of cold 2.5× MS homogenization buffer (4°C) (525 mM mannitol, 175 mM sucrose, 12.5 mM Tris-HCl pH 7.5, 2.5 mM EDTA pH 7.5) was added and mixed by inversion. The homogenate was split

between 2 microfuge tubes, and for each, the volume was bought to 1.5 mL with cold 1× MS homogenization buffer (4°C) (210 mM mannitol, 70 mM sucrose, 5 mM Tris-HCl pH 7.5, 1 mM EDTA pH 7.5). The homogenate was centrifuged twice (700 ×g, 10 min, 4°C), and then the supernatant was transferred to a new tube and centrifuged (12,000 ×g, 10 min, 4°C). The supernatant was removed, and mitochondrial pellets were stored at -80°C. Mitochondrial pellets were solubilized in 1× NativePAGE Sample Buffer (Thermo, BN2003) containing 1% Triton X-100 (Sigma, T9284-100ML) for 40 min on ice, centrifuged to remove insoluble material (20,000 ×g, 20 min, 4°C), and combined with BN-PAGE loading dye (0.5% Coomassie G-250, 50 mM 6-aminocaproic acid, 10 mM bis-Tris pH 7.0). Then 8 µg protein, as determined by BCA assay (Thermo, 23225), was separated on a NativePAGE 3- 12% Bis-Tris Gel (Thermo, BN1001BOX) with a protein standard (Thermo, LC0725), transferred to PVDF membrane (Sigma, IPVH00005), fixed by 8% acetic acid treatment, air-dried, and destained with methanol according to the manufacturer's instruction (Thermo). The membranes were probed with primary antibodies (see Supplementary Table 5 for product information) and analyzed using a Bio-Rad ChemiDoc MP and BioRad Image Lab Touch Software (version 2.4.0.03) with secondary antibodies (Cell Signaling, 7076S, 7074S) and ECL substrate (Thermo, 37071).

Oxygen Consumption Measurements

Cellular OCR measurements were performed using an Agilent Seahorse XFe96 Analyzer. HAP1 cells were seeded at 4×10^4 cells per well in XF96 microplates (Agilent, 102416-100) with IMDM (Thermo, 12440053) containing 10% FBS (Sigma, F2442) and 1× penicillin-streptomycin (Thermo, 15140122), and incubated at 37°C and 5% CO_2. The following day, the growth medium was exchanged for XF assay medium supplemented with 25 mM glucose, 1 mM pyruvate, and 4 mM glutamine (Agilent, 103680-100). OCR measurements were 3 min periods following 3 min mix periods. Cells were treated by sequential addition of 2 µM oligomycin, 250 nM FCCP, and 0.5 µM rotenone/antimycin A (Agilent, 103015-100). Mitochondrial stress test parameters were calculated using Agilent Seahorse Wave Desktop software (version 2.6.1) with OCR measurements normalized to relative cell number determined via crystal violet staining of cell nuclei as described previously[52]. Wells with injection failures, as determined by visual inspection of the injection ports and lack of drug response, were omitted from the analysis.

Calcium Assay

HAP1 or HeLa cells were seeded at 4×10^4 or 2×10^4 cells per well, respectively, in black, clear bottom poly-D-lysine coated 96-well plates (Greiner, 655946). The following day, the cells were washed once with HHBSS (1× HBSS + 20 mM HEPES) (Thermo, 14025092, 15630080), and 5 µM Fura-2 (Thermo, F1221) containing 1× (<0.2%) PowerLoad (Thermo, P10020) in HHBSS was added, and the cells were incubated at room temperature, protected from light for 40 min. The cells were then washed twice with HHBSS and incubated in HHBSS containing 2.5 mM probenecid (Thermo, P36400) without Fura-2 for 10 min at room temperature, protected from light. The cells were washed twice with calcium-free HHBSS containing 1 mM EGTA (RPI,

E14100-50.0) immediately before fluorescence measurements. Fluorescence intensity was measured every 5 seconds using excitation at 340/11 nm and 380/20 nm and emission at 508/20 nm (BioTek, 7082254, 7082228, 7082218) on a BioTek Synergy H4 plate reader and Gen5 Software (version 1.11.5) with the following settings: xenon flash light source, normal read speed, top 50% optics position, 80 sensitivity, filter switching per well, and 7.00 mM top probe vertical offset. After 2 min, 5 µM thapsigargin (Thermo, T7459) was added, and fluorescence was measured for an additional 6-8 min. Area under the curve (AUC) was calculated in GraphPad Prism 8 (version 8.4.3) with the measurement immediately before thapsigargin addition used as the baseline.

DepMap Bioinformatics Analyses

Pearson correlation coefficients and q-values from the Broad Institute Cancer Dependency Map CRISPR (DepMap 21Q2 Public+Score, Chronos)[25,53], RNAi (Achilles+DRIVE+Marcotte, DEMETER2)[54], and Proteomics[55] datasets were downloaded from depmap.org. Gene ontology (GO) enrichment analysis was performed on the Homo sapiens gene list using the analysis tool at geneontology.org (PANTHER Overrepresentation Test, released 20210224; GO Ontology database DOI: 10.5281/zenodo.5080993, released 2021-07-02). GO terms used for bar charts are rRNA binding (GO:0019843), structural constituent of ribosome (GO:0003735), mitochondrial small ribosomal subunit (GO:0005763), mitochondrial large ribosomal subunit (GO:0005762), mitochondrial degradosome (GO:0045025), mitochondrial translation (GO:0032543), mitochondrion morphogenesis (GO:0070584), positive regulation of mitochondrial RNA catabolic process (GO:0000962), proton-transporting ATP synthase activity, rotational mechanism (GO:0046933), NADH dehydrogenase (ubiquinone) activity (GO:0008137), mitochondrial proton-transporting ATP synthase complex, coupling factor F(o) (GO:0000276), mitochondrial respiratory chain complex I (GO:0005747), mitochondrial respiratory chain complex I assembly (GO:0032981), regulation of mitochondrial mRNA stability (GO:0044528), positive regulation of mitochondrial translation (GO:0070131).

LC-MS/MS Lipidomics

Sample Collection. HAP1 cells grown on 10-cm plates were washed twice with 5 mL of cold DPBS (4°C) (Thermo, 14190144) while on ice. Cells were removed from plates via scraping in 1 mL of cold DPBS containing protease inhibitors (Sigma, 11873580001), transferred to microcentrifuge tubes, and pelleted by centrifugation (3,000 ×g, 5 min, 4°C). The supernatants were removed, and the cell pellets frozen in liquid nitrogen and stored at -80°C.

Lipid Extraction. Lipid extractions were performed as previously described[56] with minor changes. Frozen cell pellets were thawed on ice and resuspended in 100 µL of cold DPBS (4°C), 10 µL of an internal standard composed of SPLASH LIPIDOMIX Mass Spec Standard + 10 µM CoQ6 + 20 µM 14:0 CL (Avanti, 330707-1EA, 900150O-1mg, 750332P-25mg) was added, and vortexed for 30 sec to mix. 900 µL of $CHCl_3$/MeOH (2:1, v/v, 4°C) (Fisher, C607SK-1; A456-500) was

added and the samples were vortexed twice for 1 min. 200 µL of HCl (1 M, 200 µL, 4°C) (Fisher, A144S-500) was added and the samples were vortexed twice for 30 sec. The samples were centrifuged (5,000 ×g, 2 min, 4 °C) to complete phase separation. All of the organic (bottom) phase was transferred to a new tube and dried under argon gas at 25°C. The organic residue was reconstituted in 100 µL of ACN/IPA/H_2O (65:30:5, v/v/v) (Fisher, A955-500, A461-500, W6500), vortexed twice for 30 sec, transferred to a new brown glass autosampler vial (Sigma, 29398-U), and stored at -80°C. Extracts from the 7 WT lines were combined, mixed, and aliquoted before storage to create a pooled WT standard to be used for normalization. The samples were thawed and vortexed before injection onto the LC-MS/MS system.

LC-MS/MS Lipidomics Data Acquisition. Lipid extracts were chromatographically separated using an Acquity CSH C18 column (50°C, 2.1 x 100 mm x 1.7 µm particle size; Waters Corporation) coupled to a Vanquish Binary Pump at a flow rate of 0.400 ml/min (Thermo Scientific). Mobile phase A consisted of 10 mM ammonium acetate in 70:30 (v/v) ACN/H2O with 250 µL/L acetic acid and mobile phase B consisted of 10 mM ammonium acetate in 90:10 (v/v) IPA/CAN with 250 µL/L acetic acid. Ten microliters of the lipid extracts were injected onto the column using the Vanquish autosampler (Thermo Scientific), ionized using a heated ESI (HESI) source (Thermo Scientific), and analyzed on the Q Exactive HF platform (Thermo Scientific). LC-MS/MS data were acquired using alternating positive and negative scan cycles consisting of a single MS scan at resolution of 60,000 at 200 m/z and AGC target of 3e6 followed by up to five data-dependent MS/MS scans at resolution of 15,000 at 200 m/z and AGC target of 5e5. Precursors were isolated using the quadruple with an isolation window of 1.4 Da and fragmented using normalized HCD of 20 and 25%.

LC-MS/MS Lipidomics Data Analysis. All lipid identification from LC-MS/MS data was performed using LipiDex (v1.0.2)[10]. Briefly, MS/MS scans were extracted from each LC-MS/MS experiment and converted to the Mascot Generic Format (MGF) using ProteoWizard (ProteoWizard Software Foundation, v3.0)[57]. The resulting MS/MS spectra were searched against the LipiDex HCD Acetate library in LipiDex to generate putative lipid identifications based on spectral similarity to *in silico* lipid fragmentation spectra. Retention time alignment and chromatographic feature finding were performed using Compound Discoverer (Thermo Scientific, v2.1). Putative MS/MS identifications were then assigned to chromatographic features and filtered in LipiDex to generate final lipid identifications.

Growth Normalization for Lipidomics Data. The quantitative values were corrected for differences in growth of each cell line based on cell density measurements. Cell density measurements were collected for each KO cell line in triplicate along with corresponding WT growth controls (see *Cell Growth (Multiomics)* above). A growth correction factor was calculated based on the mean cell density of a line divided by the mean cell density of all measured WT. This growth correction

factor was then multiplied across peak areas for metabolite values to correct for growth difference between cell lines.

GC-MS and LC-MS/MS Metabolomics

Sample Collection. HAP1 cells grown on 10-cm plates were quickly washed three times with 5 mL of warm DPBS (37°C) (Thermo, 14190144) at room temperature. Then liquid nitrogen (~2-5 mL) was added to quench metabolism[58], and the plates were transferred temporarily to dry ice before being stored in a sealed container at -80°C. Each plate was processed one-by-one from washing to quenching.

Metabolite Extraction and Derivatization. Cells were removed from plates via scraping in 1 mL cold MeOH/H2O (80:20, v/v, -80°C) (Fisher, A456-500, W6500) while on dry-ice, and transferred to microcentrifuge tubes. The extracts were clarified by centrifugation (16,000 ×g, 5 minutes, 4°C), and the supernatants were transferred to new tubes and stored at -80°C[59]. For GC-MS analysis, the samples were thawed, and 300 µL of cell extract were aliquoted into amber autosampler vials with 10 µL of a deuterated alanine/adipic acid solution (30 ppm) added as internal standard; then samples were dried under vacuum using a SpeedVac (Thermo Scientific). Samples were derivatized using 50 µL of pyridine: N-Methyl-N-(trimethylsilyl)trifluoroacetamide with 1% trimethylchlorosilane (1:1 v/v, Sigma Aldrich) and incubated at 60°C for 30 min.

For LC-MS/MS, underivatized extracts prepared as described above were used. All samples were stored at -40℃ until the day of analysis. Due to the large number of samples present in the study, samples were analyzed in 7 batches. To mitigate analytical bias in down-stream analysis, samples were distributed into batches in a manner that ensured each replicate of a cell line knockout was in a different batch, then samples were randomized within batch to further reduce bias. For each batch, samples were removed from the -40°C freezer, vortexed, placed on ice, and 200 µL of sample were aliquoted into 9 mm wide Amber glass vial with 350 µL fused insert. The samples were then dried down in the SpeedVac Concentrator (Thermo Scientific, Savant, SPD131DDA) for 70 minutes. The samples were stored dried down until the day of analysis. On the day of analysis, the controls and samples were resuspended in 35 µL of 80:20 water/methanol and vortexed on the Multi-Microplate Genie (Scientific Industries) for 30 minutes at a speed of 10 and then place in the autosampler for analysis.

GC-MS Metabolomics Data Acquisition. Metabolites samples were injected onto a TraceGOLD TG-5SILMS column (Thermo Scientific) and analyzed using Q Exactive GC-MS Orbitrap system (Thermo Scientific). Ionization mode was set to electron ionization (EI). The temperature gradient began at 100°C (hold time of one minute), increasing at a rate of 8.5°C per minute until reaching 260°C. The rate was then increased to 50°C per minute until reaching a final temperature of 320°C (hold time of four minutes). Sample gas flow was split at a 10:1 ratio, with a carrier gas flow of 1.200 mL/min. MS transfer line and ion source temperatures were set to 300°C and 250°C,

respectively. Data were acquired in full MS-SIM mode at 30,000 resolution. Scans occurred across a m/z range of 50 to 650 with an AGC target of 1e6.

LC-MS/MS Metabolomics Data Acquisition. Sample analysis was performed on an Acquity UPLC HSS T3 column held at 40 °C (150 mm x 2.1 mm x 1.8 μm particle size; Waters with a 50 mm pre-guard column of the same packing material). 1 μL of sample was injected by a Vanquish Split Sampler HT autosampler (Thermo Scientific). Separations were performed using a Vanquish Binary Pump (300 μL/min flow rate; Thermo Scientific) with the following gradient: initial conditions of 100% Mobile phase A (0.1% formic acid in water) for 0.5 min, then linear increase to 100% Mobile phase B (0.1% formic acid in methanol) over the next 8 min. 100% Mobile phase B was maintained for 0.5 min before returning to 0% mobile phase B over the next 0.75 min and equilibrating at 0% for a remaining 0.25 min.

The LC system was coupled to a Q Exactive Orbitrap mass spectrometer through a heated electrospray ionization (HESI II) source (Thermo Scientific). Source conditions were as follow: HESI II and capillary temperature at 320°C, sheath gas flow rate at 30 units, aux gas flow rate at 6 units, sweep gas flow rate at 0 units, spray voltage at |4.0 kV| for positive mode and |3.0 kV| for negative mode, and S-lens RF at 60.0 units. The MS was operated in a polarity switching mode acquiring positive and negative full MS and MS2 spectra (Top2) within the same injection. Acquisition parameters for full MS scans in both modes were 35,000 resolution, 1 × 10e6 automatic gain control (AGC) target, 100 ms ion accumulation time (max IT), and 70 to 750 m/z scan range. MS2 scans in both modes were then performed at 17,500 resolution, 1 × 10e5 AGC target, 50 ms max IT, 1.0 m/z isolation window, stepped normalized collision energy (NCE) at 20, 30, 40, and a 10.0 s dynamic exclusion.

GC-MS Metabolomics Data Analysis. An in-house software suite (https://github.com/coongroup/Y3K-GC-Quantitation-Software)[12] was used for deconvolution, peak alignment, quantitation, and identification of raw file peaks. Cutoffs for peak quantitation were set to a minimum fragment count of 10, minimum observation of a given peak across all files set to 33%, and analyte/background signal set to 10. Spectra were then matched against the unit resolution library curated by the National Institute of Standards and Technology (NIST), and the high resolution library developed in house. Metabolites lacking a confident identification were classified as ‘‘Unknown metabolites’’ and appended a unique identifier based on retention time. Peak heights of specified quant m/z were used to represent feature (metabolite) abundance.

LC-MS/MS Metabolomics Data Analysis. For data analysis, selected m/z and retention times were used to quantify metabolites, these peak areas were quantified using Thermo’s Tracefinder 4.0 application.

Combining LC-MS/MS and GC-MS Metabolomics Data. For metabolomics data, GC-MS and LC-MS data were combined into one data table with a column denoting the platform used for

quantitation. For metabolites that were detected with both platforms (n=25 metabolites), only quantitative data from one platform was retained for downstream analysis and this was selected based on the platform that yielded quantitative data with lower median within-strain relative standard deviations (RSD) for the specific metabolite.

Growth Normalization for Metabolomics Data. The quantitative values were corrected for differences in growth of each cell line based on cell density measurements. Cell density measurements were collected for each KO cell line in triplicate along with corresponding WT growth controls (see *Cell Growth (Multiomics)* above). A growth correction factor was calculated based on the mean cell density of a line divided by the mean cell density of all measured WT. This growth correction factor was then multiplied across peak areas for metabolite values to correct for growth difference between cell lines.

NanoLC-MS/MS Proteomics

Sample Collection. HAP1 cells grown on 10-cm plates were washed twice with 5 mL of cold DPBS (4°C) (Thermo, 14190144) while on ice. Cells were removed from plates via scraping in 1 mL of cold DPBS containing protease inhibitors (Sigma, 11873580001), transferred to microcentrifuge tubes, and pelleted by centrifugation (3,000 ×g, 5 min, 4°C). The supernatants were removed, and the cell pellets frozen in liquid nitrogen and stored at -80°C.

Proteomics Sample Preparation. Cell pellets were resuspended in lysis buffer containing 8 M urea, 100 mM Tris (pH = 8.0), 10 mM TCEP, and 40 mM chloroacetamide, and lyzed via vortexing for ~15 minutes and sonication in Qsonica Q700 sonicator (Newtown, CT) at amplitude of 35 and 4ºC for 20 s ON/ 10 s off with the total processing time of 10 min. Proteins were precipitated by addition of MeOH (90% final v/v) and centrifugation at 14,000 g for 10 minutes. Supernatant was discarded, and protein pellets were briefly air dried and resuspended in the 8 M urea lysis buffer described above. Lysates were diluted with 50 mM Tris to a final urea concentration of ~ 1 M before the addition of LysC in 1:50 ratio (enzyme:protein, Wako Chemicals, Richmond, VA) and overnight digestion at room temperature. In the morning trypsin in 1:50 ratio (enzyme:protein, Promega, Madison, WI) was added, and the samples were incubated for 3 hours. Then digests were acidified with 10% TFA and desalted over 10 mg StrataX solid phase extraction columns (Phenomenex, Torrance, CA). Desalted peptides were lyophilized to dryness in a SpeedVac (Thermo Fisher) and resuspended in 0.2% LC-MS grade formic acid (Pierce, Rockford, IL). Peptide concentration was measured using Quantitative Colorimetric Peptide Assay Kit (Pierce, Rockford, IL).

Shotgun Proteomics Data Acquisition. Self-pack PicoFrit 75–360 µm inner-outer diameter bare-fused silica capillary columns with 10 µm electrospray emitter tips (New Objective, Woburn, MA) were packed on an in-house built ultra-high-pressure column packing station[60] with 1.7 µm, 130 Å pore size, Bridged Ethylene Hybrid (BEH) C18 particles (Waters, Milford, MA) to a final length

of ~35 cm and installed on a Dionex Ultimate 3200 RS nano HPLC system (Thermo Fisher, Sunnyvale, CA). Mobile phase buffer A consisted of 0.2% formic acid in water; mobile phase B consisted of 0.2% formic acid in 70% HPLC-MS grade acetonitrile (Fisher Scientific, Hampton, NH). One microgram of peptides was loaded onto a column, kept at 50°C inside an in-house made heater, and separated at a flow rate of 300-350 nl/min over a 120-minute gradient, including column wash and re-equilibration time. Peptide ions were analyzed on Orbitrap Fusion Lumos® mass spectrometer (Thermo Scientific, San Jose, CA). Orbitrap survey (MS) scans were collected at a resolving power of 240,000 at 200 m/z with an AGC target of 1.5e6 ions and maximum injection time of 50 ms. The instrument functioned in Top Speed mode with 1 s cycle time using APD algorithm[61]. Precursors were isolated in the quadrupole with the isolation window of 0.7 Th. Tandem MS scans were performed in the ion trap using the turbo scan rate on precursors with 2-5 charge states using HCD fragmentation with normalized collision energy of 25 and dynamic exclusion of 20 s. The ion trap MS/MS AGC target was 3e4 with the maximum injection time of 18 ms and m/z range of 200-1,200.

Shotgun Proteomics Data Analysis. Due to computational power constrains, raw files split into two, approximately equal-sized groups. Each subset of files was processed using MaxQuant quantitative software package (version 1.6.0.13)[62] and searched against UniProt *Homo Sapiens* database including protein isoforms (20170207 download). If not specified, default MaxQuant settings were used. LFQ quantification was performed using LFQ minimum ratio count of 1 and no MS/MS requirement for LFQ comparisons. Carbamidomethylation of cysteine residues was included as fixed modifications; oxidation of methionine and acetylation of protein N-termini were set as variable modifications. Match between runs was enabled with default settings. ITMS MS/MS tolerance was 0.35 Da, and first search tolerance was 32 ppm.

Data Processing and Analysis

Lipidomics Data Processing. The list of quantified lipid features was filtered for batch-specific contaminants. Features were declared as batch-related contaminants and removed if their median RSD within a respective batch was less than 50% and the maximum difference between the average $\log_2$ feature area in each batch was greater than 2. Cardiolipin (n=73) and monolysocardiolipin (n=11) species were combined based on their chain length and saturation to improve completeness and alignment of these measurements across all samples. 15 initially unidentified lipid species were later identified as CoQ10-related adducts and subforms based on retention time and m/z matching and were removed from the final dataset.

Metabolomics Data Processing. Intraday variation was corrected using robust linear regression; rlm() function in the R package MASS[63]. The models were fit as follows, metabolite(i) intensity ~ file timestamp, for each metabolite feature i. The rlm() parameter "weight" was used to weigh the technical replicates of the wild type (WT) sample 2x more than the knockout samples, thus balancing potential sample-driven variations in the model fitting. Low quality features were

removed based on the degree of imputation and relative standard deviations (RSDs) within the WT samples; features that were imputed in ≥ 1/3 of sample or had RSDs ≥ 30 were removed. Samples analyzed between 1/3/2019 4:00pm and 1/4/2019 12:00 pm were also removed due to instrument failure.

Proteomics Data Processing. Protein groups from two MaxQuant searches were merged based on shared leading majority protein ID or by gene name. Lists of quantified proteins were filtered to remove reverse identifications, potential contaminants, and proteins which were identified only by a modification site. Measurements were further filtered where values were present in less than 50% of the combined dataset or that only ever appeared in one search and had no overlap in leading majority protein ID or gene name. Missing values from the knockout cell lines were imputed using a left-censored sampling from the tenth percentile for each measurement, based on the assumption that these values were most likely to be missing due to the low abundance of certain peptides caused by knockout related biological changes. For the wild type samples, missing values were imputed using an implementation of the k-nearest neighbors algorithm, based on the assumption that these missing values resulted from more stochastic technical issues. To control for measurements that may have been inconsistently quantified across batches (for instance due to poor retention time alignment), any protein groups with percentile rank ranges greater than 50% across all batches in the wild type samples were filtered out. The data was then batch corrected using ComBat in the R package sva (v3.30.1)[64].

Computing Mean log_2-Fold-Change Measurements, P-values, and Q-values. For lipid and metabolite measurements, mean log_2-fold-change for each cell line was computed as the difference between the average molecule log_2-intensity of the replicates of the cell line sample and the average molecule log_2-intensity over technical replicates of the wild type sample analyzed within the same batch as the cell line sample. For protein measurements, mean log_2-fold-change for each cell line was computed as the difference between the average molecule log_2-intensity over biological replicates of the cell line sample and the average molecule log_2-intensity over all technical replicates of the wild type sample across all batches. To compute p-values we used a two-sided Welch two-sample t-test comparing the set of log_2-intensities of the molecule in the replicates of the cell line samples and the set of average log_2-intensities of the molecule in the respective replicates of the wild type samples. Q-values were computed from the set of all p-values (multi-ome) using the method by Storey[65] with the qvalue R package.

KO cell line filtering

Our curated list of gene targets and associated cell lines consisted of 204 entries, corresponding to 202 single KO cell lines and two clones of a double KO cell line (Supplementary Table 1). The proteomics methodology detected 192 (94% of 204 entries) proteins that corresponded to the expression products of the KO genes. Among those, 184 proteins (90%) were quantified across a

sufficient number of cell lines and met other quality metrics (see Methods on general proteomics data processing).

We scrutinized levels of 184 KO proteins in the respective KO cell lines as measured prior to data normalization with ComBat. 101 (55% of the examined entries) contained missing values in one or more replicates that were later imputed (see Proteomics methods), meaning that originally the abundance of the KO protein was measured as zero. Among remaining 83 entries, 67 were measured as decreased by greater than two-fold, serving as confirmation of successful knocking out of the targeted gene.

Next, we estimated the practical limit of quantification (PLQ) for our proteomics methodology. As it is not possible to do this for every peptide analyte detected in our experiments, we used MaxLFQ values obtained on all proteins to generate these estimations. We approximated that to measure a protein as reduced by two-fold or greater in the KO cell line, abundance of that protein must be greater than MaxLFQ of 23.2 in wildtype cells.

Among 16 entries, remaining after the removal of proteins containing either missing values or exhibiting greater than two-fold reduction in abundance in respective KO cell lines, abundance of three proteins was below the PLQ of 23.2. Hence, we were left with 13 measurements (7% of the 184 entries) of target proteins whose abundance did not exhibit the decrease expected due to the deletion of the encoding genes for no apparent technical reason. This observation could be attributed to the ambiguity in assignment of protein groups to specific gene sequences based on the detected peptide sequences, documented imperfections of the CRISPR knockout procedure[66,67], and false identifications (up to 1%). To avoid introducing potentially faulty data into our analysis and in the absence of alternative knockout validation data, we removed measurements performed on these 13 cell lines from further consideration and did not include them into our resource.

These cell lines are the following:

FUNDC2-KO2	MTRES1-KO	SPATA20-KO1
HDHD3-KO1	PISD-KO1	SPATA20-KO2
HDHD3-KO2	PISD-KO2	TOMM7-KO2
MIPEP-KO	SFXN3-KO1	
MPC1-KO2	SFXN3-KO2	

t-SNE. T-distributed Stochastic Neighbor Embedding (t-SNE) is a technique for visualizing high-dimensional vectors by embedding them in a low-dimensional space in such a way that points (i.e. vectors) that are close in the high-dimensional space remain close in the low-dimensional space with high probability[68]. For each molecule in 191 conditions, we computed a mean $\log_2$-fold-change (KO/WT) and an associated q-value. To combine the measure of uncertainty represented by the q-value q with the magnitude of abundance change represented by the mean log-fold-change LFC value for each molecule in each cell line, we multiplied $(1 - q)$ by $(2/(1+2^{-LFC}) - 1)$, resulting in a q-adjusted relative difference value. These values represent the molecules as vectors in a 191-dimensional space. The resulting data matrix was scaled to a unit range, centered, and whitened to

20 dimensions (the number of dimensions d=20 was chosen by thresholding on a p-value of 0.05 for the variance explained by the first d principal components in permutation testing). We used the Rtsne R package to compute the t-SNE embedding with a perplexity of 110.52, the square root of the number of molecules, and theta of 0.1. Since the t-SNE optimization procedure relies on a random starting point, we executed 10 runs and used the embedding with the best objective function score for our figures, the website, and cluster analysis. When identifying radius-1 t-SNE neighbors of MXP proteins, we also note whether a molecule is consistently a radius-1 neighbor in all 10 runs.

Clustering. To identify clusters in the t-SNE embedding, we apply HDBSCAN[42] to the embedding using the dbscan R package[69] with a parameter of minPts=6. HDBSCAN is a spatial clustering algorithm which identifies clusters that are locally denser than the surrounding "background" distribution of points. Consequently, it is able identify clusters of arbitrary shape and does not require prior knowledge of the number of clusters. However, it only assigns points in dense regions to clusters, and designates points in sparser regions as "background" (see Extended Data Figure 8b). In our setting, proximity of a background point to a cluster is still indicative of similarity of differential abundance profiles, and therefore is of interest. For this reason, we also identify, for each background point, the cluster to which it is closest by single linkage (see Extended Data Figure 8c).

KO-Specific Phenotype Detection. To systematically identify significant molecule perturbations that are unique to only a few knockout strains, we applied adapted an approach previously used in Stefely et al.[12]. For each profiled molecule we considered the 191 $\log_2$ fold change measurements corresponding to all knockouts and separated them into positive and negative $\log_2$ fold-change groups. These two sets were plotted individually with $\log_2$ fold change and $-\log_{10}$(p-value) along the x and y axes respectively and normalized such that the largest magnitude $\log_2$ fold change and largest $-\log_{10}$(p-value) were set to 1. Considering the knockouts with the six largest $\log_2$ fold changes where $p < 0.05$, we calculated the Euclidean distances in this two-dimensional space to their closest neighbors among those that correspond to a knockout of a different gene and have a smaller magnitude fold_2 change (e.g. in Figure 4c, ETFB^{KO2} is not considered when searching for the closest neighbor of ETFB^{KO1} since both knockouts target ETFB). From the resulting nearest-neighbor distances we kept the largest three for each of the positive and negative $\log_2$ fold change groups for each molecule.

Protein Group and Pathway Annotation. Human MitoCarta2.0 was used to identify mitochondrially localized proteins[21]. OxPhos, mitoribosome (28S and 39S), complex I, CI assembly factors, complex Q (COQ3-COQ9), and MICOS/MIB proteins were manually curated from HGNC gene groups and literature annotations[28,68,70-74].

Targeted Analysis of CoQ_{10} and Pathway Intermediates

Lipid extracts prepared for discovery lipidomics were used to profile levels of CoQ10 and its intermediates in the targeted fashion using the same LC setup as described for discovery lipidomics. The LC system was coupled to a Q Exactive HF mass spectrometer by a HESI II heated ESI source kept at 300 °C (Thermo Scientific). The inlet capillary was kept at 300 °C, sheath gas was set to 25 units, auxiliary gas to 10 units, S-lens RF level to 50.0, and the spray voltage was set to 4,000 V for positive and 3,500 V for negative mode. The MS was operated in positive parallel reaction monitoring (PRM) mode and negative PRM mode, respectively, acquiring targeted scans to quantify key CoQ intermediates. MS acquisition parameters were 60,000 resolving power for MS2 scans, 5 x 105 automatic gain control (AGC) target, 200-ms MS2 ion accumulation time, 2.2-Th isolation width for fragmentation, and stepped HCD collision energy of 15, 25, and 35 units. The resulting data were processed to produce abundance measurements using TraceFinder 4.0 (Thermo Fisher Scientific). Mean log_2-fold-change for each molecule was computed as the difference between the average molecule log_2-intensity over technical replicates of the cell line sample and the average molecule log_2-intensity over technical replicates of the wild type sample; to compute p-values we used a two-sided Student's unequal variances t-test (Welch's t-test) comparing the set of log_2-intensities of the molecule in the technical replicates of the cell line sample and the set of log_2-intensities of the molecule in the technical replicates of the wild type sample.

LC-MS/MS PUYRF Interactome Analysis

Fresh crude mitochondria were isolated from HAP1 cells[62] (3x 15 cm plates, 75% confluency) and their protein content was measured using a BCA kit (Pierce). Next, ~5 µL of mitochondria (50 µg of total protein) was resuspended in Lysis Buffer [20 mM HEPES pH 7.4, 100 mM NaCl, 10% glycerol, 1% digitonin, 1 mM DTT, 1x protease inhibitors mix, 100 µM PMSF] to 50 µL total volume, pipetted thoroughly and incubated on ice for 20 min. with brief vortexing every 5 min. Next, the samples were spun 20 000g x 10 min. at 4°C and the clarified supernatant was transferred to a new tube. 100 µg of his6-PYURF or his6-MBP was added to the sample as a "bait" and incubated at 4°C for 2h with slow rotation. To capture the formed complexes, 50 µL of TALON resin (100 µL slurry in Wash Buffer) was added and incubated at 4°C for 2h with slow rotation. The resin with bound his6-Nd35_PYURF and its interactors was pelleted 10 min. x 100g at 4°C, supernatant containing unbound proteins was discarded, and the pellet was resuspended in 200 µL of Wash Buffer [20 mM HEPES pH 7.4, 100 mM NaCl, 10% glycerol, 0.1% digitonin, 0.1 mM DTT, 5 mM Imidazole]. Samples were transferred to 96-well MultiScreen Filter Plate (Millipore), spun 40g x 1 min. at 4°C, washed 2 times with 200 µL of the Wash Buffer, washed 8 times with 200 µL of the Wash Buffer-II [20 mM HEPES 7.4, 100 mM NaCl, 5mM Imidazole], and washed 10 times with 200 µL of Mass Spec Buffer [100 mM Tris pH 7.4, 100 mM NaCl]. Finally, the captured protein partners were eluted from the resin using Urea Buffer [8M Urea, 100 mM Tris pH 7.4, 100 mM NaCl]. The experiment was performed in triplicate.

Eluted proteins were incubated in 10 mM TCEP and 40 mM CAA for 15 minutes while vortexing. Samples were brought to pH 8 with 1 M Tris (pH = 8.5), and 1 µg of LysC was added. Samples were incubated with LysC for 4 hours at room temperature, diluted to ~1.5 M urea with 100 mM Tris (pH = 8), then incubated with 1 µg of trypsin overnight (15 hours) on a rocker at room temperature. The tryptic digests were quenched with 10% TFA, then de-salted using 10 mg Strata X columns. Following elution, samples were dried down and re-suspended in 25 µL of 0.2% formic acid for LC-MS/MS analysis, 4 µL of which were injected onto the column. Peptides were separated over 90 min gradient using a Dionex Ultimate 3200 RS nano HPLC system (Thermo Fisher, Sunnyvale, CA), as described for shotgun proteomics analysis, and analyzed on Orbitrap Eclipse® mass spectrometer (Thermo Scientific, San Jose, CA). Survey scans were collected in Orbitrap over 300-1350 scan range with resolution of 240,000, normalized AGC target of 300%, and maximum injection time (max IT) of 50 ms. Precursors were isolated using quadrupole with the isolation width of 0.5 Da and fragmented using HCD fragmentation with normalized collision energy of 25. Dependent scans were collected in the ion trap using turbo scan rate over 150-1350 scan range with normalized AGC target of 200% and max IT of 30 ms. Raw files were searched in MaxQuant against UniProt *Homo Sapiens* database including protein isoforms (20190619 download) using default settings, except the following: "match between runs" was enabled with the default settings, IT mass tolerance was set to 0.35 Da, LFQ minimum ratio count was set to 1, and no requirement of MS2 for LFQ. Reverse identifications, potential contaminants, and proteins which were identified only by a modification site were removed from final results. Mean $\log_2$-fold-change for each molecule was computed as the difference between the average molecule $\log_2$-intensity over technical replicates of the PYURF sample and the average molecule $\log_2$-intensity over technical replicates of the MBP sample; to compute p-values we used a two-sided Student's equal variances t-test comparing the set of $\log_2$-intensities of the molecule in the technical replicates of the PYURF sample and the set of $\log_2$-intensities of the molecule in the technical replicates of the MBP sample, which were then FDR-adjusted using the Benjamini–Hochberg procedure.

Recombinant PYURF Purification

E.coli RIPL (DE3) cells harboring his6-MBP-TEV-Nd35_PYURF WT or his6-MBP-TEV-Nd35_PYURF c.289_290dup coded on pVP68a plasmids were inoculated from O/N LB cultures into fresh 3 x 1L LB cultures to OD_{600}=0.1, grown to OD_{600}=0.6 at 37°C, cooled down to 20°C and induced with 200 µM IPTG. Proteins were overproduced for 17h, cells were pelleted 20 min. x 5000g, resuspended in ice-cold buffer A [25 mM HEPES pH 7.4, 400 mM NaCl, 0.5 mM β-mercaptoethanol, 10% glycerol] with 1 mM PMSF and 1 mg/g lysozyme added. The cells were incubated at 4°C for 1h and then lysed by 6 rounds of sonication (70% amplitude) for 10 sec. ON and 2 min. OFF. The lysate was centrifuged 1h x 50 000g at 4°C and then the supernatant was incubated with 5 mL of buffer A-equilibrated TALON resin O/N at 4°C. The resin was washed 3 times with 50 mL buffer A +10 mM imidazole and the protein was eluted with buffer A +200 mM imidazole buffer that was subsequently exchanged to buffer A using Amicon Ultra 30kDa cutoff

filter spin tubes. Next, his6-MBP was cleaved off O/N at 4°C with TEV protease (1:100 ratio). Cleaved MBP and added his6-TEV protease were removed by two incubations with 1 mL TALON resin. Concentration of the purified protein was measured using BCA kit (Pierce), purity was estimated using SDS-PAGE gel and the remaining supernatant was aliquoted, frozen in liquid nitrogen and stored at -80°C. The same method was used to purify his6-Nd35_PYURF (pET21a+) and his6-MBP (pVP68k); however, TEV cleavage was omitted.

Recombinant NDUFAF5 Purification

E.coli ORIGAMI cells (NOVAGEN) harboring pRSf-duet vector coding for codon optimized his6-TEV-Nd67_NDUFAF5_Cd35 and NDUFAF8 were inoculated from O/N LB cultures into fresh 6 x 1L LB cultures to OD_{600}=0.1, grown to OD_{600}=0.6 at 37°C, cooled down to 20°C and induced with 200 µM IPTG. Proteins were overproduced for 20h, cells were pelleted 20 min. x 5000g and resuspended in ice-cold buffer A [25 mM HEPES pH 7.4, 400 mM NaCl, 10% glycerol] with 5 uL DNaseI, 1 mM PMSF, 1 mM MgCl2 and 1mg/g lysozyme added. The cells were incubated at 4°C for 30 min., frozen as small droplets in liquid nitrogen and then lysed by CryoMill. The lysate in form of frozen powder was resuspended in 50 mL of Buffer A at 4°C and centrifuged 1h x 50 000g at 4°C. The clarified supernatant was incubated with 5 mL of buffer A-equilibrated TALON resin O/N at 4°C. The resin was washed 3 times with 50 mL buffer A +20 mM imidazole and the protein was eluted with buffer A +200 mM imidazole. The protein was concentrated to 1 mL using Amicon Ultra 10kDa cutoff filter spin tubes and subjected to Size Exclusion Chromatography [Bio-Rad FPLC system, Hi-Load pg200 column, 0.75 mL/min, UV detector] in buffer B [20 mM Tris pH 7.4, 200 mM NaCl, 10% glycerol]. Purity of fractions was estimated on SDS-PAGE gel and fractions of the highest purity were pooled together. Concentration of the purified protein was measured using a BCA kit (Pierce), and the sample was aliquoted and frozen in liquid nitrogen and stored at -80°C.

Differential Scanning Fluorimetry (DSF)

Proteins were thawed on ice and diluted in Buffer A [20 mM Tris 7.4, 200 mM NaCl, 10% glycerol, 1 mM DTT] to 10 µM NDUFAF5, 2 µM and 50 µM PYURF, and 2 µM and 50 µM PYURF_c.289_290dup. Fresh S-Adenosylmethionine 2 µM and 100 µM stocks were prepared in Buffer A. Peptide (APRYDMDRFGVVFRASPRQA) representing the reported NDUFAF5 substrate was purchased from GenScript, resuspended in H_2O to 1 mM and then diluted to 2 µM and 100 µM stocks in Buffer A. Proteins and ligands were pipetted into 96-well reaction plate (MicroAmp™, Thermo) at indicated concentrations and 18 µL total volume, spun 2 min. x 1000g, mixed, spun again, and incubated at room temperature for 10 min. Next, 2 µL of 50x stock of SYPRO Orange DSF dye (Thermo) prepared in Buffer A was added to each reaction to final 5x concentration. The plate was sealed with optical adhesive film (MicroAmp™, Thermo) and melting gradient (15°C to 95°C, 0.025°C/sec.) was performed on an Applied Biosystems QuantStudio 6 Flex Real-Time PCR System, and data were analyzed using Protein Thermal Shift Software (version 1.3). All reactions were performed in triplicate.

PYURF Clinical Analyses

A female infant, born to, first cousin Afghan parents. At 5 months, she presented with urosepsis and profound metabolic lactic acidosis (blood lactate 15-20 mmol/L; normal range 0.7-2.1 mmol/L). At 6 months, she showed global delayed development. Clinical features included a poor fix and follow, optic pallor, truncal hypotonia with increased peripheral tone. Initial urinary organic acids showed a marked elevation in lactate, elevated 2 hydroxybutyric, 2-hydroxyisovaleric, and 3-hydroxybutyric acids, and tyrosine metabolites. Acylcarnitines showed elevated C2- and C4-OH species with a free carnitine of 29 umol/L. A convalescent sample was normal. Plasma amino acids showed elevated alanine (> 2500 umol/L) and elevated proline (2225 umol/L).

Brain MRI at 9 months showed increased extra axial CSF spaces. Cystic, high signal cerebellar white-matter and gliosis. Decreased myelination in the internal capsule, otherwise general decreased white-matter volume/cerebellar atrophy. CSF lactate levels elevated (8 mmol/L). Echocardiogram showed septal hypertrophy. She developed electrolyte disturbance requiring supplementation.

There were further admissions due to her renal dysfunction and lactic acidosis followed by neurodevelopmental arrest. At 16 months, she had a further episode of severe lactic acidosis and electrolyte imbalance and significant persistent clinical deterioration. In light of the severity of her likely mitochondrial disease presentation and her predicted severe course, the family were counseled, management was redirected to end of life care and she passed away. Mother has since delivered a healthy sibling who remains well.

Written informed consent was obtained from the parents of the child involved in the study in accordance with the Declaration of Helsinki protocols and approved by local institutional review boards.

WES

Whole exome sequencing was performed as follows; coding regions were enriched with Nextera Rapid Exome Capture (Illumina) and sequenced with 100 bp paired-end reads on an Illumina NextSeq500 platform. An in-house bioinformatics pipeline was applied to the raw FASTQ files. Briefly, PCR duplicates were removed using FastUniq[75]; Burrows-Wheeler Aligner [76] was used to align the resultant reads to the human reference genome (UCSC hg38); genetic variants were detected using Freebayes[77] and functionally annotated using ANNOVAR[78]. In addition to the variant identified for *PYURF*, two further homozygous changes were filtered and further investigated. A homozygous missense *AIFM3* (GenBank: NM_144704) variant, c.1123C>T (p.Arg375Cys), was excluded due to homozygotes being present in gnomAD. A homozygous missense *MTHFD1L* (GenBank: NM_001242767) variant, c.283T>G (p.Phe95Val), gnomAD (MAF 0.0003211) was not predicted to be pathogenic according to any of the *in silico* prediction tools applied. WES was performed following referral to the NHS Highly Specialised Mitochondrial Services, Newcastle upon Tyne. Written informed consent was obtained from the

parents of the child involved in the study in accordance with the Declaration of Helsinki protocols and approved by local institutional review boards.

Genomic Analyses of *RAB5IF*-related CFSMR Family

Genome-wide SNP genotyping by Affymetrix Human Mapping 250K Nsp1 arrays had already been completed in Alanay et al. 2013[47]. The individual genotypes for the CFSMR family have been reevaluated for genome-wide identity-by-decent regions, with an emphasis on *RAB5IF* locus on chromosome 20. Sanger sequencing for four coding exons of *RAB5IF* (ENST00000344795.3, NM_018840.5) has been performed using BigDye Terminator v3.1 according to manufacturer's protocols. GRCh37/hg19 assembly was used as the human reference genome. CFSMR family samples used in this study have been collected and studied in the scope of Cranirare project, under Hacettepe University Non-Interventional Clinical Research Ethics Board approval (TBK 09/4-42; revised at 15th June 2021, Approval number: GO 21/748).

Human Research Participants Ethics Oversight

All procedures were in accordance with the ethical principles of the Declaration of Helsinki. Written patient consent was obtained, and all the studies were performed in agreement with the approved guidelines of local ethics committees of each institution participating in this study.

Data Availability

All associated mass spectrometry RAW files and search results were deposited into the MassIVE data repository (accession number MSV000086685) and can be accessed using the following link: https://massive.ucsd.edu/ProteoSAFe/dataset.jsp?task=9a79911bd36e4f02baf0ea1d108511e6

MITOMICS data can be accessed and explored via the webtool MITOMICS.app, which is an adaptation of the online data processing and visualization platform Argonaut[79].

Relevant ClinVar accession numbers for the disease-causing variants described in this manuscript are as follows:
PYURF, NM_032906.4; c.289_290dup (p.Gln97Hisfs*6); ClinVar: SCV001470705.
RAB5IF, NM_018840.5; c.75G>A (p.Trp25*); ClinVar: SCV002059946

Code Availability

All associated code for data processing, analysis, and the companion webtool can be found in the GitHub repository https://github.com/coongroup/MITOMICS.

Methods references

49 Clayton, D. A. & Shadel, G. S. Isolation of mitochondria from tissue culture cells. *Cold Spring Harb Protoc* **2014**, pdb prot080002, doi:10.1101/pdb.prot080002 (2014).

50 Folco, E. G., Lei, H., Hsu, J. L. & Reed, R. Small-scale nuclear extracts for functional assays of gene-expression machineries. *J Vis Exp*, doi:10.3791/4140 (2012).

51 McKenzie, M., Lazarou, M., Thorburn, D. R. & Ryan, M. T. Analysis of mitochondrial subunit assembly into respiratory chain complexes using Blue Native polyacrylamide gel electrophoresis. *Anal Biochem* **364**, 128-137, doi:10.1016/j.ab.2007.02.022 (2007).

52 Rensvold, J. W. *et al.* Complementary RNA and protein profiling identifies iron as a key regulator of mitochondrial biogenesis. *Cell Rep* **3**, 237-245, doi:10.1016/j.celrep.2012.11.029 (2013).

53 Meyers, R. M. *et al.* Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nat Genet* **49**, 1779-1784, doi:10.1038/ng.3984 (2017).

54 McFarland, J. M. *et al.* Improved estimation of cancer dependencies from large-scale RNAi screens using model-based normalization and data integration. *Nat Commun* **9**, 4610, doi:10.1038/s41467-018-06916-5 (2018).

55 Nusinow, D. P. *et al.* Quantitative Proteomics of the Cancer Cell Line Encyclopedia. *Cell* **180**, 387-402 e316, doi:10.1016/j.cell.2019.12.023 (2020).

56 Jha, P. *et al.* Systems Analyses Reveal Physiological Roles and Genetic Regulators of Liver Lipid Species. *Cell Syst* **6**, 722-733 e726, doi:10.1016/j.cels.2018.05.016 (2018).

57 Adusumilli, R. & Mallick, P. Data Conversion with ProteoWizard msConvert. *Methods Mol Biol* **1550**, 339-368, doi:10.1007/978-1-4939-6747-6_23 (2017).

58 Lorenz, M. A., Burant, C. F. & Kennedy, R. T. Reducing time and increasing sensitivity in sample preparation for adherent mammalian cell metabolomics. *Anal Chem* **83**, 3406-3414, doi:10.1021/ac103313x (2011).

59 Fan, J. *et al.* Quantitative flux analysis reveals folate-dependent NADPH production. *Nature* **510**, 298-302, doi:10.1038/nature13236 (2014).

60 Shishkova, E., Hebert, A. S., Westphall, M. S. & Coon, J. J. Ultra-High Pressure (>30,000 psi) Packing of Capillary Columns Enhancing Depth of Shotgun Proteomic Analyses. *Anal Chem* **90**, 11503-11508, doi:10.1021/acs.analchem.8b02766 (2018).

61 Hebert, A. S. *et al.* Improved Precursor Characterization for Data-Dependent Mass Spectrometry. *Anal Chem* **90**, 2333-2340, doi:10.1021/acs.analchem.7b04808 (2018).

62 Cox, J. *et al.* Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics* **13**, 2513-2526, doi:10.1074/mcp.M113.031591 (2014).

63 Venables, W. N., Ripley, B. D. & Venables, W. N. *Modern applied statistics with S*. 4th edn, (Springer, 2002).

64 Leek, J. T. *et al.* sva: Surrogate variable analysis. *R package version* **3**, 882-883 (2017).

65 Storey, J. D. The positive false discovery rate: A Bayesian interpretation and the q-value. *Ann Stat* **31**, 2013-2035, doi:DOI 10.1214/aos/1074290335 (2003).

66 Smits, A. H. *et al.* Biological plasticity rescues target activity in CRISPR knock outs. *Nat Methods* **16**, 1087-1093, doi:10.1038/s41592-019-0614-5 (2019).

67 Tuladhar, R. *et al.* CRISPR-Cas9-based mutagenesis frequently provokes on-target mRNA misregulation. *Nat Commun* **10**, 4056, doi:10.1038/s41467-019-12028-5 (2019).

68 van der Maaten, L. & Hinton, G. Visualizing Data using t-SNE. *J Mach Learn Res* **9**, 2579-2605 (2008).
69 Hahsler, M., Piekenbrock, M. & Doran, D. dbscan: Fast Density-Based Clustering with R. *J Stat Softw* **91**, 1-30, doi:10.18637/jss.v091.i01 (2019).
70 Amunts, A., Brown, A., Toots, J., Scheres, S. H. W. & Ramakrishnan, V. Ribosome. The structure of the human mitochondrial ribosome. *Science* **348**, 95-98, doi:10.1126/science.aaa1193 (2015).
71 Brown, A. *et al.* Structure of the large ribosomal subunit from human mitochondria. *Science* **346**, 718-722, doi:10.1126/science.1258026 (2014).
72 Formosa, L. E., Dibley, M. G., Stroud, D. A. & Ryan, M. T. Building a complex complex: Assembly of mitochondrial respiratory chain complex I. *Semin Cell Dev Biol* **76**, 154-162, doi:10.1016/j.semcdb.2017.08.011 (2018).
73 Greber, B. J. *et al.* Ribosome. The complete structure of the 55S mammalian mitochondrial ribosome. *Science* **348**, 303-308, doi:10.1126/science.aaa3872 (2015).
74 Huynen, M. A., Muhlmeister, M., Gotthardt, K., Guerrero-Castillo, S. & Brandt, U. Evolution and structural organization of the mitochondrial contact site (MICOS) complex and the mitochondrial intermembrane space bridging (MIB) complex. *Biochim Biophys Acta* **1863**, 91-101, doi:10.1016/j.bbamcr.2015.10.009 (2016).
75 Xu, H. *et al.* FastUniq: a fast de novo duplicates removal tool for paired short reads. *PLoS One* **7**, e52249, doi:10.1371/journal.pone.0052249 (2012).
76 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).
77 Garrison, E. & Marth, G. Haplotype-based variant detection from short-read sequencing. arXiv:1207.3907 (2012). <https://ui.adsabs.harvard.edu/abs/2012arXiv1207.3907G>.
78 Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**, e164, doi:10.1093/nar/gkq603 (2010).
79 Brademan, D. R. *et al.* Argonaut: A Web Platform for Collaborative Multi-omic Data Visualization and Exploration. *Patterns (N Y)* **1**, doi:10.1016/j.patter.2020.100122 (2020).

Supplementary Table Guide

Supplementary Table 1: Knockout targets, cell density measurements, KO cell class categories, and level of KO protein target in respective KO cells

Supplementary Table 2: MITOMICS quantitative dataset (log_2 biomolecule intensities)

Supplementary Table 3: MITOMICS quantitative dataset (log_2 fold-changes, standard deviations, and p-values)

Supplementary Table 4: IP-MS PYURF interactome

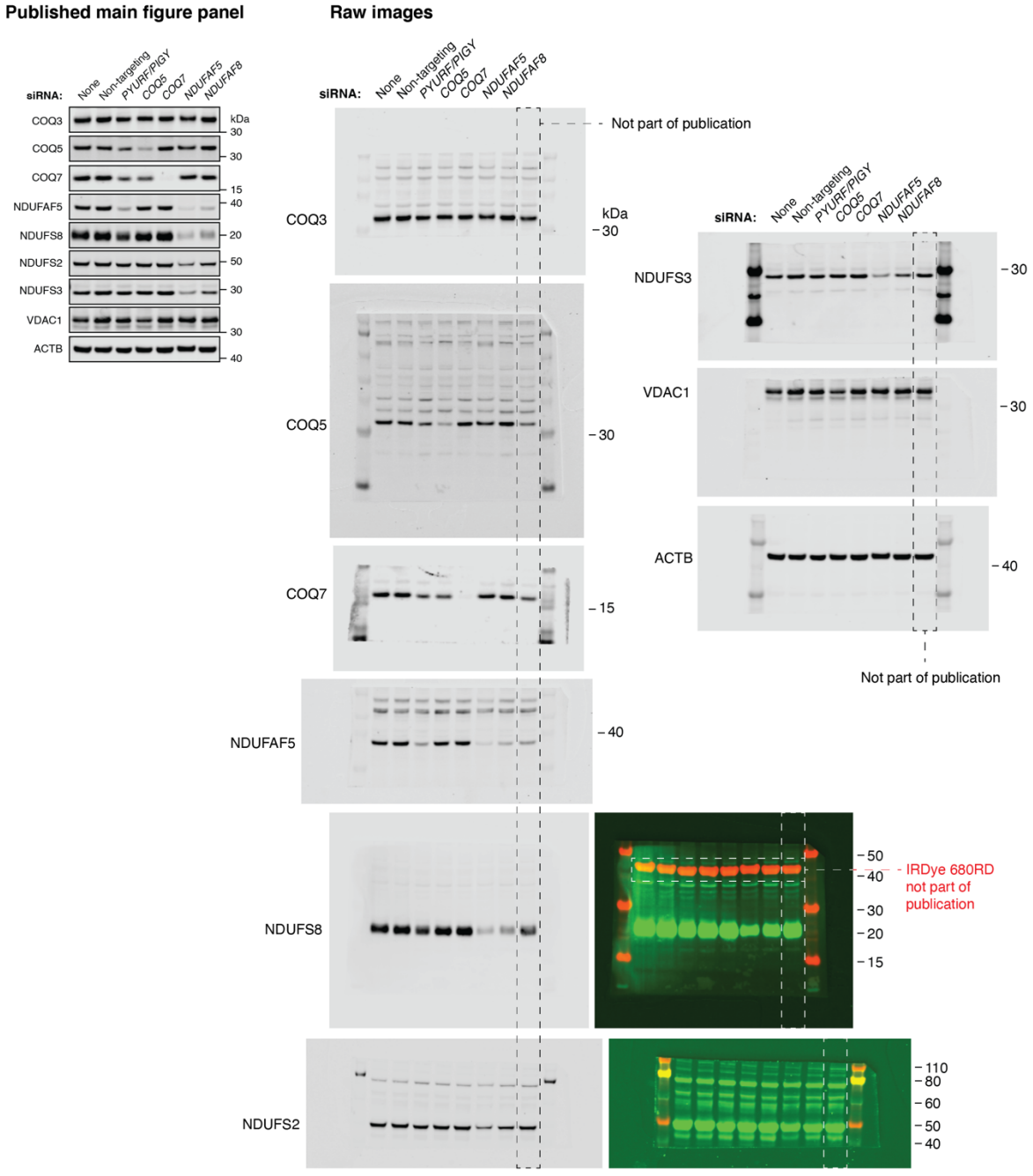
Supplementary Table 5: t-SNE MXP analysis results

Supplementary Table 6: KO-specific phenotype (outlier) analysis results

Supplementary Table 7: Knockout cell lines and reagent lists

The tables above contain data underlying graphical representations used in figures.

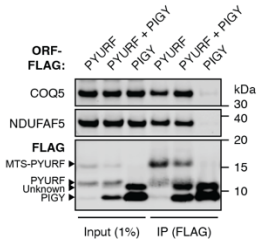
Supplementary Figure 1
Related to: Figure 3e



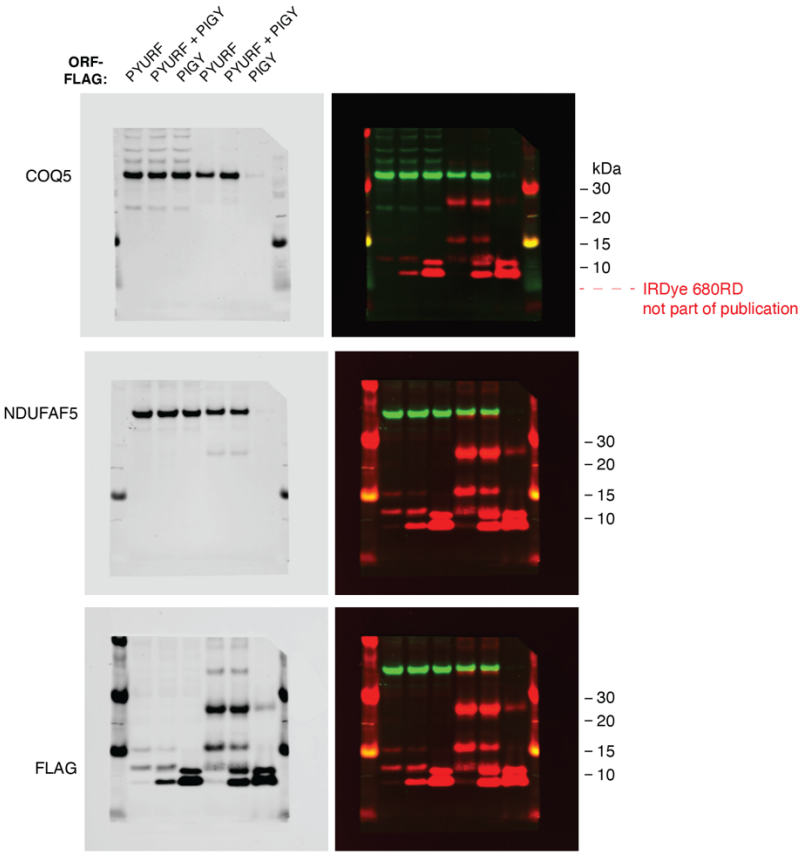
Supplementary Figure 1 | The uncropped versions of the cropped blots and gels in this study

Supplementary Figure 1, continued
Related to: Figure 3g

Published main figure panel

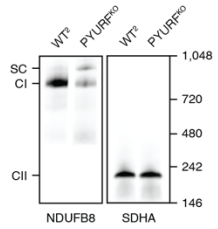


Raw images

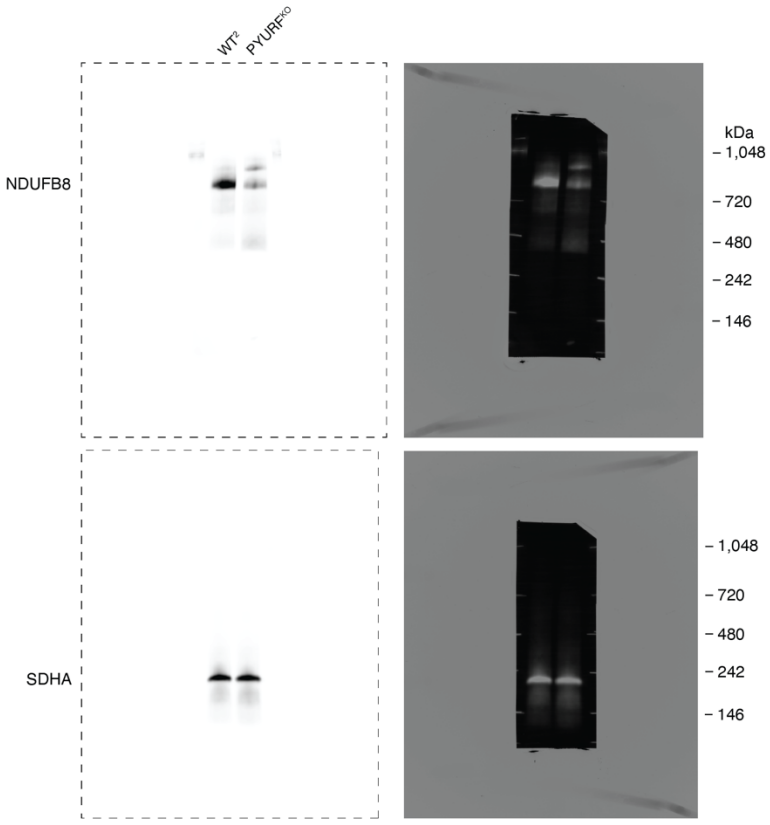


Supplementary Figure 1, continued
Related to: Figure 3i

Published main figure panel

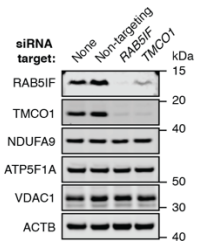


Raw images

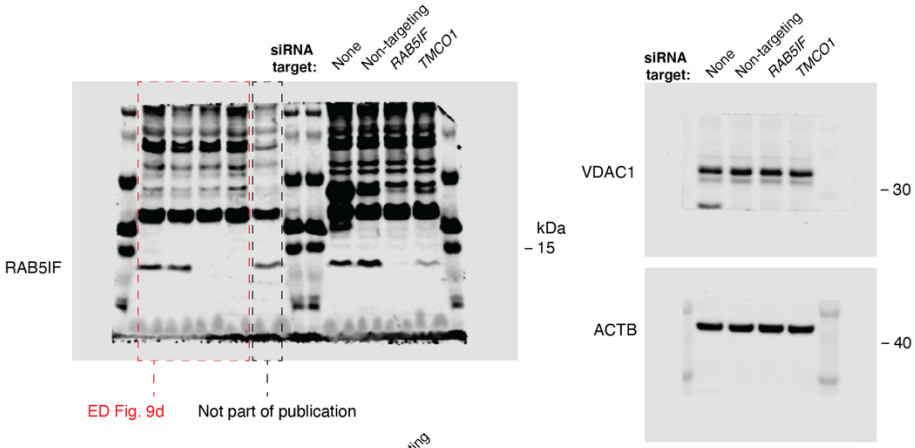


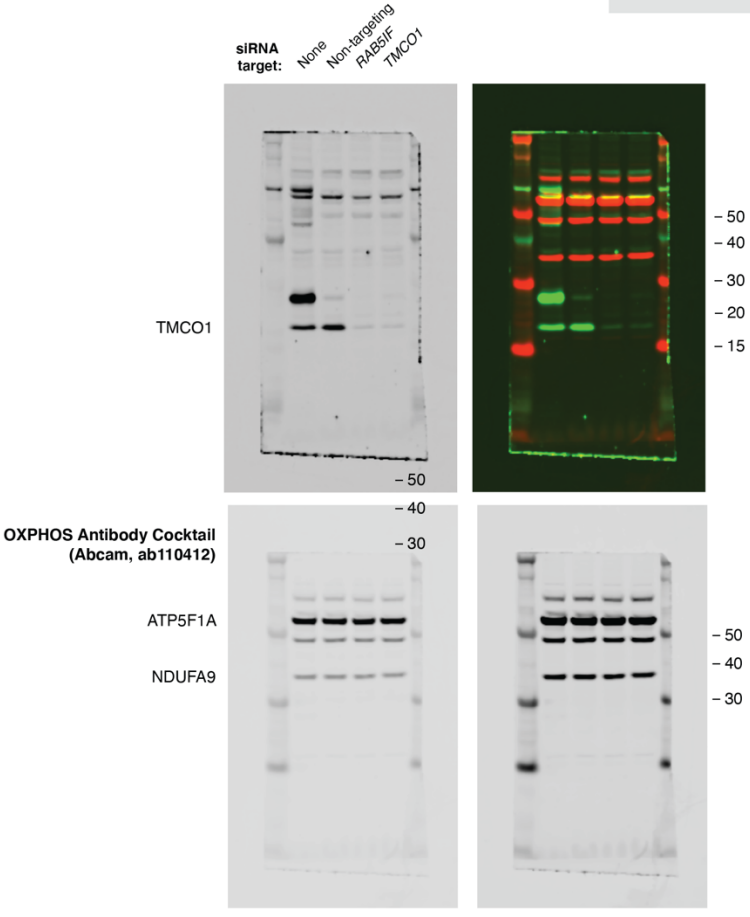
Supplementary Figure 1, continued
Related to: Figure 4h

Published main figure panel



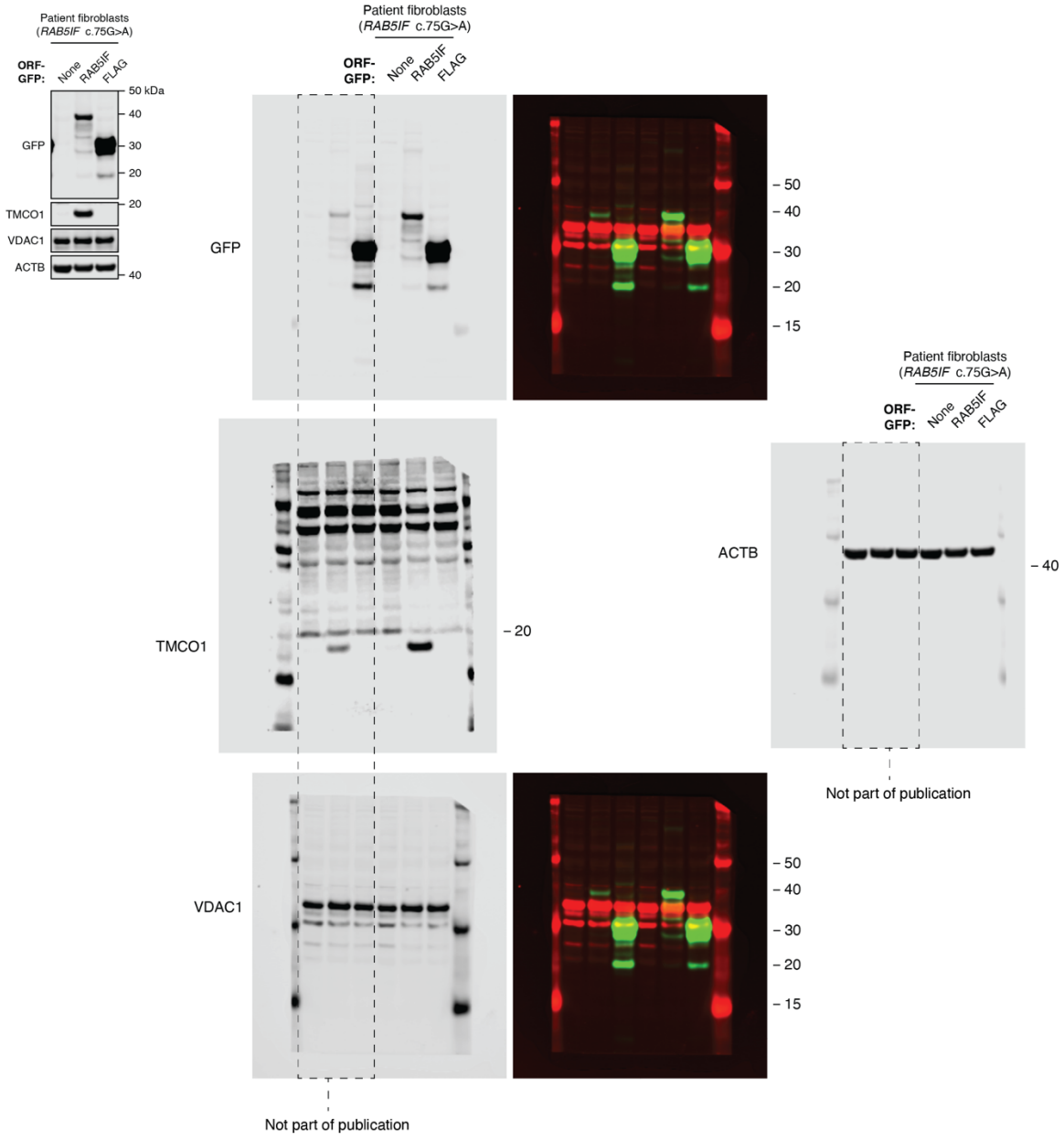
Raw images





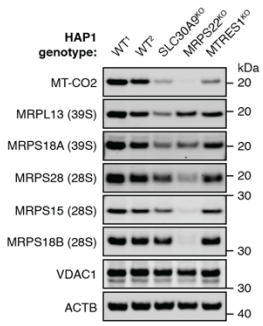
Supplementary Figure 1, continued
Related to: Figure 4l

Published main figure panel **Raw images**

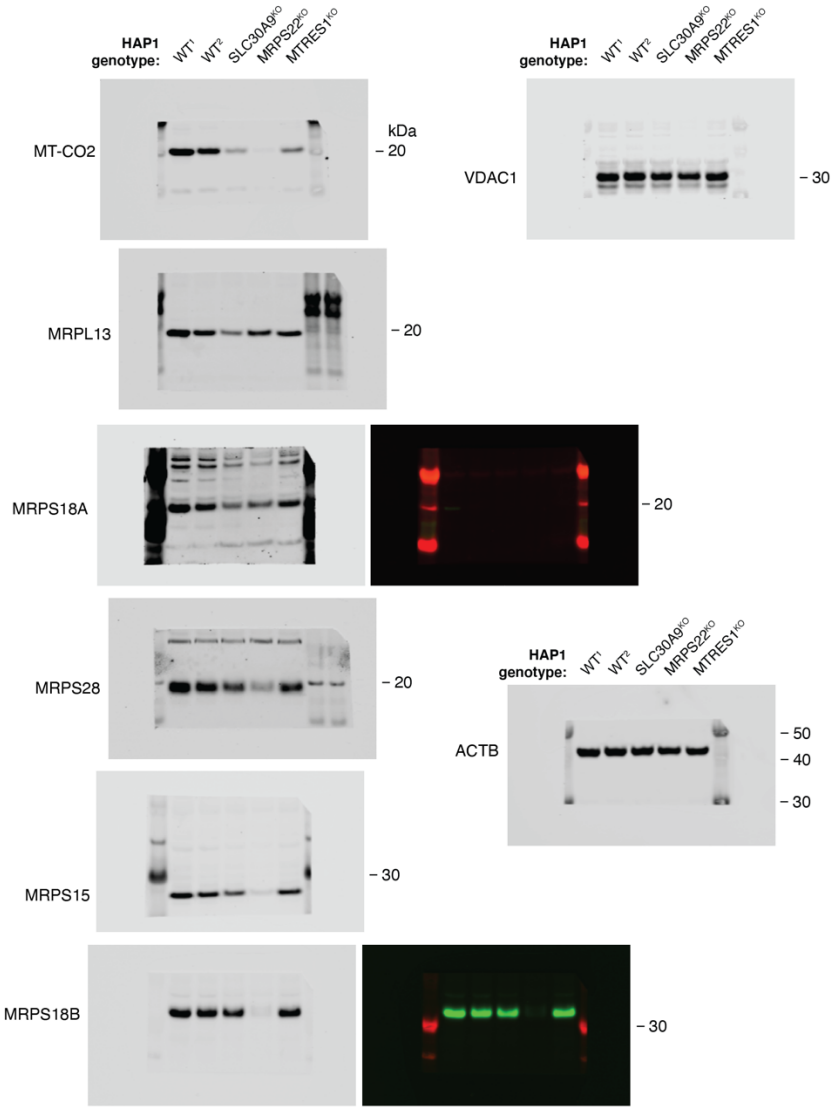


Supplementary Figure 1, continued
Related to: Extended Data Figure 3b

Published figure panel

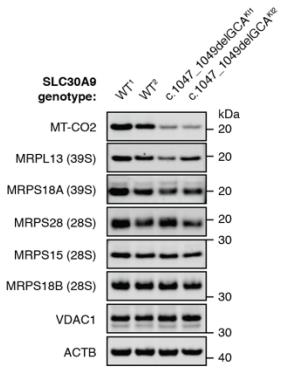


Raw images

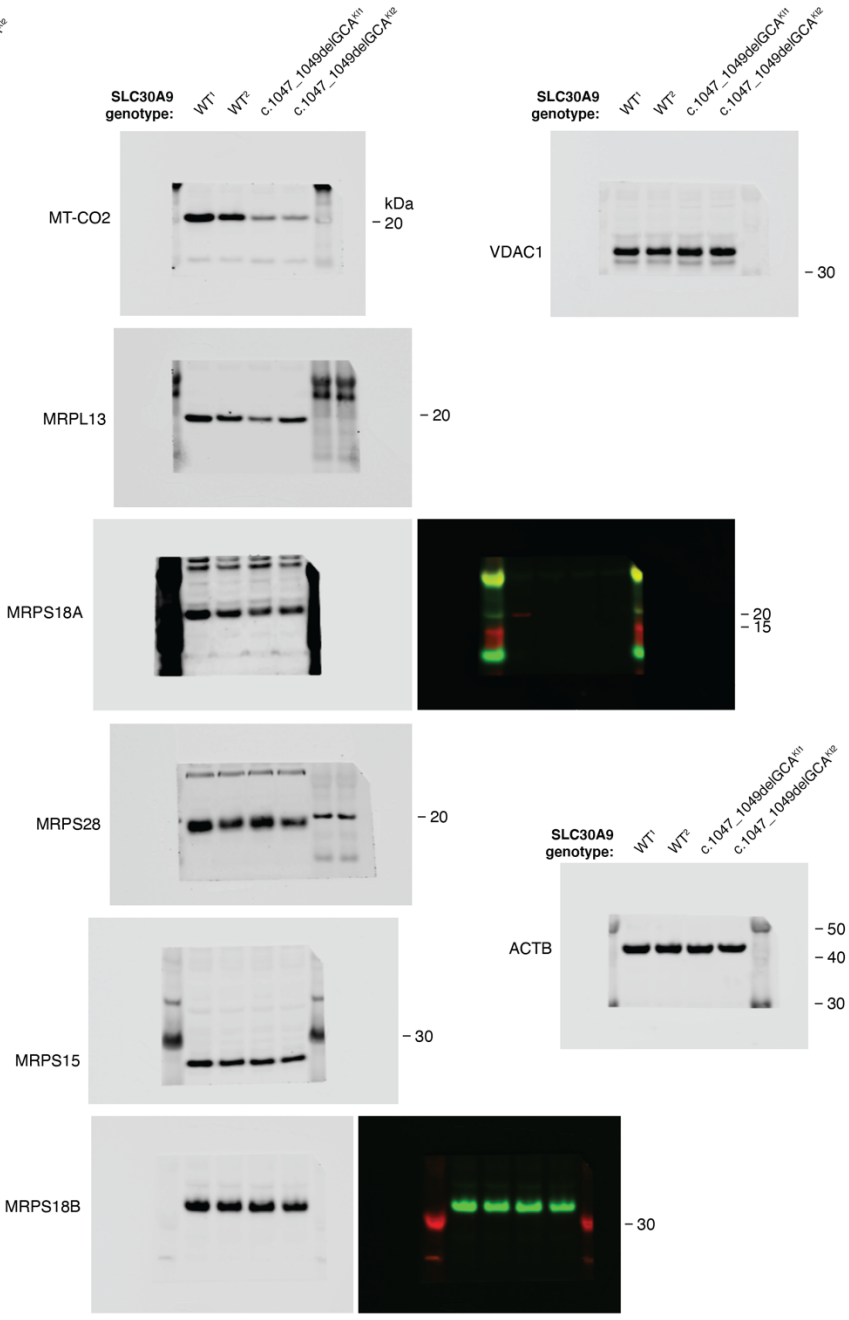


Supplementary Figure 1, continued
Related to: Extended Data Figure 3c

Published figure panel

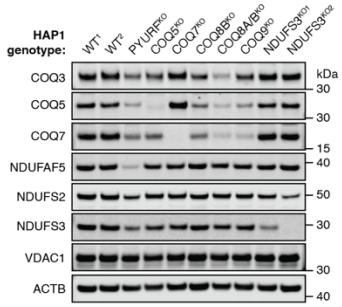


Raw images

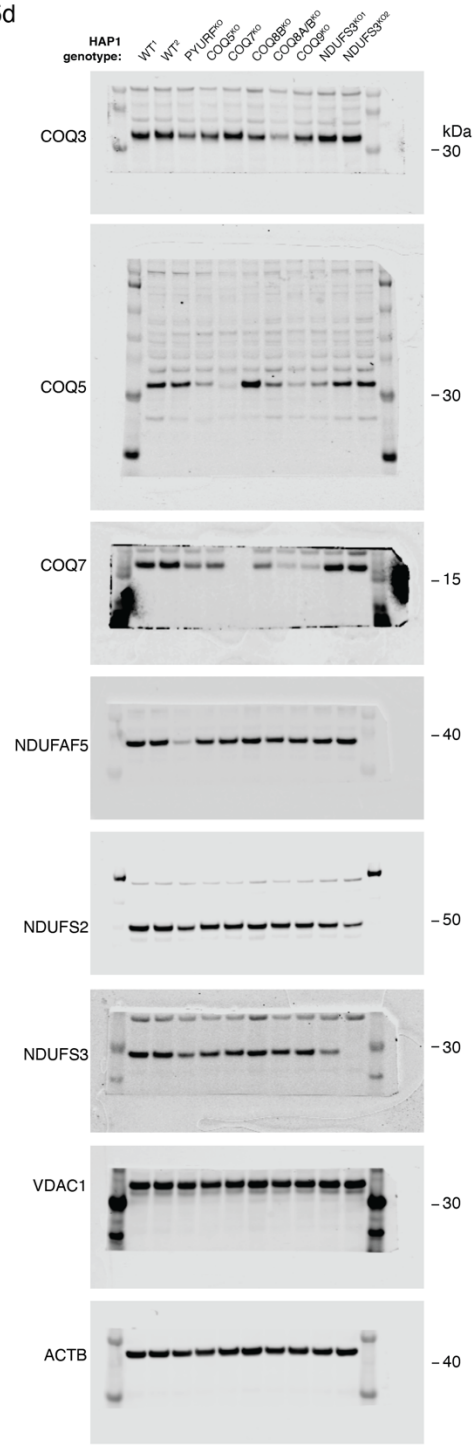


Supplementary Figure 1, continued
Related to: Extended Data Figure 5d

Published figure panel

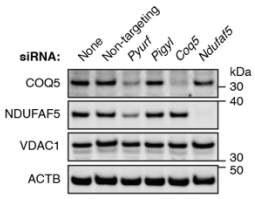


Raw images

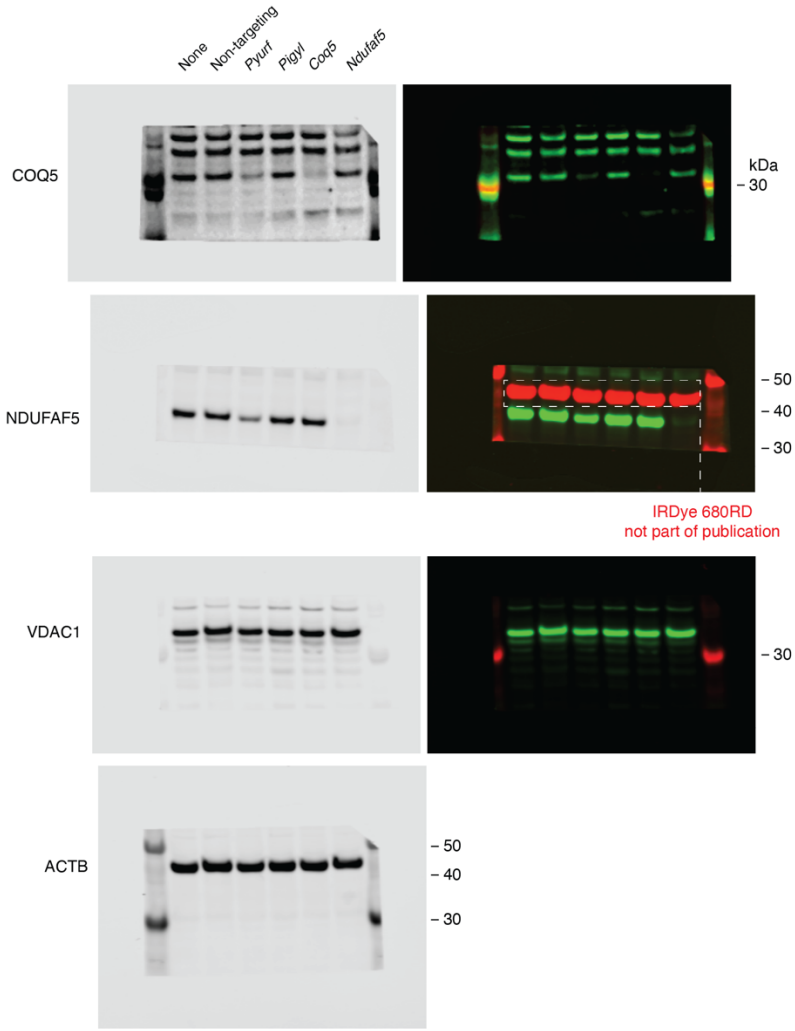


Supplementary Figure 1, continued
Related to: Extended Data Figure 5h

Published figure panel

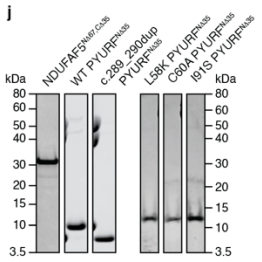


Raw images

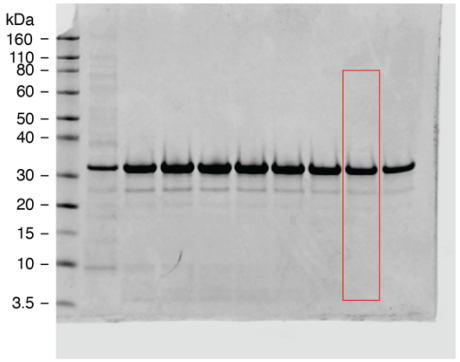


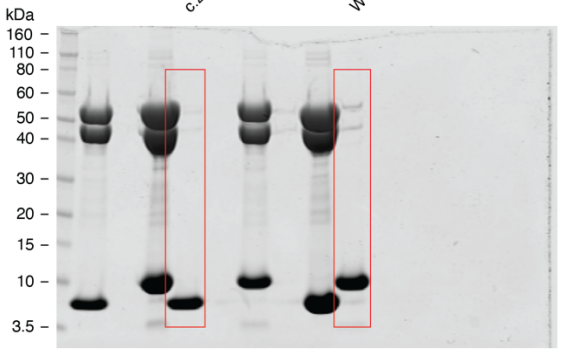
Supplementary Figure 1, continued
Related to: Extended Data Figure 5j

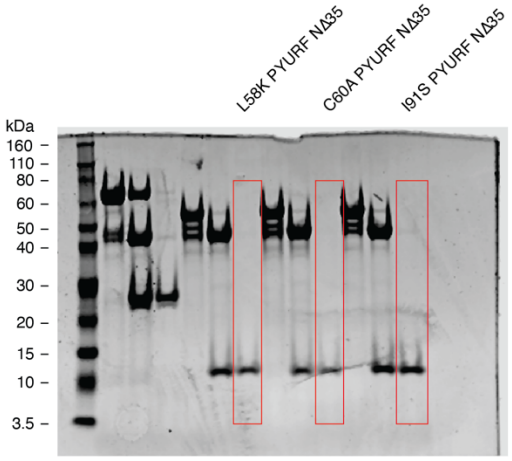
Published figure panel



Raw images

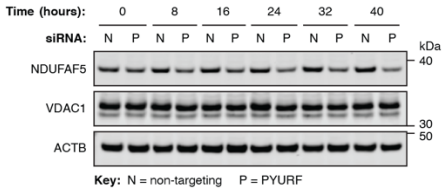




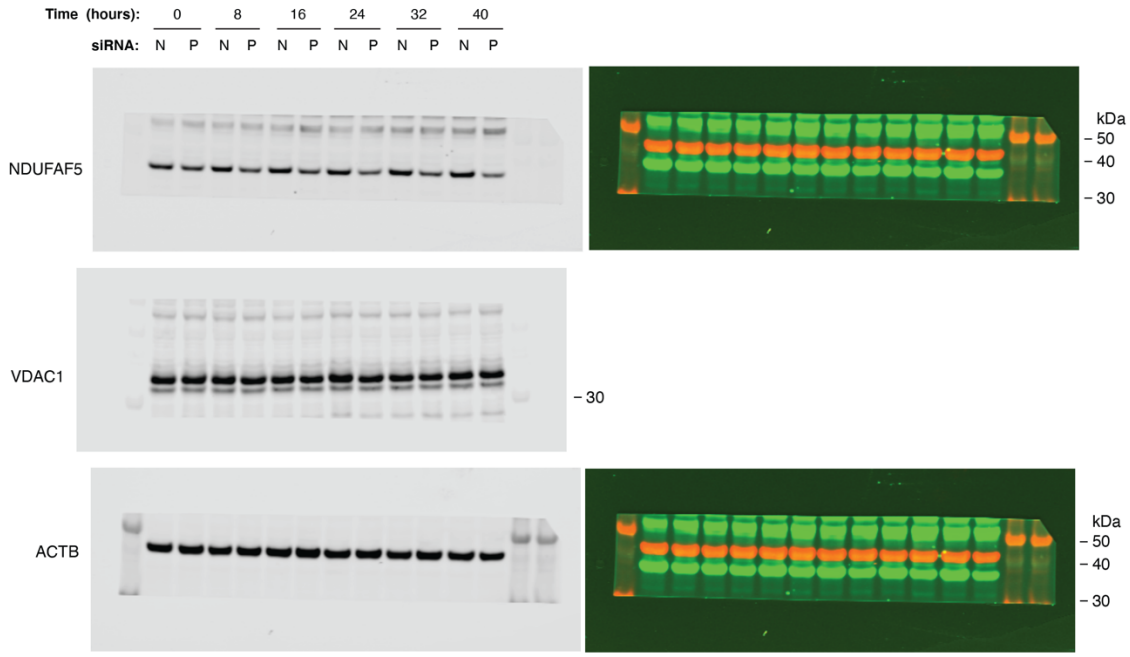


Supplementary Figure 1, continued
Related to: Extended Data Figure 7a

Published figure panel

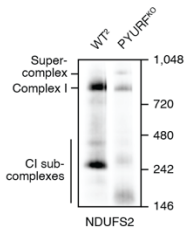


Raw images

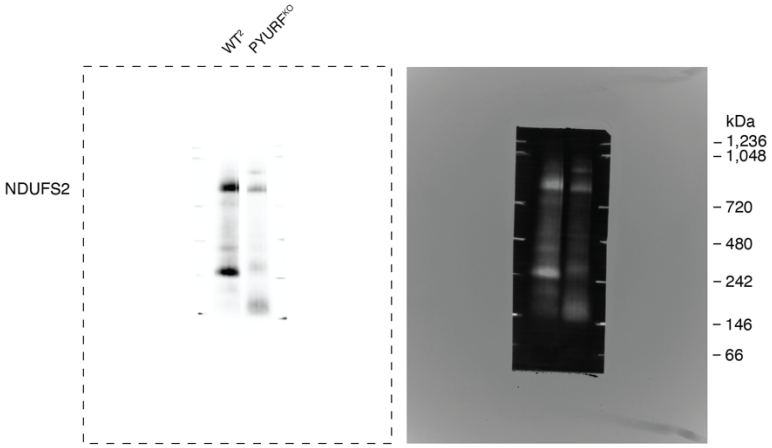


Supplementary Figure 1, continued
Related to: Extended Data Figure 7c

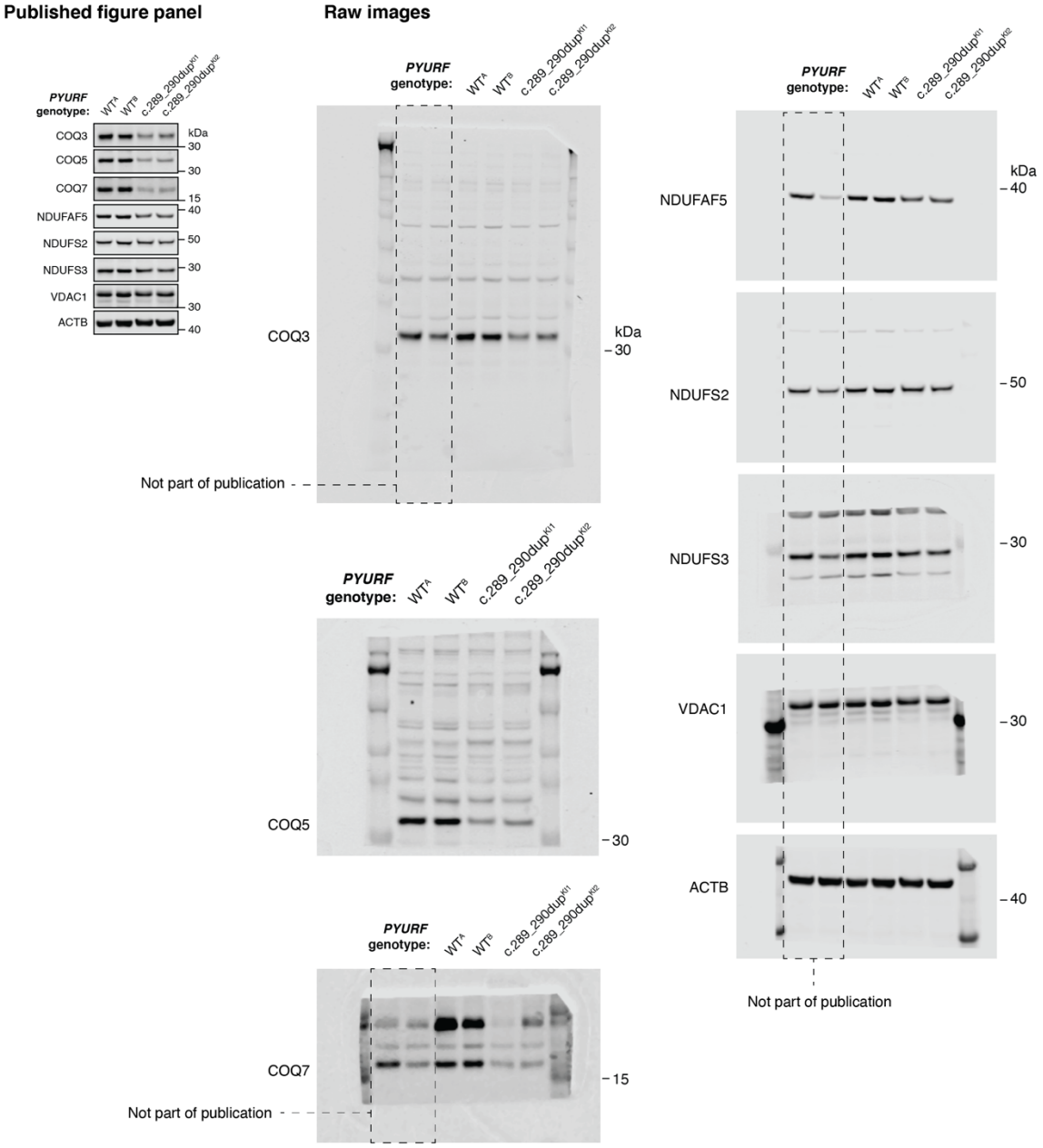
Published figure panel



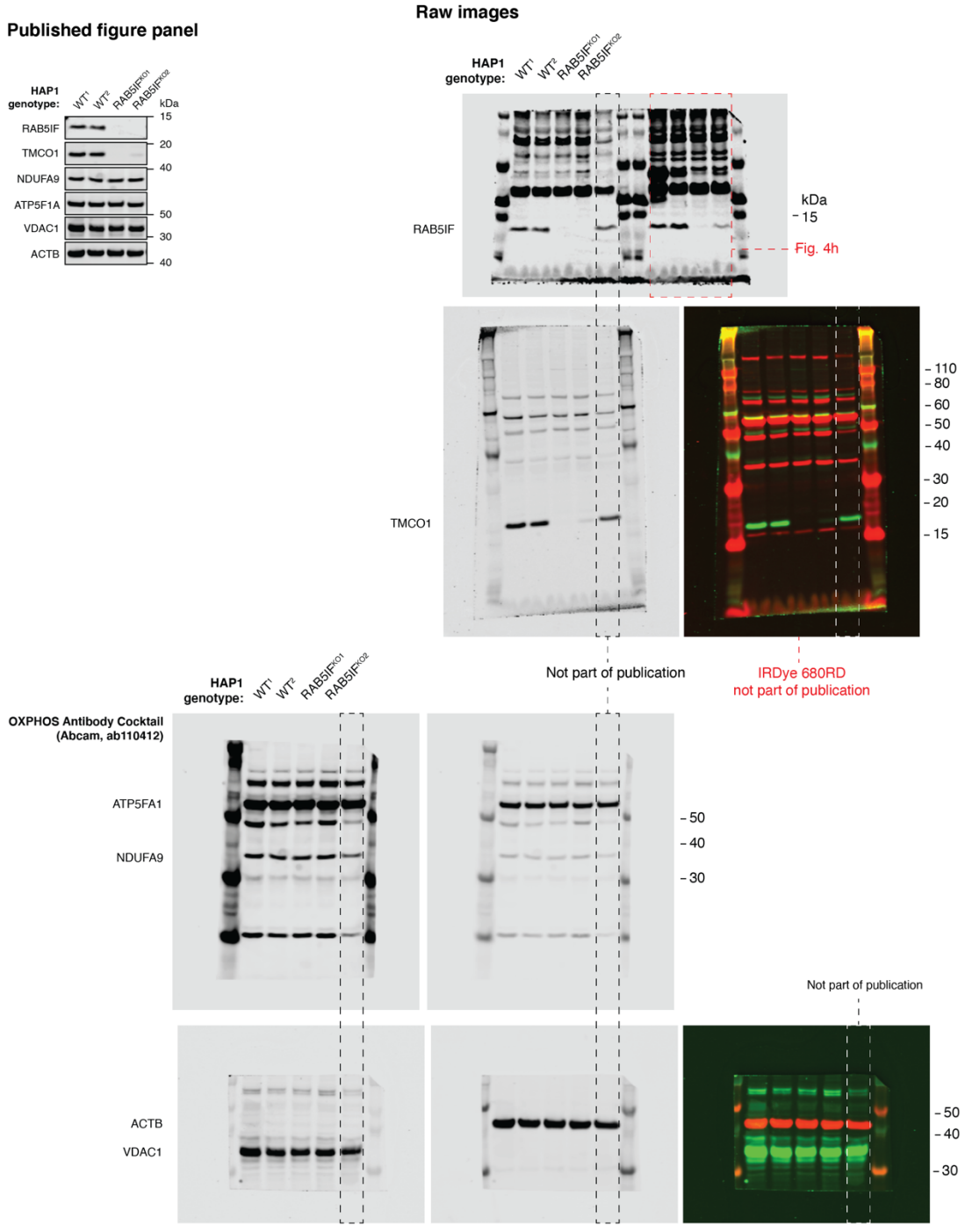
Raw images



Supplementary Figure 1, continued
Related to: Extended Data Figure 7g

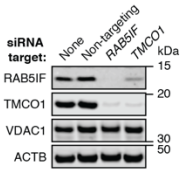


Supplementary Figure 1, continued
Related to: Extended Data Figure 9d

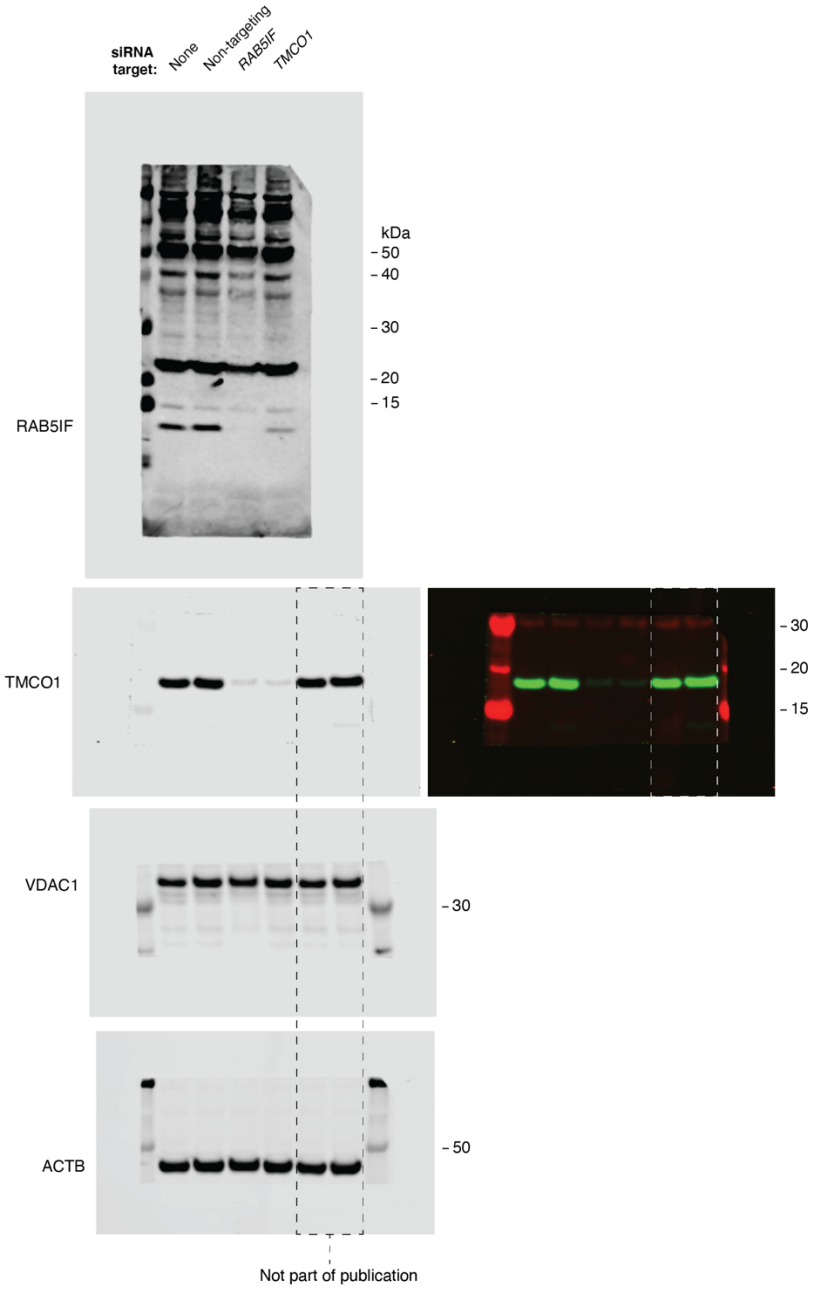


Supplementary Figure 1, continued
Related to: Extended Data Figure 9g

Published figure panel

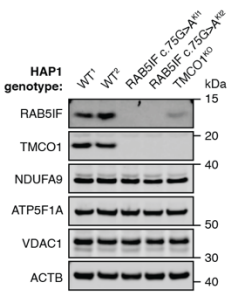


Raw images

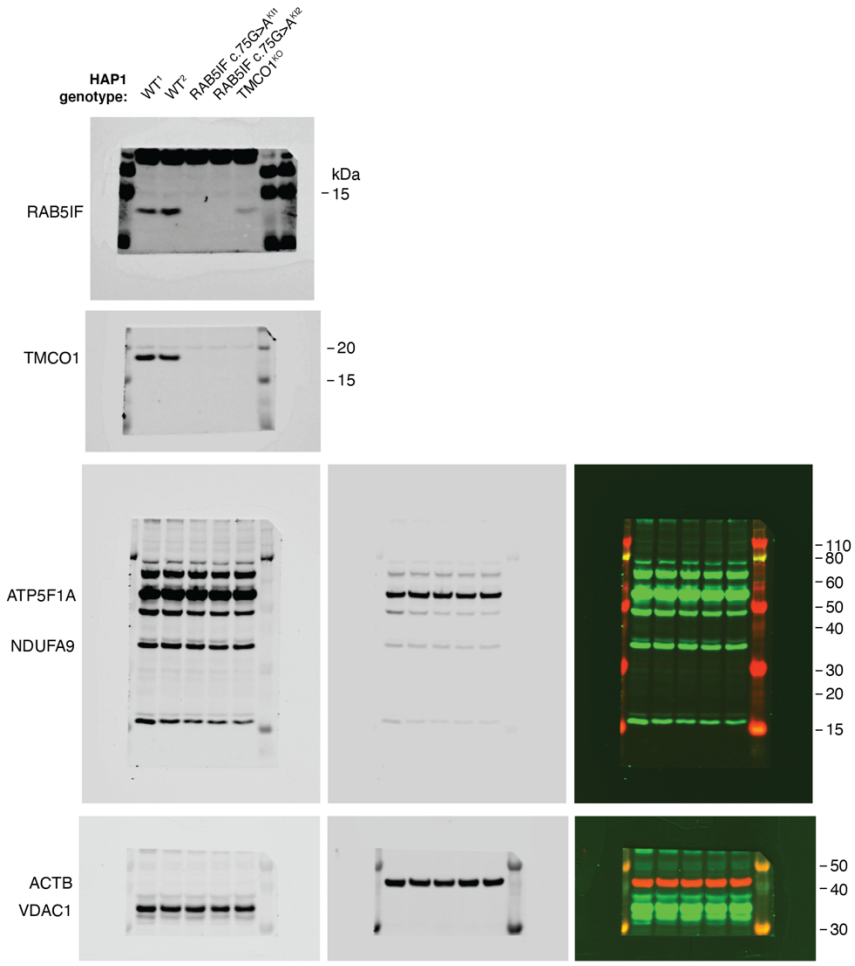


Supplementary Figure 1, continued
Related to: Extended Data Figure 9k

Published figure panel



Raw images



Supplementary Figure 1, continued
Related to: Extended Data Figure 9l